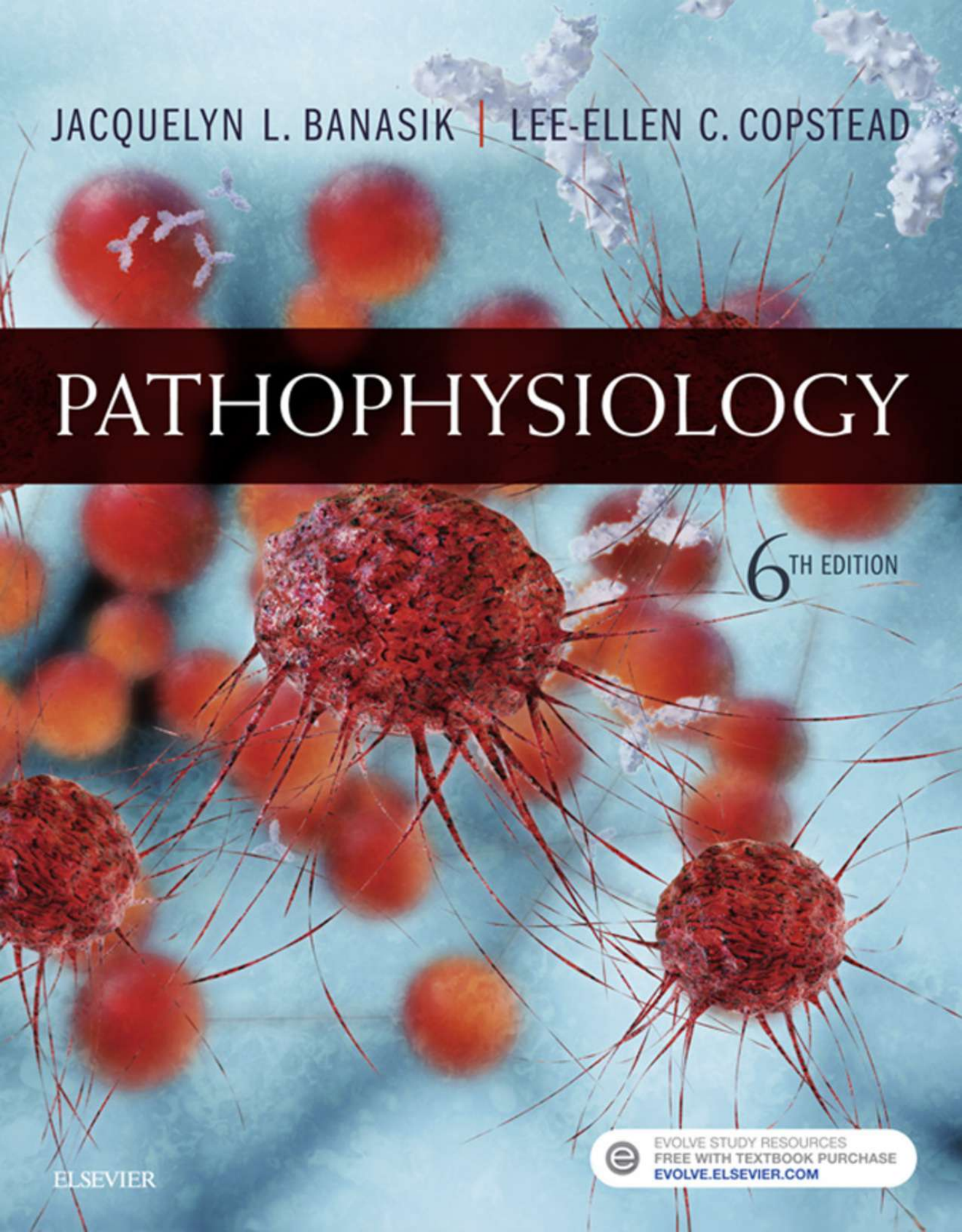

JACQUELYN L. BANASIK | LEE-ELLEN C. COPSTEAD
PATHOPHYSIOLOGY
6TH EDITION
EVOLVE STUDY RESOURCES
FREE WITH TEXTBOOK PURCHASE
EVOLVE.ELSEVIER.COM
ELSEVIER

BRIEF CONTENTS

UNIT I
Pathophysiologic Processes

1 Introduction to Pathophysiology, 1

2 Homeostasis, Allostasis, and Adaptive Responses to Stressors, 12

UNIT II
Cellular Function

3 Cell Structure and Function, 26

4 Cell Injury, Aging, and Death, 59

5 Genome Structure, Regulation, and Tissue Differentiation, 77

6 Genetic and Developmental Disorders, 94

7 Neoplasia, 117

UNIT III
Defense

8 Infectious Processes, 141

9 Inflammation and Immunity, 158

10 Alterations in Immune Function, 194

11 Malignant Disorders of White Blood Cells, 215

12 HIV Disease and AIDS, 233

UNIT IV
Oxygen Transport, Blood Coagulation, Blood Flow, and Blood Pressure

13 Alterations in Oxygen Transport, 259

14 Alterations in Hemostasis and Blood Coagulation, 298

15 Alterations in Blood Flow, 313

16 Alterations in Blood Pressure, 337

UNIT V
Cardiac Function

17 Cardiac Function, 354

18 Alterations in Cardiac Function, 382

19 Heart Failure and Dysrhythmias: Common Sequelae of Cardiac Diseases, 411

20 Shock, 434

UNIT VI
Respiratory Function

21 Respiratory Function and Alterations in Gas Exchange, 451

22 Obstructive Pulmonary Disorders, 478

23 Restrictive Pulmonary Disorders, 499

UNIT VII
Fluid, Electrolyte, and Acid-Base Homeostasis

24 Fluid and Electrolyte Homeostasis and Imbalances, 521

25 Acid–Base Homeostasis and Imbalances, 541

UNIT VIII
Renal and Bladder Function

26 Renal Function, 551

27 Intrarenal Disorders, 575

28 Acute Kidney Injury and Chronic Kidney Disease, 593

29 Disorders of the Lower Urinary Tract, 609

UNIT IX
Genital and Reproductive Function

30 **Male Genital and Reproductive Function,** 626

31 **Alterations in Male Genital and Reproductive Function,** 641

32 **Female Genital and Reproductive Function,** 656

33 **Alterations in Female Genital and Reproductive Function,** 671

34 **Sexually Transmitted Infections,** 689

UNIT X
Gastrointestinal Function

35 **Gastrointestinal Function,** 697

36 **Gastrointestinal Disorders,** 720

37 **Alterations in Function of the Gallbladder and Exocrine Pancreas,** 742

38 **Liver Diseases,** 754

UNIT XI
Endocrine Function, Metabolism, and Nutrition

39 **Endocrine Physiology and Mechanisms of Hypothalamic-Pituitary Regulation,** 783

40 **Disorders of Endocrine Function,** 799

41 **Diabetes Mellitus,** 815

42 **Nutritional and Metabolic Disorders,** 838

UNIT XII
Neural Function

43 **Structure and Function of the Nervous System,** 850

44 **Acute Disorders of Brain Function,** 891

45 **Chronic Disorders of Neurologic Function,** 915

46 **Alterations in Special Sensory Function,** 936

47 **Pain,** 955

UNIT XIII
Neuropsychological Function

48 **Neurobiology of Psychotic Illnesses,** 971

49 **Neurobiology of Nonpsychotic Illnesses,** 989

UNIT XIV
Musculoskeletal Support and Movement

50 **Structure and Function of the Musculoskeletal System,** 1001

51 **Alterations in Musculoskeletal Function: Trauma, Infection, and Disease,** 1020

52 **Alterations in Musculoskeletal Function: Rheumatic Disorders,** 1042

UNIT XV
Integumentary System

53 **Alterations in the Integumentary System,** 1058

54 **Burn Injuries,** 1093

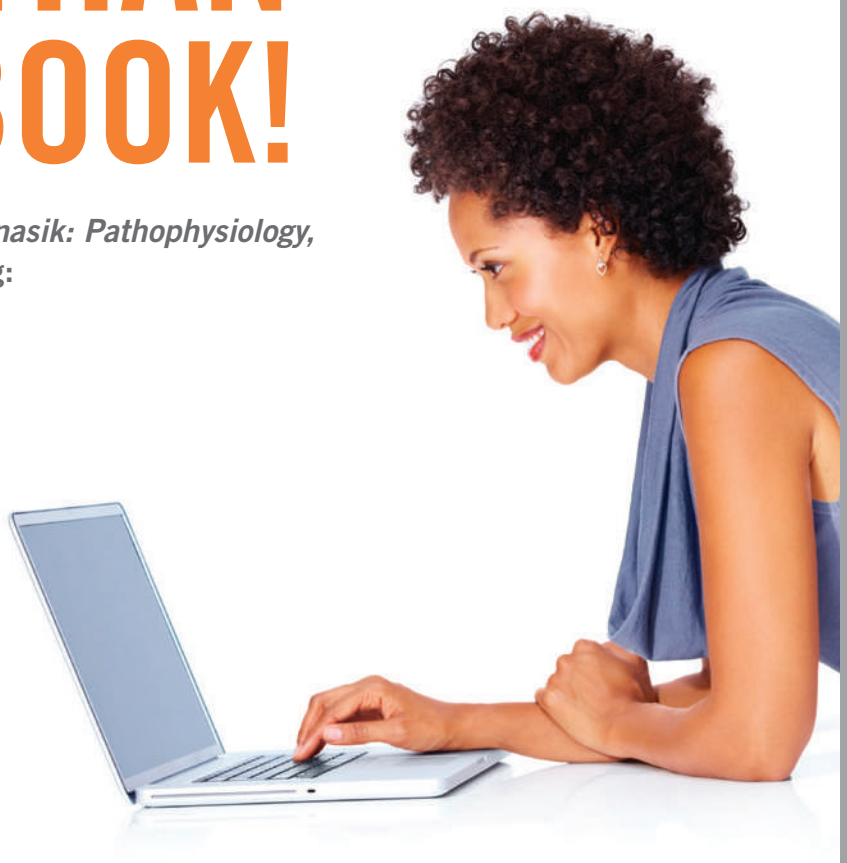
evolve

ELSEVIER

YOU'VE JUST PURCHASED MORE THAN A TEXTBOOK!

Evolve Student Resources for *Banasik: Pathophysiology, 6th Edition*, include the following:

- Animations.
- Review questions with rationales.
- Key Points.
- Glossary with audio pronunciations.
- Answers to key questions.



Activate the complete learning experience that comes with each textbook purchase by registering at

http://evolve.elsevier.com/Banasik/pathophysiology

REGISTER TODAY!

You can now purchase Elsevier products on Evolve!
Go to evolve.elsevier.com/html/shop-promo.html to search and browse for products.

PATHOPHYSIOLOGY

6TH EDITION

JACQUELYN L. BANASIK, PhD, ARNP

Associate Professor
College of Nursing
Washington State University
Spokane, Washington

LEE-ELLEN C. COPSTEAD, PhD, RN

Professor Emerita
Department of Nursing
College of Nursing and Health Sciences
University of Wisconsin—Eau Claire
Eau Claire, Wisconsin

ELSEVIER

ELSEVIER

3251 Riverport Lane
St. Louis, Missouri 63043

PATHOPHYSIOLOGY, SIXTH EDITION ISBN: 978-0-323-35481-3

Copyright © 2019 by Elsevier Inc. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds or experiments described herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made. To the fullest extent of the law, no responsibility is assumed by Elsevier, authors, editors or contributors for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Previous editions copyrighted 2013, 2010, 2005, 2000, and 1995.

International Standard Book Number: 978-0-323-35481-3

Executive Content Strategist: Kellie White
Content Development Specialist: Jennifer Wade
Publishing Service Manager: Jeffrey Patterson
Book Production Specialist: Carol O'Connell
Book Designer: Renee Duenow

Printed in Canada

Last digit is the print number: 9 8 7 6 5 4 3 2 1

To:
Loved ones, past and present,
who give meaning to the work.

JLB

This page intentionally left blank

CONTRIBUTORS

Brent A. Banasik, PhD, MA, BS
Scientist
Chemistry
Banasik Consulting
Seattle, Washington

Brianne N. Banasik, BS Environmental Science, MS Marine Resources Management
Research Associate
Pediatric Vaccinology
University of Texas Medical Branch
Galveston, Texas

Jacquelyn L. Banasik, PhD, ARNP
Associate Professor
College of Nursing
Washington State University
Spokane, Washington

Robin Y. Beeman, BSN, MSN, PhD
Chair and Professor
Nursing
University of Wisconsin—Eau Claire
Eau Claire, Wisconsin

Cheryl L. Brandt, PhD, RN, ACNS-BC
Professor
College of Nursing and Health Sciences
University of Wisconsin—Eau Claire
Eau Claire, Wisconsin

Ann Futterman Collier, PhD
Associate Professor and Chair
Psychological Sciences
Northern Arizona University
Flagstaff, Arizona

Lee-Ellen C. Copstead, PhD, RN
Professor Emerita
Department of Nursing
College of Nursing and Health Sciences
University of Wisconsin—Eau Claire
Eau Claire, Wisconsin

Carol L. Danning, MD
Staff Rheumatologist
Rheumatology Department
Gundersen Health Systems
La Crosse, Wisconsin

Michael R. Diestelmeier, MD
Fellow American Academy of Dermatology
Dermatologist, Retired
Mayo Clinic Health System
Eau Claire, Wisconsin

Ruth E. Diestelmeier, RN, MSN
Clinical Instructor
Department of Nursing
University of Wisconsin—Eau Claire
Eau Claire, Wisconsin

Roberta J. Emerson, PhD, RN
Associate Professor, Retired
Washington State University
College of Nursing
Spokane, Washington

Linda Felver, PhD, RN
Associate Professor
School of Nursing
Oregon Health & Science University
Portland, Oregon

Rosemary A. Jadack, PhD, RN
Professor
Nursing
University of Wisconsin—Eau Claire
Eau Claire, Wisconsin

Debra A. Jansen, PhD, RN
Associate Dean, Professor
College of Nursing and Health Sciences
University of Wisconsin—Eau Claire
Eau Claire, Wisconsin

Marie L. Kotter, PhD, MS, BS
Professor Emeritus
Health Sciences
Weber State University
Ogden, Utah

Teresa Grigsby Loftsgaarden, MSN, RN, OCN,ONN-CG
Oncology Nurse Navigator
Regional Cancer Center
Sacred Heart Hospital
Eau Claire, Wisconsin

Joni D. Marsh, BSN, MN
Advanced Registered Nurse Practitioner
Medical Oncology
Summit Cancer Centers
Spokane, Washington

Benjamin J. Miller, PhD, ARNP, FNP-C, ACNPC, ENP-C
Assistant Professor
Seattle University
Seattle, Washington

Sarah Ogle, DO, MS
Banner University Medical Center-Phoenix
University of Arizona College of Medicine Phoenix
Phoenix, Arizona

Nirav Patel, MD
Assistant Professor, Infectious Diseases and Critical Care Medicine
Internal Medicine
Saint Louis University School of Medicine;
Chief Medical Officer
SSM Health Saint Louis University Hospital;
Infection Control Officer/Director of Antibiotic Stewardship
SSM Health Saint Louis University Hospital
Saint Louis, Missouri

Faith Young Peterson, BSN, MSN, MPA, CFNP
Family Nurse Practitioner
Marsing, Idaho

Cheryl Rockwell, RN, MSN
Clinical Assistant Professor
Nursing Department
Indiana University-Purdue University at Fort Wayne;
Staff/Clinical Nurse
Surgical Trauma Intensive Care Unit
Parkview Health Systems
Fort Wayne, Indiana

Samantha Cody Russell, Psychology, MA
Graduate Student
Psychological Sciences
Northern Arizona University
Flagstaff, Arizona

Jeffrey S. Sartin, MD
Consulting Physician
Infectious Diseases
Infectious Disease and Epidemiology Associates;
Consulting Physician
Infectious Diseases
Nebraska Medicine;
Consulting Physician
Infectious Diseases
CHI Hospitals
Omaha, Nebraska

Lorna L. Schumann, PhD
Heritage UGM Women and Children's Clinic
Medical Clinic
Heritage Health
Coeur d'Alene, Idaho

Susan G. Trevithick, RN, MS, NE-BC
Compliance Officer
VA Salt Lake City Healthcare System
Salt Lake City, Utah

Marvin Van Every, MD
Staff Urologist
Urology Department
Gundersen Health Systems
La Crosse, Wisconsin

Linda D. Ward, PhD, FNP-C
Assistant Professor
College of Nursing
Washington State University
Spokane, Washington

REVIEWERS

Brianne N. Banasik, MS
Research Associate
Pediatric Vaccinology
University of Texas Medical Branch
Galveston, Texas

Deb Cipali, RN, BSN, MSN, EdD(c)
Des Moines Area Community College
Nursing Lab Coordinator/Adjunct Professor
Ankeny, Iowa

Janie Corbitt, RN, MLS
Milledgville, Georgia

Maria Fleurdeliz Cuyco, BS
Instructor
Preferred College of Nursing
Los Angeles, California

Abimbola Farinde, PhD
Professor
Columbia Southern University
Orange Beach, Alabama

Annette Gunderman, DEd, MSN, RN
Associate Professor of Nursing
Bloomsburg University
Bloomsburg, Pennsylvania

Barbara Hunter, RN, MSN
School of Nursing, Klamath Falls Campus
Oregon Health & Science University
Klamath Falls, Oregon

Sandra L. Kaminski, MS, PA-C
Assistant Professor
Seton Hall University
School of Health and Medical Sciences;
Physician Assistant
VA NJ Healthcare System
Medical Service/Infections Disease
Clark, New Jersey

Steven Krau, PhD, RN, CNE
Associate Professor
Vanderbilt School of Nursing
Nashville, Tennessee

Clarice Perry, MS
Research Associate
Pediatric Vaccinology
University of Texas Medical Branch
Galveston, Texas

Janet Pinkelman, MSN, RNC-Maternal Newborn Nursing
Professor of Nursing
Owens Community College
Toledo, Ohio

Linda Turchin, RN, MSN, CNE
Associate Professor of Nursing
Fairmont State University
Fairmont, West Virginia

Kim Webb, MN, RN
Adjunct Nursing Instructor
Pioneer Technology Center
Ponca City, Oklahoma

Janice Williams, RN, ACNS-BC, CDE
Professor of Nursing and Program Director
Armstrong McDonald School of Nursing
College of the Ozarks
Point Lookout, Missouri

PREFACE

The pace of scientific discovery in health and medicine continues to transform our understanding of physiology and disease. To be clinically relevant and useful to health care students and professionals, a text must synthesize a vast amount of detailed knowledge into overarching concepts that can be applied broadly. As in previous editions, the goal of the sixth edition of *Pathophysiology* is to include recent and relevant information on anatomy, biochemistry, cell physiology, genomics, and pathophysiology while not overwhelming the reader. Attention is given to major concepts relevant to clinical practice while still providing enough detail for deep understanding.

ORGANIZATION

Pathophysiology uses a systems approach to content, beginning with a review of normal anatomy and physiology, followed by pathophysiology and application of concepts to specific disorders. The text is organized into 15 units, each of which includes a particular body system or group of interrelated body systems and the pertinent pathophysiologic concepts and disorders.

FEATURES

An understanding of normal structure and function of the body is necessary for any detailed understanding of its abnormalities and pathophysiology. The first chapter in most units includes a fully illustrated *review of normal physiology.* Age-related concepts are highlighted in boxes titled *Geriatric Considerations* and *Pediatric Considerations*.

Each chapter opens with *Key Questions,* which are designed to alert the reader to important conceptual questions that will be discussed in the chapter. Although the chapters are meant to be read from beginning to end to develop an understanding of the material, the text also serves as a reference for looking up specific content. *Chapter Outlines* are included at the beginning of each chapter to help the reader locate specific content. Within every chapter, *Key Points* are identified at the end of every major discussion and are presented in short bulleted lists. These recurring summaries help readers to focus on the main points.

Nearly 900 illustrations elucidate both normal physiology and pathophysiologic changes. The entire book is in *full color,* with color used generously in the illustrations to better explain pathophysiologic concepts.

A study of pathophysiology requires a new vocabulary, and many of these terms are defined in a comprehensive *Glossary,* which appears at the end of the text. Common prefixes and suffixes as well as root words are included in the back matter to help with basic understanding of the language of pathophysiology.

ANCILLARIES

Student Learning Resources on Evolve

The student section of the book's website hosted on Evolve offers nearly 700 *Student Review Questions* in a variety of question formats, an *Audio Glossary, Animations* to help readers visualize pathophysiologic processes, *Case Studies with questions,* and *Key Points Review.* Visit the Evolve website at http://evolve.elsevier.com/Banasik/pathophysiology.

Study Guide

Pathophysiology can be a daunting subject for students because of the large volume of factual material to be learned. The student **Study Guide** is designed to help students focus on important pathophysiologic concepts. Questions to check recall of normal anatomy and physiology are included for each chapter. A number of activities that help the student focus on similarities and differences between often-confused pathologic processes are included. More than 1500 self-assessment test questions with answers are included to help students check their understanding and build confidence for examinations. Case studies, with more than 250 questions including rationales for correct and incorrect answers, are used to help students begin to apply pathophysiologic concepts to clinical situations.

Instructor Learning Resources on Evolve

The **Instructor Learning Resources** on Evolve provide a number of teaching aids for instructors who require the text for their students. The materials include a *Test Bank* presented in Exam View with approximately 1200 test items, a *Teach for Nurses instructor manual* detailing the resources available to instructors for their lesson planning, a *PowerPoint lecture guide* with more than 4000 slides with integrated case studies and audience response questions to facilitate classroom presentations, and an *Image Collection* of more than 900 color images from the text.

ACKNOWLEDGMENTS

Revising this 6th edition of the text has been possible because of the tremendous dedication of authors, artists, reviewers, and editors. Sincere gratitude goes to all who helped with this and previous editions. In particular, grateful appreciation is extended to all of the contributing authors who have given exhaustively of their time over many editions over the decades. Thank you to the many thoughtful experts who gave their time to read and critique manuscripts and help ensure excellence in chapter content throughout the text.

Grateful recognition is made to the staff at Elsevier: Kellie White, Executive Content Strategist; Jennifer Wade, Content Development Specialist; Jeffrey Patterson, Publishing Service Manager; Carol O'Connell, Book Production Specialist; Renee Duenow, Book Designer; and Vikraman Palani, Multimedia Producer.

We would like to recognize those who provided a foundation for the revised text through their contributions to earlier editions: Arnold A. Asp, Donna Bailey, Barbara Bartz, Linda Belsky-Lohr, Tim Brown, Carolyn Spenee Cagle, Karen Carlson, Katherina P. Choka, Arnold Norman Cohen, Cynthia F. Corbett, Lorri Dawson, Leslie Evans, Patricia Garber, Jane Georges, Karen Groth, Christine M. Henshaw, Carolyn Hoover, Jo Annalee Irving, Marianne Genge Jagmin, Debby Kaaland, Naomi Lungstrom, Rick Madison, Anne Roe Mealey, David Mikkelsen, Carrie Miller, Linda Denise Oakley, Maryann Pranulis, Mark Puhlman, Edith Randall, Bridget Recker, Cleo Richard, Dawn Rondeau, Mary Sanguinetti-Baird, Billie Marie Severtsen, Jacqueline Siegel, Gary Smith, Sheila Smith, Martha Snider, Pam Springer, Angela Starkweather, Patti Stec, Julie Symes, Lorie Wild, and Debra Winston-Heath.

CONTENTS

UNIT I Pathophysiologic Processes

1 Introduction to Pathophysiology, 1
Lee-Ellen C. Copstead
- **Framework for Pathophysiology, 2**
  - *Etiology, 2*
  - *Pathogenesis, 2*
  - *Clinical Manifestations, 3*
  - *Treatment Implications, 3*
- **Concepts of Normality in Health and Disease, 4**
  - *Statistical Normality, 4*
  - *Individual Factors Influencing Normality, 5*
- **Patterns of Disease in Populations, 6**
  - *Concepts of Epidemiology, 6*

2 Homeostasis, Allostasis, and Adaptive Responses to Stressors, 12
Debra A. Jansen and Roberta J. Emerson
- **Homeostasis and Allostasis, 12**
  - *Homeostasis, 12*
  - *Allostasis, 13*
- **Stress as a Concept, 13**
  - *The General Adaptation Syndrome and Allostasis, 14*
  - *Stressors, Gender and Developmental Influences, and Risk Factors, 16*
- **Neurohormonal Mediators of Stress and Adaptation, 17**
  - *Catecholamines: Norepinephrine and Epinephrine, 17*
  - *Adrenocortical Steroids: Cortisol and Aldosterone, 18*
  - *Endorphins, Enkephalins, and Immune Cytokines, 19*
  - *Sex Hormones: Estrogen, Testosterone, and Dehydroepiandrosterone, 19*
  - *Growth Hormone, Prolactin, and Oxytocin, 19*
- **Adaptation, Coping, and Illness, 20**
  - *Adaptation, Coping, and Resilience, 20*
  - *Allostatic Overload and Illness, 21*

UNIT II Cellular Function

3 Cell Structure and Function, 26
Jacquelyn L. Banasik and Brianne N. Banasik
- **Plasma Membrane, 27**
  - *Membrane Structure, 27*
  - *Lipid Bilayer, 27*
  - *Membrane Proteins, 29*
- **Organization of Cellular Compartments, 30**
  - *Cytoskeleton, 30*
  - *Nucleus, 30*
  - *Endoplasmic Reticulum, 31*
  - *Golgi Apparatus, 32*
  - *Lysosomes and Peroxisomes, 33*
  - *Mitochondria, 34*
- **Cellular Metabolism, 34**
  - *Glycolysis, 36*
  - *Citric Acid Cycle, 36*
  - *Oxidative Phosphorylation, 36*
- **Functions of the Plasma Membrane, 39**
  - *Membrane Transport of Macromolecules, 39*
  - *Membrane Transport of Small Molecules, 40*
  - *Cellular Membrane Potentials, 45*
- **Intercellular Communication and Growth, 48**
  - *Cell Signaling Strategies, 48*
  - *Cell Surface Receptor–Mediated Responses, 49*
  - *Intracellular Receptor–Mediated Responses, 53*
  - *Regulation of Cellular Growth and Proliferation, 54*

4 Cell Injury, Aging, and Death, 59
Jacquelyn L. Banasik
- **Reversible Cell Injury, 59**
  - *Hydropic Swelling, 60*
  - *Intracellular Accumulations, 60*
- **Cellular Adaptation, 63**
  - *Atrophy, 63*
  - *Hypertrophy, 63*
  - *Hyperplasia, 63*
  - *Metaplasia, 63*
  - *Dysplasia, 64*
- **Irreversible Cell Injury, 64**
  - *Necrosis, 64*
  - *Apoptosis, 66*
- **Etiology of Cellular Injury, 68**
  - *Ischemia and Hypoxic Injury, 68*
  - *Nutritional Injury, 70*
  - *Infectious and Immunologic Injury, 70*
  - *Chemical Injury, 72*
  - *Physical and Mechanical Injury, 72*
- **Cellular Aging, 74**
  - *Cellular Basis of Aging, 74*
  - *Physiologic Changes of Aging, 75*
- **Somatic Death, 75**

5 Genome Structure, Regulation, and Tissue Differentiation, 77
Jacquelyn L. Banasik
- **Structure and Function of DNA, 78**
  - *Structure of DNA, 78*
  - *DNA Replication, 79*
  - *Genetic Code, 80*
  - *Transcription, 81*
  - *Translation, 82*
- **Regulation of the Genome, 84**
  - *Transcriptional Controls, 84*
- **Differentiation of Tissues, 86**
  - *Cell Diversification and Cell Memory, 86*
  - *Mechanisms of Development, 86*
  - *Differentiated Tissues, 87*

6 Genetic and Developmental Disorders, 94
Linda D. Ward
- **Principles of Inheritance, 95**
  - *DNA Mutation and Repair, 96*

GENETIC DISORDERS, 98
Chromosomal Abnormalities, 99
Aberrant Number of Chromosomes, 99
Abnormal Chromosome Structure, 100
Examples of Autosomal Chromosome Disorders, 100
Examples of Sex Chromosome Disorders, 101
Mendelian Single-Gene Disorders, 102
Autosomal-Dominant Disorders, 103
Autosomal-Recessive Disorders, 104
Sex-Linked (X-Linked) Disorders, 106
Nonmendelian Single-Gene Disorders, 107
Anticipation, 109
Mitochondrial Gene Mutations, 109
Genomic Imprinting, 110
Polygenic and Multifactorial Disorders, 110
Environmentally Induced Congenital Disorders, 111
Periods of Fetal Vulnerability, 111
Teratogenic Agents, 111
Other Disorders of Infancy, 113
Diagnosis, Counseling, and Gene Therapy, 113
Prenatal Diagnosis and Counseling, 113
Genetic Analysis and Therapy, 114
Recombinant DNA Technology, 114

7 Neoplasia, 117
Jacquelyn L. Banasik
Benign Versus Malignant Growth, 118
Characteristics of Benign and Malignant Tumors, 118
Tumor Terminology, 118
The Malignant Phenotype, 118
Epidemiology and Cancer Risk Factors, 120
Tobacco Use, 120
Nutrition, 120
Genetic Mechanisms of Cancer, 123
Proto-Oncogenes, 124
Tumor Suppressor Genes, 127
Multistep Nature of Carcinogenesis, 129
Initiation, 129
Promotion, 131
Progression, 131
Metastasis, 132
Patterns of Spread, 132
Angiogenesis, 133
Grading and Staging of Tumors, 134
Effects of Cancer on the Body, 136
Cancer Therapy, 137
Surgery, 137
Radiation Therapy, 138
Drug Therapy, 138
Immunotherapy, 138
Gene and Molecular Therapy, 138
Stem Cell Transplantation, 139

UNIT III Defense

8 Infectious Processes, 141
Brent A. Banasik
Host–Microbe Relationship, 142
The Human Microbiome, 142
Host Characteristics, 142
Pathogen Characteristics, 144
Transmission of Infection, 147
Routes of Transmission, 148
Emerging Infectious Diseases, 148
Weapons of Bioterrorism, 149
Types of Pathogenic Organisms, 149
Bacteria, 149
Viruses, 150
Fungi, 151
Parasites, 152

9 Inflammation and Immunity, 158
Jacquelyn L. Banasik
COMPONENTS OF THE IMMUNE SYSTEM, 159
Epithelial Barriers, 159
Mononuclear Phagocyte System, 159
Lymphoid System, 160
Primary Lymphoid Organs, 160
Secondary Lymphoid Organs, 161
Leukocytes, 162
Neutrophils, 162
Eosinophils, 163
Basophils and Mast Cells, 164
Monocytes and Macrophages, 164
Dendritic Cells, 165
Lymphocytes, 165
Chemical Mediators of Immune Function, 167
Complement, 167
Kinins, 169
Clotting Factors, 169
Cytokines and Chemokines, 169
INNATE DEFENSES AND INFLAMMATION, 169
Inflammation, 171
Increased Vascular Permeability, 171
Emigration of Leukocytes, 172
Phagocytosis, 172
Chronic Inflammation, 174
Healing, 174
Inflammatory Exudates, 175
Systemic Manifestations of Inflammation, 175
SPECIFIC ADAPTIVE IMMUNITY, 176
Major Histocompatibility Complex, 176
Antigen Presentation by MHC, 176
MHC Class I Presentation, 177
MHC Class II Presentation, 177
Mechanisms of Cell-Mediated Immunity, 178
T Helper Cells ($CD4^+$), 178
Cytotoxic T Cells ($CD8^+$), 181
Mechanisms of Humoral Immunity, 181
Antigen Recognition by B Cells, 181
Antibody Structure, 183
Class Switching and Affinity Maturation, 185
Antibody Functions, 186
Passive and Active Immunity, 187
Passive Immunity, 187
Active Immunity, 187
INTEGRATED FUNCTION AND REGULATION OF THE IMMUNE SYSTEM, 189
Integrated Response to Microbial Antigen, 189
Integrated Response to Viral Antigen, 191
Regulation of Immune Function, 192

10 Alterations in Immune Function, 194
Faith Young Peterson
EXCESSIVE IMMUNE RESPONSES, 195
Autoimmunity, 195
Genetic Factors, 196
Environmental Triggers, 196
Pharmacotherapies, 197
Hypersensitivity, 198
Type I Hypersensitivity, 199
Type IIa Hypersensitivity, 201
Type IIb Hypersensitivity, 204
Type III Hypersensitivity, 204
Type IV Hypersensitivity, 207
DEFICIENT IMMUNE RESPONSES, 210
Primary Immunodeficiency Disorders, 210
B-Cell and T-Cell Combined Disorders, 210
T-Cell Disorders, 211
B-Cell Disorders, 212
Secondary Immunodeficiency Disorders, 212
11 Malignant Disorders of White Blood Cells, 215
Marie L. Kotter and Jacquelyn L. Banasik
Classification of Hematologic Neoplasms, 215
Etiology of Myeloid and Lymphoid Neoplasms, 216
General Principles of Management, 217
Diagnosis of Hematologic Neoplasms, 217
Principles of Treatment, 218
Prevention and Management of Complications, 219
Myeloid Neoplasms, 221
Chronic Myeloid Leukemia, 221
Acute Myeloid Leukemia, 222
Lymphoid Neoplasms, 223
Chronic Lymphoid Leukemia, 223
Acute Lymphoblastic Leukemia/Lymphoma, 223
Hairy Cell Leukemia, 224
Plasma Cell Myeloma (Multiple Myeloma), 224
Hodgkin Disease, 227
B-Cell, T-Cell, and NK-Cell Lymphoma (Non-Hodgkin), 229
12 HIV Disease and AIDS, 233
Faith Young Peterson
Epidemiology, 234
History, 234
Types of HIV, 234
Transmission, 236
Prevention of Transmission, 238
Etiology, 239
HIV Structure, 239
HIV Binding and Infection, 239
Pathogenesis, 242
Effect of HIV on Immune Cells at the Cellular Level, 242
Viral Production and Cell Death, 242
Progression of HIV Infection From Seroconversion to AIDS, 244
CDC HIV Classification System, 245
Diagnostic Testing, 245
Monitoring the Progression of HIV, 246
Clinical Manifestations, 248
Systemic Manifestations, 248
Gastrointestinal Manifestations, 249
Pulmonary Manifestations, 249
Mucocutaneous Manifestations, 250
Gynecologic Manifestations, 251
Neurologic Manifestations, 252
Ocular Manifestations, 253
Cardiovascular Manifestations, 253
Manifestations in Other Systems, 253
Manifestations in Children, 254
Treatment, 254
Antiretroviral Therapy Recommendations, 254
Nucleoside Reverse Transcriptase Inhibitors, 255
Nucleotide Reverse Transcriptase Inhibitors, 255
Nonnucleoside Reverse Transcriptase Inhibitors, 255
Protease Inhibitors, 257
Fusion Inhibitors, 257
CCR5 Inhibitors, 257
Integrase Strand Transfer Inhibitors, 257
Other Treatments and Vaccines, 257

UNIT IV Oxygen Transport, Blood Coagulation, Blood Flow, and Blood Pressure

13 Alterations in Oxygen Transport, 259
Susan G. Trevithick
Composition of Blood, 260
Organic and Inorganic Components, 261
Cellular Components, 261
Structure and Function of Red Blood Cells, 262
Hematopoiesis, 263
Hemoglobin Synthesis, 265
Nutritional Requirements for Erythropoiesis, 266
Energy and Maintenance of Erythrocytes, 267
Red Cell Production, 267
Red Cell Destruction, 268
Gas Transport and Acid–Base Balance, 268
Oxygen Transport, 268
Carbon Dioxide Transport, 270
Alterations in Oxygen Transport, 270
Anemia, 272
General Effects of Anemia, 272
Anemia Related to Decreased Red Cell Production, 277
Aplastic Anemia, 277
Anemia of Chronic Renal Failure, 277
Anemia Related to Vitamin B_{12} (Cobalamin) or Folate Deficiency, 278
Iron Deficiency Anemia, 279
Anemia Related to Inherited Disorders of the Red Cell, 279
Thalassemia, 279
Sickle Cell Anemia, 281
Hereditary Spherocytosis, 282
Glucose-6-Phosphate Dehydrogenase Deficiency, 282
Anemia Related to Extrinsic Red Cell Destruction or Loss, 284
Hemolytic Disease of the Newborn, 284
Antibody-Mediated Drug Reactions, 284
Acute Blood Loss, 285
Other Extrinsic Abnormalities, 285
Transfusion Therapy, 286
Polycythemia, 286

Polycythemia Vera, 286
Secondary Polycythemia, 292
Relative Polycythemia, 293

14 Alterations in Hemostasis and Blood Coagulation, 298
Cheryl Rockwell
The Process of Hemostasis, 298
Stages of Hemostasis, 298
Platelets, 299
Blood Coagulation Factors, 299
Fibrin Clot, 299
Fibrinolysis, 301
Evaluation of Hemostasis and Coagulation, 301
Clinical Assessment, 301
Laboratory Tests, 304
Vascular and Platelet Disorders, 304
Vascular Disorders, 304
Platelet Disorders, 306
Coagulation Disorders, 308

15 Alterations in Blood Flow, 313
Teresa Grigsby Loftsgaarden
Organization of the Circulatory and Lymphatic Systems, 315
Vessel Structure, 315
Lymphatic Structure, 317
Principles of Flow, 317
Hemodynamics of the Circulatory System, 317
Control of Flow, 322
Control of Blood Flow, 322
Control of Lymphatic Flow, 323
General Mechanisms That Cause Altered Flow, 323
Blood Vessels: Obstructions, 323
Blood Vessels: Structural Alterations, 325
Lymphatic Vessels, 326
Alterations in Arterial Flow, 326
Arteriosclerosis/Atherosclerosis, 326
Thromboangiitis Obliterans (Buerger Disease), 330
Raynaud Syndrome, 330
Aneurysms, 330
Acute Arterial Occlusion, 331
Alterations in Venous Flow, 332
Valvular Incompetence, 332
Varicose Veins, 332
Chronic Venous Insufficiency, 333
Deep Vein Thrombosis, 333
Alterations in Lymphatic Flow, 333
Lymphedema, 333

16 Alterations in Blood Pressure, 337
Benjamin J. Miller
Arterial Blood Pressure, 337
Determinants of Systemic Blood Pressure, 337
Measurement of Blood Pressure, 338
Mechanisms of Blood Pressure Regulation, 341
Short-Term Regulation of Systemic Blood Pressure, 341
Long-Term Regulation of Systemic Blood Pressure, 341
Normal Fluctuations in Systemic Blood Pressure, 343
Hypertension, 343
Definition and Classification, 343
Primary Hypertension, 344
Secondary Hypertension, 348
Hypertensive Emergencies and Urgency, 349
Low Blood Pressure, 350

UNIT V Cardiac Function

17 Cardiac Function, 354
Jacquelyn L. Banasik
Cardiovascular Anatomy, 355
Heart, 355
Circulatory System, 356
Cardiac Cycle, 358
Isovolumic Contraction, 359
Ventricular Ejection, 360
Isovolumic Relaxation, 360
Atrial Events, 360
Aortic and Pulmonary Artery Events, 360
Coronary Circulation, 360
Anatomy of the Coronary Vessels, 360
Regulation of Coronary Blood Flow, 361
Cardiac Myocytes, 363
Myocyte Structure, 363
Structure of the Contractile Apparatus, 363
Characteristics of Contractile Filaments, 365
Molecular Basis of Contraction, 366
Overview of Contraction, 366
Sliding Filament/Cross-Bridge Theory of Muscle Contraction, 366
Role of Calcium in Muscle Contraction, 367
Energy of Muscle Relaxation, 367
Cardiac Energy Metabolism, 368
Oxygen Utilization, 368
Substrate Utilization, 369
Cardiac Electrophysiology, 369
Cardiac Resting Potential, 369
Cardiac Action Potential, 369
Rhythmicity of Myocardial Cells, 370
Specialized Conduction System of the Heart, 371
Autonomic Regulation of Rhythmicity, 372
Electrocardiography, 372
Determinants of Cardiac Output, 374
Determinants of Heart Rate, 374
Determinants of Stroke Volume, 375
Cardiac Workload, 376
Endocrine Function of the Heart, 376
Tests of Cardiac Function, 376
Electrocardiography, 376
Magnetic Resonance Imaging and Computed Tomography, 378
Echocardiography, 378
Nuclear Cardiography, 379
Cardiac Catheterization/Coronary Angiography, 379

18 Alterations in Cardiac Function, 382
Jacquelyn L. Banasik
Coronary Heart Disease, 383
Etiology of Coronary Heart Disease, 383
Risk Factors and Mechanisms of Coronary Atherosclerosis, 383
Pathophysiology of Ischemia, 385
Clinical Features and Management of Coronary Syndromes, 388
Endocardial and Valvular Diseases, 394
Disorders of the Mitral Valve, 394
Disorders of the Aortic Valve, 397
Diseases of the Endocardium, 398

Myocardial Diseases, 399
Myocarditis, 399
Cardiomyopathy, 400
Pericardial Diseases, 402
Pericardial Effusion, 402
Pericarditis, 402
Congenital Heart Diseases, 403
Embryologic Development, 403
Etiology and Incidence of Congenital Heart Disease, 404
Pathophysiology of Congenital Heart Disease, 405
Acyanotic Congenital Defects, 406
Cyanotic Congenital Defects, 408

19 Heart Failure and Dysrhythmias: Common Sequelae of Cardiac Diseases, 411
Benjamin J. Miller and Jacquelyn L. Banasik
Heart Failure, 412
Pathogenesis and Diagnosis, 412
Compensatory Mechanisms, Remodeling, and Progression, 413
Clinical Manifestations, 417
Class and Stage of Heart Failure, 420
Treatment, 421
Cardiac Dysrhythmias, 421
Dysrhythmia Mechanisms, 421
Dysrhythmia Analysis, 423
Abnormal Rates of Sinus Rhythm, 424
Abnormal Site of Impulse Initiation, 425
Conduction Pathway Disturbances, 428
Treatment, 431

20 Shock, 434
Benjamin J. Miller
Pathogenesis of Shock, 434
Impaired Tissue Oxygenation, 435
Compensatory Mechanisms and Stages of Shock, 437
Types of Shock, 439
Cardiogenic Shock, 439
Obstructive Shock, 441
Hypovolemic Shock, 442
Distributive Shock, 443
Assessment and Hemodynamic Monitoring, 447
Cardiac Output, 447
Arterial Oxygen Content, 447
Distribution of Blood Flow, 448
Hemodynamic Monitoring, 448
Complications of Shock, 449
Acute Respiratory Distress Syndrome, 449
Disseminated Intravascular Coagulation, 449
Acute Renal Failure, 449
Multiple Organ Dysfunction Syndrome, 449

UNIT VI Respiratory Function

21 Respiratory Function and Alterations in Gas Exchange, 451
Lorna L. Schumann
Functional Anatomy, 452
Development of the Pulmonary System, 452
Upper Airway Structures, 452
Lower Airway Structures, 453
Pulmonary Circulation, 457
Age-Related Variations, 457
Ventilation, 460
Lung Volumes and Capacities, 460
Dead Space, 460
Minute Ventilation, 460
Alveolar Ventilation/Oxygenation, 461
Mechanics of Breathing, 461
Airway Resistance, 461
Lung Compliance, 462
Distribution of Ventilation, 462
Neurologic Control of Ventilation, 462
Pulmonary Blood Flow, 465
Pulmonary Vasculature, 465
Distribution of Blood Flow, 465
Ventilation–Perfusion Ratios, 465
Hypoxic Vasoconstriction, 466
Diffusion and Transport of Respiratory Gases, 466
Barriers to Diffusion, 466
Oxygen Transport, 467
Carbon Dioxide Transport, 467
Alterations in Pulmonary Function, 468
Hypoventilation and Hyperventilation, 468
Hypoxemia and Hypoxia, 468
Acute Respiratory Failure, 469
Diagnostic Tests, 471
Pulmonary Function Testing, 471
Bronchial Provocation Tests, 472
Alterations in Pulmonary Vasculature, 472
Pulmonary Hypertension, 472
Pulmonary Venous Thromboembolism, 473
Pulmonary Malignancies, 475
Etiology, 475
Pathogenesis, 475
Clinical Manifestations, 476
Diagnosis, 476
Treatment, 476

22 Obstructive Pulmonary Disorders, 478
Benjamin J. Miller and Lorna L. Schumann
Obstruction From Conditions in the Wall of the Lumen, 479
Asthma, 479
Acute Bronchitis, 483
Chronic Bronchitis, 485
Obstruction Related to Loss of Lung Parenchyma, 488
Emphysema, 488
Obstruction of the Airway Lumen, 491
Bronchiectasis, 491
Bronchiolitis, 493
Cystic Fibrosis, 494
Acute Tracheobronchial Obstruction, 495
Epiglottitis, 496
Croup Syndrome, 496

23 Restrictive Pulmonary Disorders, 499
Lorna L. Schumann and Benjamin J. Miller
Lung Parenchyma Disorders, 500
Fibrotic Interstitial Lung Diseases, 500
Atelectatic Disorders, 504
Pleural Space Disorders, 508

Neuromuscular, Chest Wall, and Obesity Disorders, 511
Neuromuscular Disorders, 511
Chest Wall Deformities, 511
Disorders of Obesity, 513
Infection or Inflammation of the Lung, 514

UNIT VII Fluid, Electrolyte, and Acid-Base Homeostasis

24 Fluid and Electrolyte Homeostasis and Imbalances, 521
Linda Felver
Body Fluid Homeostasis, 522
Fluid Intake and Absorption, 522
Fluid Distribution, 523
Fluid Excretion, 524
Fluid Loss Through Abnormal Routes, 524
Fluid Imbalances, 525
Extracellular Fluid Volume, 525
Body Fluid Concentration, 526
Both Volume and Concentration, 528
Interstitial Fluid Volume, 528
Principles of Electrolyte Homeostasis, 529
Electrolyte Intake and Absorption, 529
Electrolyte Distribution, 531
Electrolyte Excretion, 531
Electrolyte Loss Through Abnormal Routes, 531
Electrolyte Imbalances, 531
Plasma Potassium, 531
Plasma Calcium, 533
Plasma Magnesium, 534
Plasma Phosphate, 535

25 Acid–Base Homeostasis and Imbalances, 541
Linda Felver
Acid–Base Homeostasis, 541
Buffers, 542
Respiratory Contribution, 542
Renal Contribution, 543
Acid–Base Imbalances, 545
Mixed Acid–Base Imbalances, 549

UNIT VIII Renal and Bladder Function

26 Renal Function, 551
Jacquelyn L. Banasik
Renal Anatomy, 552
Renal Parenchyma, 552
Renal Lymphatics and Innervation, 552
Renal Blood Supply, 553
Overview of Nephron Structure and Function, 554
Glomerulus, 555
Proximal Convoluted Tubule, 557
Loop of Henle, 558
Distal Convoluted Tubule, 559
Collecting Duct, 559
Regulation of Glomerular Filtration, 560
Physics of Filtration, 560
Factors Affecting Filtration Pressure, 562
Tubuloglomerular Feedback, 563
Effects of Glucose and Amino Acids, 563
Role of Mesangial Cells, 564
Transport Across Renal Tubules, 565
Reabsorption of Glucose, 565
Regulation of Acid–Base Balance, 565
Secretion of Potassium, 566
Regulation of Blood Volume and Osmolality, 567
Antidiuretic Hormone, 567
Aldosterone, Angiotensin II, Natriuretic Peptides, Urodilatin, Uroguanylin, and Guanylin, 568
Diuretic Agents, 569
Endocrine Functions, 570
Erythropoietin, 570
Vitamin D, 570
Age-Related Changes in Renal Function, 570
Infant, 570
Adult and Elderly, 570
Tests of Renal Structure and Function, 570
Urine and Blood Studies, 570
Diagnostic Tests, 572

27 Intrarenal Disorders, 575
Jacquelyn L. Banasik and Roberta J. Emerson
Common Manifestations of Kidney Disease, 575
Pain, 575
Abnormal Urinalysis Findings, 576
Other Diagnostic Tests, 576
Congenital Abnormalities, 577
Renal Agenesis and Hypoplasia, 577
Cystic Kidney Diseases, 578
Neoplasms, 579
Infection, 581
Obstruction, 583
Glomerular Disorders (Glomerulopathies), 586
Glomerulonephritis, 587
Nephrotic Syndrome, 589

28 Acute Kidney Injury and Chronic Kidney Disease, 593
Cheryl Rockwell and Robin Y. Beeman
Acute Kidney Injury, 593
Etiology and Pathophysiology, 594
Clinical Presentation of Acute Kidney Injury, 596
Chronic Kidney Disease, 601
Risk Factors, 601
Pathophysiology of Progression of Chronic Kidney Disease, 602
Stages of Chronic Kidney Disease, 602
Complications of Chronic Kidney Disease, 602
Clinical Management, 604

29 Disorders of the Lower Urinary Tract, 609
Cheryl L. Brandt
Lower Urinary Tract, 609
Functional Anatomy, 609
Physiology of Micturition, 610
Diagnostic Tests, 611
Lower Urinary Tract Symptoms and Syndromes, 611
Neurogenic Bladder, 614
Congenital Disorders, 615
Neoplasms, 617
Inflammation and Infection, 619
Obstruction, 622
Lower Urinary Tract Calculi, 622

UNIT IX Genital and Reproductive Function

30 Male Genital and Reproductive Function, 626
Marvin Van Every
- **Anatomy, 626**
  - *Upper Genitourinary Tract, 626*
  - *Lower Genitourinary Tract, 627*
  - *Auxiliary Genital Glands, 628*
  - *External Genitalia, 629*
- **Embryology, 632**
  - *Nephric System, 632*
  - *Vesicourethral Unit, 633*
  - *Gonads, 633*
  - *Genital Duct System, 633*
  - *External Genitalia, 633*
- **Male Reproductive Physiology, 633**
  - *Hypothalamic-Pituitary-Testicular Axis, 633*
  - *Spermatogenesis, 637*
  - *Anatomy of Spermatozoa, 637*
  - *Transport of Spermatozoa, 637*

31 Alterations in Male Genital and Reproductive Function, 641
Marvin Van Every
- **Disorders of the Penis and Male Urethra, 641**
  - *Congenital Anomalies, 641*
  - *Acquired Disorders, 643*
  - *Infectious Disorders, 645*
  - *Neoplastic Disorders, 647*
- **Disorders of the Scrotum and Testes, 647**
  - *Congenital Disorders, 647*
  - *Acquired Disorders, 648*
  - *Infectious Disorders, 650*
  - *Neoplastic Disorders, 650*
- **Disorders of the Prostate, 651**

32 Female Genital and Reproductive Function, 656
Rosemary A. Jadack
- **Reproductive Structures, 656**
  - *Organization of the Female Reproductive Organs, 656*
- **Menstrual Cycle, 659**
- **Breast, 662**
  - *Structure of the Breast, 662*
  - *Breast Development, 662*
  - *Lactation, 663*
- **Pregnancy, 663**
  - *Early Human Development, 663*
  - *Implantation, 663*
  - *Fetal Membranes and Placenta, 663*
  - *Development of the Human Embryo and Fetus, 664*
  - *Parturition, 665*
  - *Response of the Mother's Body to Pregnancy, 667*
- **Menopause, 668**

33 Alterations in Female Genital and Reproductive Function, 671
Rosemary A. Jadack
- **Menstrual Disorders, 672**
- **Alterations in Uterine Position and Pelvic Support, 674**
- **Inflammation and Infection of the Female Reproductive Tract, 676**
- **Benign Growths and Aberrant Tissue of the Female Reproductive Tract, 678**
- **Cancer of the Female Genital Structures, 679**
- **Disorders of Pregnancy, 681**
- **Disorders of the Breast, 682**
  - *Reactive-Inflammatory Breast Disorders, 682*
  - *Benign Breast Disorders, 683*
  - *Malignant Disorder of the Breast, 684*

34 Sexually Transmitted Infections, 689
Rosemary A. Jadack
- **Urethritis, Cervicitis, Salpingitis, and Pelvic Inflammatory Disease, 690**
- **Diseases With Systemic Involvement, 691**
- **Diseases With Localized Lesions, 694**
  - *Ulcerative Lesions, 694*
  - *Nonulcerative Lesions, 694*
- **Enteric Infections, 695**

UNIT X Gastrointestinal Function

35 Gastrointestinal Function, 697
Jeffrey S. Sartin
- **Structure and Organization of the Gastrointestinal Tract, 698**
  - *Embryology, 698*
  - *Functional Anatomy, 699*
- **Gastrointestinal Motility, 704**
  - *Characteristics of the Intestinal Wall, 704*
  - *Neural Control, 704*
  - *Hormonal Control, 706*
  - *Movement in the Gastrointestinal Tract, 706*
  - *Movement of Nutrients, 707*
- **Secretory Function, 712**
  - *Secretion of Gastrointestinal Juices, 712*
  - *Gastrointestinal Hormones, 712*
- **Digestion and Absorption, 712**
  - *Digestion of Carbohydrates, 713*
  - *Digestion of Lipids, 713*
  - *Digestion of Proteins, 714*
  - *Absorption, 715*
- **Gastrointestinal Function Across the Life Span, 717**
  - *Maturation, 717*
  - *Age-Related Changes, 717*

36 Gastrointestinal Disorders, 720
Jeffrey S. Sartin
- **Manifestations of Gastrointestinal Tract Disorders, 721**
  - *Dysphagia, 721*
  - *Esophageal Pain, 721*
  - *Abdominal Pain, 721*
  - *Vomiting, 723*
  - *Intestinal Gas, 723*
  - *Alterations in Bowel Patterns, 723*
- *DISORDERS OF THE MOUTH AND ESOPHAGUS, 724*
- **Oral Infections, 724**
- **Esophageal Disorders, 724**
- *ALTERATIONS IN THE INTEGRITY OF THE GASTROINTESTINAL TRACT WALL, 726*
- **Inflammation of the Stomach and Intestines, 726**

Inflammatory Bowel Disease, 728
Enterocolitis, 731
ALTERATIONS IN MOTILITY OF THE GASTROINTESTINAL TRACT, 733
Motility Disorders, 733
Disorders of Malabsorption, 735
Mucosal Disorders, 735
Malabsorption Disorders After Surgical Intervention, 735
NEOPLASMS OF THE GASTROINTESTINAL TRACT, 737
Esophageal, Gastric, and Small Intestinal Cancers, 737
Colonic Polyps and Colon Cancer, 738
Psychosocial Aspects of Gastrointestinal Disorders, 739
Stress of Lifestyle Changes, 739

37 Alterations in Function of the Gallbladder and Exocrine Pancreas, 742
Jeffrey S. Sartin
Structure and Function of the Pancreaticobiliary System, 742
Embryology of the Pancreaticobiliary System, 743
Physiology of Bile, 743
Functional Anatomy of the Pancreas, 744
Disorders of the Gallbladder, 744
Pathophysiology of Cholesterol Gallstone Formation, 744
Cholelithiasis and Cholecystitis, 745
Disorders of the Pancreas, 748
Pancreatitis, 748

38 Liver Diseases, 754
Jeffrey S. Sartin
Structure and Function of the Liver, 755
General Manifestations of Liver Disease, 755
Hepatocellular Failure, 755
Portal Hypertension, 759
Portal Systemic Encephalopathy, 763
Complications of Advanced Liver Disease, 764
Disorders of the Liver, 767
Hepatitis, 767
Chronic Hepatitis, 772
Cirrhosis, 773
Alcoholic Liver Disease, 773
Toxic Liver Disorders, 774
Metal Storage Diseases, 774
Toxic Metabolic Agents, 775
Other Structural Liver Conditions, 775
Transplantation, 776
Evaluation of the Transplantation Patient, 777
Posttransplantation Management, 777
Age-Related Liver Disorders, 778
Liver Diseases and Pediatric Considerations, 778
Abnormal Bilirubin Metabolism in the Neonatal Period, 778
Infectious and Acquired Hepatitides in Children, 778
Congenital Liver Disease, 779
Liver Diseases and Geriatric Considerations, 780

UNIT XI Endocrine Function, Metabolism, and Nutrition

39 Endocrine Physiology and Mechanisms of Hypothalamic-Pituitary Regulation, 783
Jacquelyn L. Banasik
Hormone Structure and Action, 784
Chemical Structure of Hormone Classes, 784
Mechanisms of Hormone Action, 784
Hormone Regulation, 786
Hormone Synthesis, Secretion, and Metabolism, 786
Regulation of Receptor Responses, 787
Hypothalamic-Pituitary Endocrine System, 788
Hormones of the Posterior Pituitary Gland, 788
Hormones of the Hypothalamus and Anterior Pituitary Gland, 790
Thyroid Hormones, 793
Thyroid Hormone Synthesis and Secretion, 793
Thyroid Action on Target Cells, 793
Steroid Hormones, 794
Steroid Hormone Synthesis and Secretion, 794
Steroid Action on Target Cells, 795
Categories of Endocrine Disease, 797
Hyposecretion, 797
Hypersecretion, 797
Hyporesponsiveness, 797

40 Disorders of Endocrine Function, 799
Jacquelyn L. Banasik
Basic Concepts of Endocrine Disorders, 799
Etiology of Endocrine Disorders, 799
Classification of Endocrine Disorders, 800
Growth Hormone Disorders, 801
Thyroid Hormone Disorders, 803
Adrenocortical Hormone Disorders, 806
Adrenal Medulla Disorder, 810
Parathyroid Gland Disorders, 811
Regulation and Actions of Parathyroid Hormone, 811
Antidiuretic Hormone Disorders, 812

41 Diabetes Mellitus, 815
Benjamin J. Miller
Regulation of Glucose Metabolism, 816
Hormonal Regulation, 816
Neural Regulation, 817
Exercise, 818
Stress, 818
Glucose Intolerance Disorders, 820
Classification of Glucose Intolerance Disorders, 820
Prediabetes, 820
Diabetes Mellitus, 821
Screening for Diabetes, 824
Clinical Manifestations and Complications, 825
Acute Hyperglycemia, 825
Diabetic Ketoacidosis, 825
Nonketotic Hyperglycemic Hyperosmolar Syndrome, 826
Chronic Hyperglycemia, 826
Vascular Complications, 826
Neuropathic Complications, 827
Complications in Pregnancy, 827

Treatment and Education, 827
Nutrition, 828
Obesity and Eating Disorders, 828
Exercise, 829
Pharmacologic Agents, 829
Stress Management, 831
Assessment of Efficacy, 831
Pediatric Considerations, 833
Goals of Therapy, 834
Acute Complications, 834
Chronic Complications, 834
Treatment, 834
Geriatric Considerations, 834
Goals of Therapy, 835
Acute Complications, 835
Chronic Complications, 835
Treatment, 835

42 Nutritional and Metabolic Disorders, 838
Brent A. Banasik
Metabolic Processes, 838
Anabolism and Catabolism, 839
Metabolic Rate, 839
Nutrient Metabolism, 839
Carbohydrates, 839
Lipids, 840
Proteins, 841
Regulation of Appetite and Nutrient Metabolism, 841
Role of Genetics, Epigenetics, and Environment, 841
Hormonal Regulation of Nutrient Intake and Appetite, 842
Hormonal Regulation of Nutrient Storage, Distribution, and Metabolism, 843
Obesity and Metabolic Syndrome, 844
Obesity, 844
Metabolic Syndrome, 844
Metabolic Responses to Starvation and Physiologic Stress, 844
Starvation and Protein-Energy Malnutrition, 845
Physiologic Stress, 845
Nutritional Considerations for Aging and Altered Health States, 846
Aging, 846
Infection, Sepsis, and Fever, 846
Surgery, 848
Trauma, 848
Burns, 848
Cancer, 848
Immobility, 848

UNIT XII Neural Function

43 Structure and Function of the Nervous System, 850
Jacquelyn L. Banasik
STRUCTURAL ORGANIZATION, 851
Central Nervous System, 851
Support and Protection of the Central Nervous System, 851
The Brain, 853
The Spinal Cord, 861
Peripheral Nervous System, 863
Cranial Nerves, 863
Spinal Nerves, 864
Autonomic Nervous System, 867
NEURONAL STRUCTURE AND FUNCTION, 868
Neurons and Supportive Cells, 868
Neurons, 868
Glia, 868
Neuronal Communication, 872
Membrane Potentials, 872
Synaptic Transmission, 874
Neurotransmitters, 875
Neuronal Circuits, 880
Neural Development, Aging, and Injury, 880
Development, 880
Aging, 881
Injury, 881
SENSORY FUNCTION, 882
Sensory Receptors, 882
Sensory Pathways, 883
Somatosensory Cortex, 883
MOTOR FUNCTION, 883
Motor Neurons, 884
Spinal Reflexes, 884
Central Control of Motor Function, 885
CONSCIOUSNESS, MEMORY, AND SLEEP, 886
Consciousness and Memory, 886
Sleep, 888

44 Acute Disorders of Brain Function, 891
Joni D. Marsh and Jacquelyn L. Banasik
Mechanisms of Brain Injury, 892
Ischemia and Hypoxia, 892
Increased Intracranial Pressure, 895
Manifestations of Brain Injury, 900
Level of Consciousness, 900
Glasgow Coma Scale, 900
Cranial Nerve Reflexes, 901
TRAUMATIC BRAIN INJURY, 902
Epidemiology, 902
Types of Traumatic Brain Injury, 903
Primary Injury, 903
Intracranial Hematomas, 903
Secondary Injury, 905
Treatment, 905
CEREBROVASCULAR DISEASE AND STROKE, 905
Epidemiology, 906
Ischemic Stroke, 906
Hemorrhagic Stroke, 906
Treatment, 906
Stroke Sequelae, 908
Motor and Sensory Deficits, 908
Language Deficits, 908
Cognitive Deficits, 908
CREBRAL ANEURYSM AND ARTERIOVENOUS MALFORMATION, 909
Cerebral Aneurysm, 909
Etiology, 909
Pathogenesis and Manifestations, 910
Treatment, 910
Arteriovenous Malformation, 911
Etiology, 911

Pathogenesis and Manifestations, 911
Treatment, 911
CENTRAL NERVOUS SYSTEM INFECTIONS, 911
Meningitis, 911
Etiology, 911
Pathogenesis and Clinical Manifestations, 911
Treatment, 911
Encephalitis, 912
Etiology, 912
Pathogenesis and Manifestations, 912
Treatment, 912
Brain Abscess, 912
Etiology, 912
Pathogenesis and Manifestations, 912
Treatment, 913
45 Chronic Disorders of Neurologic Function, 915
Joni D. Marsh
Brain and Cerebellar Disorders, 916
Seizure Disorder, 916
Dementia, 918
Parkinson Disease, 920
Cerebral Palsy, 922
Hydrocephalus, 923
Cerebellar Disorders, 924
Spinal Cord and Peripheral Nerve Disorders, 925
Multiple Sclerosis, 925
Spina Bifida, 927
Amyotrophic Lateral Sclerosis, 928
Spinal Cord Injury, 928
Guillain-Barré Syndrome, 931
Bell Palsy, 932
46 Alterations in Special Sensory Function, 936
Joni D. Marsh
HEARING AND BALANCE, 937
Structure and Function of the Ear, 937
External Ear, 937
Middle Ear, 937
Inner Ear, 937
Balance, 938
General Manifestations of Hearing Impairment, 938
Hearing Impairment Disorders, 939
Conductive Hearing Impairment, 939
Sensorineural Hearing Impairment, 939
Otitis Media, 941
Interventions for Individuals With Hearing Impairment, 942
VISION, 942
Structure of the Eye, 942
Visual Pathways, 943
General Manifestations of Visual Impairment, 944
Disorders of the Eye, 944
Errors of Refraction, 944
Age-Related Disorders, 945
Retinopathy, 947
Glaucoma, 948
Visual Field Deficits, 950
Interventions for Individuals With Vision Impairment, 950
SMELL AND TASTE, 952
47 Pain, 955
Joni D. Marsh
Physiology of Pain, 955
Transduction, 956
Transmission, 956
Perception, 958
Modulation, 958
TYPES OF PAIN, 960
Acute Pain, 961
Headache, 962
Chronic Pain, 962
Fibromyalgia Syndrome, 963
Cancer-Related Pain, 964
Neuropathic Pain, 964
Trigeminal Neuralgia, 965
Diabetic Neuropathy, 965
Postherpetic Neuralgia, 966
Ischemic Pain, 966
Referred Pain, 966
Physiologic Responses to Pain, 967
Pain in the Young and the Elderly, 967
TREATMENT MODALITIES, 967
Pharmacologic and Nonpharmacologic Pain Management, 967
Interrupting Peripheral Transmission of Pain, 968
Modulating Pain Transmission at the Spinal Cord, 968
Altering the Perception and Integration of Pain, 968

UNIT XIII Neuropsychological Function

48 Neurobiology of Psychotic Illnesses, 971
Ann Futterman Collier and Samantha Cody Russell
Schizophrenia, 972
Etiology and Neurobiology, 972
Clinical Manifestations, 976
Pharmacologic Treatment, 976
Nonpharmacologic Treatment, 977
Major Depressive and Persistent Depressive Disorders, 978
Etiology and Neurobiology, 979
Clinical Manifestations, 980
Pharmacologic Treatment, 981
Nonpharmacologic Treatment, 981
Bipolar Disorder, 982
Etiology and Neurobiology, 982
Clinical Manifestations, 983
Pharmacologic Treatment, 983
Nonpharmacologic Treatment, 984
Population Considerations, 984
Women and Mental Illness, 984
Cultural Considerations, 984
Geriatric Considerations, 984
49 Neurobiology of Nonpsychotic Illnesses, 989
Ann Futterman Collier and Samantha Cody Russell
Anxiety Disorders, 989
Panic Disorder, 990
Generalized Anxiety Disorder, 991
Obsessive-Compulsive Disorder, 992
Posttraumatic Stress Disorder, 994
Neurodevelopmental Disorders, 995
Attention-Deficit/Hyperactivity Disorder, 996
Autism Spectrum Disorder, 997

UNIT XIV Musculoskeletal Support and Movement

50 Structure and Function of the Musculoskeletal System, 1001
Carol L. Danning
- **Structure and Function of Bone, 1002**
  - *Composition, 1002*
  - *Functional Properties, 1003*
  - *Response to Injury, Stress, and Aging, 1004*
- **Structure and Function of Joints, 1006**
  - *Synarthroses, 1007*
  - *Diarthroses, 1007*
- **Structure and Function of Articular Cartilage, 1011**
  - *Composition, 1011*
  - *Functional Properties, 1011*
  - *Response to Injury, Stress, and Aging, 1012*
- **Structure and Function of Tendons and Ligaments, 1012**
  - *Composition, 1012*
  - *Functional Properties, 1013*
  - *Response to Injury, Stress, and Aging, 1013*
- **Structure and Function of Skeletal Muscle, 1013**
  - *Composition, 1014*
- **Mechanics of Muscle Contraction, 1015**
  - *Sliding Filament Theory, 1015*
  - *Role of Calcium, 1015*
  - *Electromechanical Coupling, 1015*
  - *Types of Muscle Contraction, 1016*
  - *Mechanical Principles, 1017*
  - *Response to Movement and Exercise, 1017*

51 Alterations in Musculoskeletal Function: Trauma, Infection, and Disease, 1020
Carol L. Danning
- ***SOFT TISSUE INJURIES, 1021***
- **Inert Soft Tissue Injuries, 1021**
  - *Ligament Injuries, 1021*
  - *Joint Capsule Injuries, 1022*
  - *Internal Joint Derangement, 1023*
  - *Injuries to Fasciae and Bursae, 1023*
  - *Injuries to Nerves, Nerve Roots, or Dura Mater, 1023*
- **Contractile Soft Tissue Injuries, 1023**
  - *Injury to Tendons, 1023*
  - *Muscle and Tendon Strains, 1024*
  - *Blunt Trauma, 1024*
  - *Compartment Syndrome, 1024*
  - *Soft Tissue Healing After Trauma, 1024*
- ***BONE INJURIES AND INFECTIONS, 1026***
- **Bone and Joint Trauma, 1026**
  - *Types of Bone, 1026*
  - *Fracture, 1026*
  - *Dislocations and Subluxations, 1030*
- **Infections of the Bone, 1031**
  - *Osteomyelitis, 1031*
  - *Tuberculosis, 1032*
- ***ALTERATIONS IN BONE STRUCTURE AND MASS, 1032***
- **Bone Structure Disorders, 1032**
  - *Scoliosis, 1032*
- **Metabolic Bone Diseases, 1033**
  - *Osteoporosis, 1033*
  - *Rickets and Osteomalacia, 1035*
  - *Paget Disease, 1035*
- **Bone Tumors, 1036**
  - *Benign Tumors, 1036*
  - *Malignant Bone Tumors, 1037*
- ***DISEASES OF SKELETAL MUSCLE, 1038***
- **Idiopathic Inflammatory Myopathy, 1038**
  - *Polymyositis and Dermatomyositis, 1038*
- **Muscular Dystrophy, 1039**
  - *Duchenne Muscular Dystrophy, 1039*
  - *Becker Muscular Dystrophy, 1039*
  - *Facioscapulohumeral Muscular Dystrophy, 1039*
  - *Myotonic Dystrophies, 1039*
- ***OTHER DISORDERS OF MUSCLE, 1039***
- **Myasthenia Gravis, 1039**
  - *Treatment, 1039*
- **Chronic Muscle Pain, 1040**
  - *Fibromyalgia Syndrome, 1040*

52 Alterations in Musculoskeletal Function: Rheumatic Disorders, 1042
Carol L. Danning
- **Local Disorders of Joint Function, 1042**
  - *Osteoarthritis, 1042*
  - *Infectious Arthritis, 1045*
  - *Lyme Disease, 1045*
- **Systemic Disorders of Joint Function, 1046**
  - *Immune-Mediated Disorders, 1046*
  - *Postinfectious Systemic Disorders, 1052*
- **Joint Dysfunction Secondary to Other Diseases, 1053**
  - *Psoriatic Arthritis, 1053*
  - *Enteropathic Arthritis, 1053*
  - *Neuropathic Osteoarthropathy, 1054*
  - *Hemophilic Arthropathy, 1054*
  - *Gout, 1054*
  - *Adult-Onset Still Disease, 1055*
- **Pediatric Joint Disorders, 1056**
  - *Nonarticular Rheumatism, 1056*
  - *Hypermobility of Joints, 1056*
  - *Juvenile Idiopathic Arthritis, 1056*

UNIT XV Integumentary System

53 Alterations in the Integumentary System, 1058
Lee-Ellen C. Copstead, Ruth E. Diestelmeier, and Michael R. Diestelmeier
- **Age-Related Changes, 1059**
  - *Epidermis, 1060*
  - *Dermis and Subcutaneous Tissue, 1060*
  - *Appendages, 1061*
- **Evaluation of the Integumentary System, 1061**
  - *Primary and Secondary Lesions, 1061*
  - *Lesion Descriptors, 1062*
- **Selected Skin Disorders, 1063**
- **Infectious Processes, 1064**
  - *Viral Infections, 1064*
  - *Fungal Infections, 1065*
  - *Bacterial Infections, 1067*
- **Inflammatory Conditions, 1069**
- **Allergic Skin Responses, 1072**
- **Parasitic Infestations, 1075**

Other Disorders of the Dermis, 1076
Scleroderma, 1076
Sunburn and Photosensitivity, 1077
Ulcers, 1078
Altered Cell Growth: Epidermal Proliferation, 1078
Pigmentation Alterations, 1080
Special Characteristics of Dark Skin, 1081
Integumentary Manifestations of Systemic Disease, 1083
Skin, 1083
Hair, 1084
Nails, 1084
Treatment Implications, 1086
Topical Treatment, 1086
Intralesional Injection, 1086
Selection of a Delivery System, 1087
Corticosteroids, 1087
Developmental Considerations, 1087
Infancy, 1087
Childhood Skin Disorders, 1088
Adolescence and Young Adulthood, 1090
Geriatric Considerations, 1090

54 Burn Injuries, 1093
Sarah Ogle and Nirav Patel
Thermal Injury, 1094
Etiology, Incidence, and Mortality, 1094
Risk Factors, 1094
Integument Effects, 1094
Depth Classification, 1096
Extent of Injury, 1097
Severity Classification, 1097
Acute Management, 1097
Assessment, 1099
Burn Shock and Acute Resuscitation, 1099
Organ Dysfunction, 1101
Metabolic Changes, 1102
Cellular Changes, 1102
Immune Response, 1103
Elements of Burn Injury Survival, 1103
Rehabilitation Phase, 1106
Electrical Injury, 1107
Incidence and Mortality, 1107
Pathophysiology, 1107
Management and Complications, 1108
Chemical Injury, 1109
Management and Complications, 1109
Common Agents and Treatment, 1109
Special Populations, 1110
Introduction, 1110

Glossary, 1114
Index, 1143

1

Introduction to Pathophysiology

Lee-Ellen C. Copstead

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- What is pathophysiology?
- How are etiology and pathogenesis used to predict clinical manifestations and response to therapy?
- How are normal and abnormal physiologic parameters defined?
- What general factors affect the expression of disease in a particular person?
- What kinds of information about disease can be gained through understanding concepts of epidemiology?

CHAPTER OUTLINE

Framework for Pathophysiology, 2
Etiology, 2
Pathogenesis, 2
Clinical Manifestations, 3
Stages and Clinical Course, 3
Treatment Implications, 3
Concepts of Normality in Health and Disease, 4
Statistical Normality, 4
Reliability, Validity, and Predictive Value, 5
Individual Factors Influencing Normality, 5
Cultural Considerations, 5
Age Differences, 6
Gender Differences, 6
Situational Differences, 6
Time Variations, 6
Patterns of Disease in Populations, 6
Concepts of Epidemiology, 6
Endemic, Pandemic, and Epidemic Diseases, 7
Aggregate Factors, 7
Levels of Prevention, 9

Pathophysiology derives from the intersection of two older, related disciplines: pathology (from *pathos,* suffering) and physiology (from *physis,* nature). Pathology is the study and diagnosis of *disease* through examination of *organs, tissues, cells,* and *bodily fluids.* Physiology is the study of the mechanical, physical, and biochemical functions of living organisms. Together, as *pathophysiology,* the term refers to the study of abnormalities in physiologic functioning of living beings.

Pathophysiology seeks to reveal physiologic responses of an organism to disruptions in its internal or external environment. Because humans exhibit considerable diversity, healthy structure and function are not precisely the same in any two individuals. However, discovering the common and expected responses to abnormalities in physiologic functioning is useful, and it allows a general prediction of clinical progression, identification of possible causes, and selection of interventions that are most likely to be helpful. Thus pathophysiology is studied in terms of common, or "classic," presentations of disorders.

Historically, descriptions of diseases were based on observations of those individuals who attracted medical attention because they exhibited abnormal signs or complained of symptoms. Over time, cases with similar presentations were noted and treatments that had been successful before were used again. In some cases, similarities among individuals pointed to possible common causes. With the advent of more sophisticated measurements of physiologic and biochemical function, such as blood pressure measurements, blood chemistry values, x-ray images, and DNA analysis, the wide variability in the expression of diseases and disorders in the population became apparent, as did the opportunity to discover diseases at earlier stages, before they were clinically obvious. Screening programs that evaluated large segments of the population

revealed the complexity and diversity of disease expression, even in persons with the same genetic defect. Thus although the study of pathophysiology is necessarily a study of the usual and expected responses of the body to a given disruption, individuals often vary significantly from a classic presentation, making the diagnostic process complex and challenging.

Advances in genomic and epigenomic characterization, innovative technologies, and revolutionary approaches to the analysis of genetic variation and function have made studies and treatments possible that were not even imaginable just a few years ago. As a result, definitions of the living world have been virtually transformed and permeate every branch of biological science. Benefits of this new biology include a deeper understanding of evolution, greater insights into immune mechanisms, and advances against cancer and AIDS.

Genetic manipulation also raises sensitive and complex ethical and moral questions that did not exist half a century ago. Scientists are able to experiment with genetic manifestations and their mechanisms of action, dramatically altering medical practice, especially the management of inherited diseases. New capabilities have led to experimental treatments such as gene therapy–molecular surgery powerful enough to cure and alter the next generation. The study of pathophysiology assumes even greater significance as genetic research shows fresh insights and hopeful new treatments for human diseases.

Pathophysiology examines disturbances of normal mechanical, physical, and biochemical functions, either caused by a disease or resulting from a disease or abnormal syndrome or condition. For example, the study of a toxin released by a bacterium has evolved from the science of infectious diseases, as well as the harmful effects of that toxin on the body, with one possible result being sepsis. Another example is the study of the chemical changes that take place in body tissue as the result of inflammation.

Although individual study of specific diseases undertaken in medical pathology textbooks helps students identify subtle differences between similar diseases, the study of pathophysiology is dynamic and conceptual, seeking to explain processes and relationships common to a number of pathologies. For example, the pathophysiology of inflammation, hypotension, fluid volume deficit, hypoxia, and ischemia is important to the understanding of a large number of different pathologies, but each separate process is not necessarily a specific disease.

Pathophysiology includes four interrelated topics: etiology, pathogenesis, clinical manifestations, and treatment implications—the framework used throughout this textbook. Specific diseases will be used as illustrative examples of conditions in which particular pathophysiologic processes may occur.

FRAMEWORK FOR PATHOPHYSIOLOGY

Etiology

Etiology, in its most general definition, is the study of the causes or reasons for phenomena. A description of etiologic process includes the identification of those causal factors that, acting in concert, provoke a particular disease or injury. When the cause is unknown, a condition is said to be *idiopathic*. If the cause is the result of an unintended or unwanted medical intervention, the resulting condition is said to be *iatrogenic*. Most disorders are multifactorial, having several different etiologic factors that contribute to their development. For example, coronary heart disease is a result of the interaction of genetic predisposition, diet, exposure to cigarette smoke, elevated blood pressure, and perhaps numerous other lifestyle and hormonal factors acting in concert. None of these individual factors can be said to cause the disease. When the link between an etiologic factor and development of a disease is less than certain but the probability is increased when the factor is present, it is termed a *risk factor*. The identification of risk factors is important for disease prevention, and various levels of prevention provide a focus for the epidemiology section at the end of this chapter.

Some diseases are closely linked with etiologic factors, such that they are said to be the causative agents in the disease. For example, microbial pathogens are considered to be causative agents for infectious diseases: Ebola virus disease—a rare and deadly disease—is caused by infection with a virus of the family *Filoviridae,* genus *Ebolavirus* (Fig. 1.1); human immunodeficiency virus causes HIV disease, and *Mycobacterium tuberculosis* causes pulmonary tuberculosis. These diseases do not occur unless the pathogen is present in the body; however, this does not mean that the infection will have the same consequences in each case, because many host factors affect the clinical course.

Even when the link between disease and etiologic agent is strong, only a portion of the population exposed to the factor may develop the disease. For example, in persons who consume large quantities of alcohol and develop liver cirrhosis, it is the alcohol consumption that is considered to be the cause, yet only a portion of persons who drink heavily will develop cirrhosis. Thus categorizing the probable etiologies for diseases is a long, difficult research process and, not surprisingly, the exact causes of most disorders remain incompletely understood. Several classification schemes have been proposed to categorize diseases according to etiology. Box 1.1 summarizes an example of an etiologic classification system. No classification system is truly comprehensive, and some diseases fall into multiple categories. Some diseases may receive different designations in the future, as further research reveals new data.

Pathogenesis

Pathogenesis refers to the development or evolution of a disease, from the initial stimulus to the ultimate expression of the manifestations of the disease. The sequence of physiologic events that occurs in response to an etiologic agent is a dynamic interplay of changes in cell, tissue, organ, and systemic function. As the ways in which intricate intercellular communication networks control physiologic function are discovered,

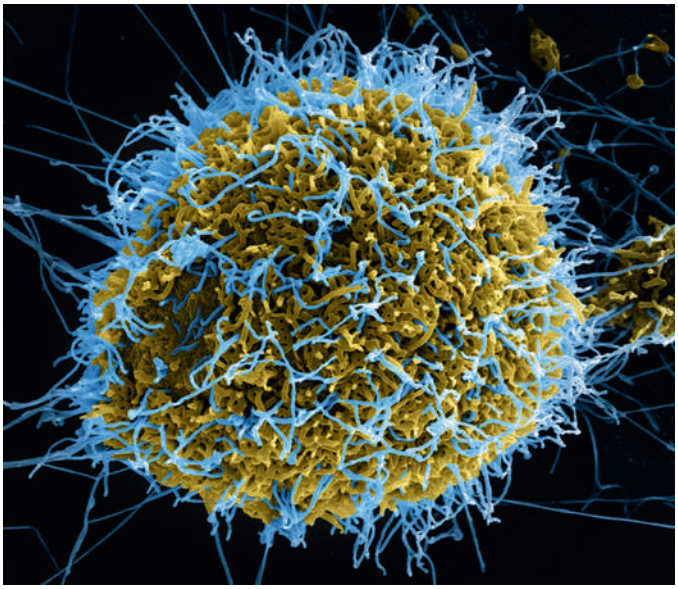


FIG 1.1 Colorized scanning electron micrograph of filamentous Ebola virus particles (blue) budding from a chronically infected VERO E6 cell (yellow-green). (From National Institute of Allergy and Infectious Diseases at http://flickr.com/photos/54591706@N02/14440817981. March 20, 2017.)

BOX 1.1 Etiologic Classification of Diseases

Congenital (inborn) diseases or birth defects
Degenerative diseases
Iatrogenic diseases
Idiopathic diseases
Immunologic diseases
Infectious diseases
Inherited diseases
Metabolic diseases
Neoplastic diseases
Nutritional deficiency diseases
Physical agent–induced diseases
Psychogenic diseases

pathogenesis is being increasingly understood on the cellular level. One of the best examples of this communication network is the immune system and its interactions with essentially every other cell in the body. A disruption in the delicate system of checks and balances between immune tolerance of normal cells and immune surveillance for abnormal cells and foreign antigens is at the root of a large number of degenerative and inflammatory diseases.

Pathologic disruptions in cellular behavior lead, in turn, to changes in organ and system function that may be detected by clinical or laboratory examination. Most pathophysiology texts take a systems approach to presenting information. This approach builds on the way in which students learn anatomy and physiology and has its roots in medical specialization. Usually the clinical examination of a patient is also conceptualized by a systems approach. Although the division into systems is useful for dividing the content into conceptual pieces, it is important to remember that the organism functions as an integrated whole and the intercellular communication networks are not confined within single systems. In summary, pathogenesis is a description of how etiologic factors are thought to alter physiologic function and lead to the development of clinical manifestations that are observed in a particular disorder or disease.

Clinical Manifestations

Manifestations of disease that are observed are termed *signs of disease.* Such objective data may be gathered by clinical examination or by biochemical analysis, diagnostic imaging, and other laboratory tests. The subjective feelings of an abnormality in the body are termed *symptoms.* By definition, symptoms are subjective and can only be reported by the affected individual to an observer. For example, the feeling of nausea is a symptom, whereas vomiting is objectively observed and is a sign. Some signs and symptoms, such as fever and headache, are nonspecific and, although they designate that something is amiss, they do not indicate a specific cause. In this case further examination and, often, laboratory tests are needed to focus on the possible causes of the signs and symptoms. Many diseases and disorders are characterized by a particular constellation of signs and symptoms, the knowledge of which is essential for accurate detection and diagnosis. When the etiology of a particular set of signs and symptoms has not yet been determined, the disorder may be termed a *syndrome.* For example, AIDS was originally detected as a set of signs and symptoms related to a deficiency of helper T cells of unknown cause, now known to be a late stage of HIV infection.

The clinical manifestations of some diseases may change significantly over time, resulting in a completely different clinical presentation at different stages. Knowledge of the possible stages of a disease is helpful in making an appropriate diagnosis and anticipating the clinical course.

Stages and Clinical Course

Early in the development of a disease, the etiologic agent or agents may provoke a number of changes in biological processes that can be detected by laboratory analysis, although no recognition of these changes by the patient has occurred. The interval between exposure of a tissue to an injurious agent and the first appearance of signs and symptoms may be called a *latent period* or, in the case of infectious diseases, an *incubation period*. The prodromal period, or prodrome, refers to the appearance of the first signs and symptoms indicating the onset of a disease. Prodromal symptoms often are nonspecific, such as headache, malaise, anorexia, and nausea. During the stage of manifest illness, or the acute phase, the disease reaches its full intensity, and signs and symptoms attain their greatest severity. Sometimes during the course of a disease, the signs and symptoms may become mild or even disappear for a time. This interval may be called a *silent period* or *latent period.* For example, in the total-body irradiation syndrome, a latent period may occur between the prodrome and the stage of manifest illness. Another example is syphilis, which may have two latent periods: one occurring between the primary and secondary clinical stages and another occurring between the secondary and tertiary stages.

A number of diseases have a subclinical stage, during which the patient functions normally, although the disease processes are well established. It is important to understand that the structure and function of many organs provide a large reserve or safety margin, so that functional impairment may become evident only when organ damage has become advanced. For example, chronic renal disease can completely destroy one kidney and partly destroy the other before any symptoms related to a decrease in renal function are perceived.

The clinical course of a disease is often classified as acute or chronic. An acute condition has relatively severe manifestations but runs a short course measured in hours, days, or a few weeks. A chronic condition lasts for months to years. Sometimes chronic disease processes begin with an acute phase and become prolonged when the body's defenses are insufficient to overcome the causative agent or stressor. In other cases, chronic conditions develop insidiously and never have an acute phase.

Some diseases (e.g., some types of autoimmune diseases) follow a course of alternating exacerbations and remissions. An exacerbation is a relatively sudden increase in the severity of a disease or any of its signs and symptoms. A remission is an abatement or decline in severity of the signs and symptoms of a disease. If a remission is permanent (sometimes defined as longer than 5 years), the person is said to be cured.

Convalescence is the stage of recovery after a disease, injury, or surgical operation. Occasionally a disease produces a subsequent pathologic condition called a *sequela* (plural: sequelae). For example, the sequela of an inflammatory process might be scarring. The sequelae of acute rheumatic inflammation of the heart might be scarring and deformation of cardiac valves. In contrast, a complication of a disease is a new or separate process that may arise secondarily because of some change produced by the original problem. For example, bacterial pneumonia may be a complication of viral infection of the respiratory tract.

Treatment Implications

An understanding of the etiology, pathogenesis, and clinical consequences of a particular disorder may suggest, or "imply," that certain treatments could be helpful. For example, understanding that a person with septic shock has excessive dilation of blood vessels that contributes to

hypotension implies that fluid administration would likely be helpful. In contrast, most patients with cardiogenic shock have fluid overload, and hypotension in this case is unlikely to improve with fluid administration. Care must be taken not to rely on theoretical implications when evidence-based treatment recommendations are available. When subjected to evaluation by rigorous randomized clinical trials, many treatments that seem as though they should help based on pathophysiology fail to pass the test of application.

The treatment implications discussed in pathophysiology texts usually are general statements rather than specific prescriptions. For example, the pathophysiology of heart failure is characterized by fluid overload, which implies that diuretic therapy would be useful; however, the exact selection of a drug and the dosing schedule would depend on a number of factors particular to the individual patient. Specific treatment recommendations are beyond the scope of a pathophysiology text and can be found in pharmacology and clinical practice textbooks.

KEY POINTS

- Pathophysiology includes four interrelated topics: etiology, pathogenesis, clinical manifestations, and treatment implications.
- Etiology refers to study of the proposed cause or causes of a particular disease process. Etiology is a complex notion because most diseases are multifactorial, resulting from interplay between genetic constitution and environmental influences.
- Pathogenesis refers to the proposed mechanisms whereby an etiologic stimulus leads to typically observed clinical manifestations. Pathogenesis describes the direct effects of the initiating event, as well as the usual physiologic responses and compensatory mechanisms.
- Clinical manifestations describe the signs and symptoms that typically accompany a particular pathophysiologic process. Manifestations may vary depending on the stage of the disorder, individual variation, and acuity or chronicity.
- An understanding of the etiology, pathogenesis, and clinical consequences of a particular disorder implies that certain treatments may be helpful.

CONCEPTS OF NORMALITY IN HEALTH AND DISEASE

The ability to measure numerous structural, physiologic, biochemical, and genetic parameters in an individual allows the evaluation of information that is helpful in the diagnosis and monitoring of clinical diseases. Many of these same measures are commonly used to screen for disease or to evaluate the risks of a disease occurring in the future. To determine whether a certain finding is indicative of disease or "abnormal," it must be compared with what is "normal." The obviousness of this statement belies the difficulty in determining what is normal and the degree of deviation from normal that would be considered abnormal. Many clinical parameters are evaluated by direct observation by the examiner. Skin color and warmth, quality of pulses, briskness of pupil reactions to light, mental acuity, muscle strength, joint mobility, heart sounds, lung sounds, bowel sounds, balance, psychological affect, and level of consciousness are but a few examples of assessments that are subjectively interpreted based on the examiner's observations. Deciding whether a clinical finding is normal, a normal variation, or an abnormality indicative of a disorder is essential. Reliability of data obtained from observation is dependent on the examiner's skill and experience. Often the clinical examination is not sufficient to determine definitively the underlying pathophysiologic processes, and diagnostic testing is undertaken to provide more information.

Statistical Normality

Some of the variables that are measured to diagnose disease are relatively easy to declare as normal or abnormal because they occur in only two states; for example, a bone is either broken or not broken on x-ray examination. However, most diagnostic variables occur in the population according to a "bell curve," or normal distribution. This means that a large enough sample taken from the population should give a good estimate of the range of values in the population. Statistics are often used to determine the standard deviation of the variable in question, and then a normal range is suggested as the mean ±2 standard deviations. This means that 95% of the values in the population are expected to fall in the normal range and 5% will be either higher or lower (Fig. 1.2). The "population" chosen to serve as the normal reference population

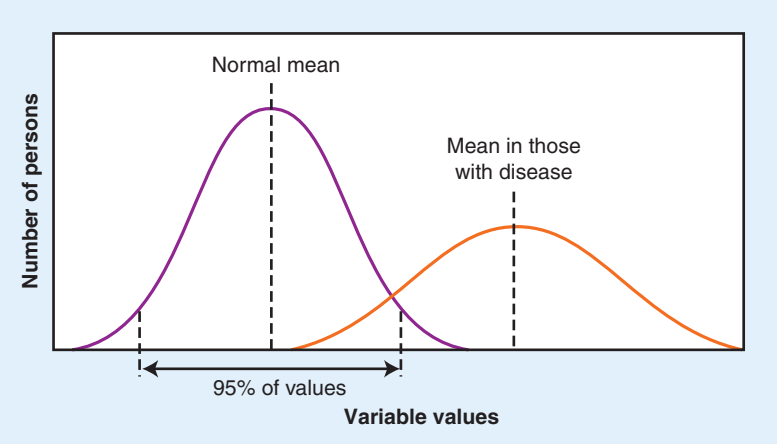


FIG 1.2 Representative example of a normal bell curve for a physiologic variable. Many physiologic variables are normally distributed within the population, so the mean ±2 standard deviations include 95% of the normal values in the sample. Approximately 2.5% of values will be above the normal range and 2.5% will be below it. There may be overlap between the values in a normal sample and those in the population with a disease, making interpretation difficult in some cases.

must be carefully selected to represent the individual to be tested for disease, because many variables are influenced by age and gender.

For example, bone density can be measured in the population by radiologic imaging, and then a mean and standard deviation can be calculated. Women typically have lower bone density than men, and older women have lower bone density than younger women. If an elderly woman's bone density is compared with women of her own age group, it may fall within the normal range, but compared with a group of younger women, it is more than 2 standard deviations below the mean. Which is the right comparison group to use to determine whether she has osteoporosis? There is controversy on this point because, in this situation, it is difficult to determine the difference between disease and the effects of normal aging.

Often, when assessing a person's health status, a *change* in some value or factor is more significant than the actual value of the factor. A blood pressure of 90/70 mm Hg may not be significant if that is the usual value. However, if a person usually has a blood pressure of 120/80 mm Hg, a reading of 90/70 mm Hg could indicate a significant change. Individuals are typically evaluated more than once—generally two or three times—to establish deviation from their usual value.

Reliability, Validity, and Predictive Value

The accurate determination of whether a specific condition is present or absent depends on the quality and adequacy of the data collected, as well as the skill of interpretation. Decisions about the data needed are based on the initial clinical presentation and a working knowledge of pathophysiology, which guide hypothesis generation about probable etiologies. During the clinical examination, data are analyzed, and a number of likely explanations for the clinical presentation may emerge. These possible explanations are "probabilities" based on knowledge and past experience with similar cases. The purpose of further data collection, particularly laboratory and diagnostic testing, is to refine the initial probability estimates and identify the most likely diagnosis. The success of this approach depends on the selection of appropriate tests based on the pretest probabilities, as well as on the validity, reliability, and predictive value of the tests.

Validity, or accuracy, is the degree to which a measurement reflects the true value of the object it is intended to measure. For example, a pulse oximeter is designed to measure arterial oxygen saturation, and the closeness of the reading to a direct measurement of oxygen saturation in an arterial blood sample reflects its accuracy. Reliability, or precision, is the ability of a test to give the same result in repeated measurements. An instrument or laboratory test can be reliable, yet inaccurate. Repeated measurements with the pulse oximeter could give the same result each time, but if those values are significantly different from the "gold standard" of an arterial blood sample, the oximeter data would have poor validity.

Some measurements vary according to the reagents and laboratory methods used. For example, prothrombin time (PT) is sensitive to the reagent used. In one method of determining PT, the reagent—a substance composed of thromboplastin and calcium—is added to decalcified plasma to create a reaction resulting in clot formation. The PT is then determined by measuring the length of time it takes for clotting to occur after this reagent is added and compared with the normative average. Portions of the same blood sample sent to several different laboratories could return significantly different PT results. In fact, this is such a problem that laboratories now use a correction procedure to normalize the PT values across labs. The corrected PT value is reported as the international normalized ratio (INR), which has higher reliability than the PT.

The predictive value of a test is the extent to which the test can differentiate between the presence and absence of a condition in an individual. The positive predictive value is an estimate of the probability that disease is present if the test is positive. The negative predictive value is an estimate of the probability that disease is absent if the test is negative. The predictive value of a test depends in part on the sensitivity and specificity of the test and in part on the probability of the disease being present before the test is obtained. Most tests are not perfectly specific and sensitive so the results must be interpreted probabilistically in view of the diagnostic hypotheses being tested.

Sensitivity and specificity are measures of how well a given test can discriminate between persons with and without a given condition. Sensitivity is the probability that the test will be positive when applied to a person with the condition. For example, if a kit for testing a throat swab for the presence of streptococcal infection has a sensitivity of 80%, then 20% of a group of people with streptococcal throat infection would erroneously test negative for the condition (false-negative rate). Another example is the blood test for HIV antibodies, which has a sensitivity of 99% and would fail to detect the condition in only 1% of a group of individuals who had HIV antibodies in their blood. Specificity is the probability that a test will be negative when applied to a person who does not have a given condition. If the streptococcal throat swab kit has a specificity of 95%, then 5% of those tested who do not actually have the condition would erroneously test positive (false-positive rate). The importance of evaluating the accuracy and precision of data is paramount because inappropriate diagnoses and clinical management could occur if decisions are predicated on invalid or unreliable data.

The positive predictive value of a test is improved when sensitivity and specificity are high and the test is applied to individuals who have a high probability of having the condition being tested. If the likelihood of a condition in the population being tested is low (e.g., a 2% prevalence rate), then a positive result in a test with 99% specificity and 99% sensitivity would only have a 67% positive predictive value. This means that testing low-likelihood or low-risk individuals would produce a high percentage of false-positive results (33% in the preceding example). Therefore deciding who to test for a given condition based on the probability of the condition being present is as important as the sensitivity and specificity of the test. A good working knowledge of pathophysiology is necessary to generate the hypotheses that guide collection of appropriate data and facilitate the diagnostic process.

Individual Factors Influencing Normality

Variations in physiologic processes may be a result of factors other than disease or illness. Age, gender, genetic and ethnic background, geographic area, and time of day may influence various physiologic parameters. Care must be taken to interpret "abnormal" findings with consideration of these possible confounding factors. In addition, the potential for spurious findings always exists. Thus trends and changes in a particular individual are more reliable than single observations. Single measurements, observations, or laboratory results that seem to indicate abnormality must always be judged in the context of the entire health picture of the individual. One slightly elevated blood glucose level does not mean clinical diabetes, a single high blood pressure reading does not denote hypertension, and a temporary feeling of hopelessness does not indicate clinical depression.

Cultural Considerations

Each culture defines health and illness in a manner that reflects its experience. Cultural factors determine which signs, symptoms, or behaviors are perceived as abnormal. An infant from an impoverished culture with endemic chronic diarrhea and a degree of malnutrition would be viewed as abnormal in a progressive culture, such as a well-baby

clinic in Sweden. Given cultural variations that affect definitions of normal and abnormal, the resulting pattern of behaviors or clinical manifestations affects what the culture labels as illness.

Age Differences

Many biological factors vary with age, and the normal value for a person at one age may be abnormal at another. Physiologic changes, such as hair color, skin turgor (tension), and organ size, vary with age. In general, most organs shrink; exceptions are the male prostate and the heart, which enlarge with age. Special sensory changes, such as severely diminished near-sight, high-tone hearing loss, and loss of taste discriminations for sweet and salty, are normal in an elderly adult and abnormal in a middle-aged adult or child. There are fewer sweat glands and less thirst perception in an elderly person than in a young adult or child. Elderly persons have diminished temperature sensations and can therefore sustain burn injuries—from a heating pad or bath water—because they do not perceive heat with the same intensity as do middle-aged adults. A resting heart rate of 120 beats per minute is normal for an infant but not for an adult.

Gender Differences

Some laboratory values, such as levels of sex and growth hormones, show gender differences. The complete blood cell count shows differences by gender in hematocrit, hemoglobin, and red blood cell (RBC) count. For example, the normal range of hemoglobin concentration for adult women is lower than that for adult men—for adult women, the normal hemoglobin range is 12 to 16 g/100 mL of blood, whereas for adult men the normal range is 13 to 18 g/100 mL of blood. There are also gender differences in the erythrocyte sedimentation rate (ESR). Normally, in males, the ESR is less than 13 mm/hr; it is slightly higher in females. There are differences by gender in creatinine values too. For females, the normal serum creatinine level is 0.4 to 1.3 mg/dL; for males, the normal range is 0.6 to 1.5 mg/dL. Research into gender differences also suggests that, on average, males snore more; have longer vocal cords, better daylight vision, and higher metabolic rates; and are more likely to be left-handed than females. Research suggests, too, that females and males may have different communication styles and respond differently to similar conditions.

Situational Differences

In some cases, a deviation from the usual value may occur as an adaptive mechanism, and whether the deviation is considered abnormal depends on the situation. For example, the RBC count increases when a person moves to a high altitude. The increase is a normal adaptive response to the decreased availability of oxygen at a high altitude and is termed *acclimatization*. A similar increase in the RBC count at sea level would be abnormal.

Time Variations

Some factors vary according to the time of day; that is, they exhibit a circadian rhythm or diurnal variation. In interpreting the result of a particular test, it may be necessary to know the time at which the value was determined. For example, body temperature and plasma concentrations of certain hormones (such as growth hormone and cortisol) exhibit diurnal variation. Reflecting fluctuation in plasma levels, the peak rate in urinary excretion for a particular steroid (17-ketosteroid) occurs between 8 AM and 10 AM for persons who customarily rise early in the morning and is about two to three times greater than the lowest rate in the same people, which occurs between midnight and 2 AM, usually during sleep. The urinary excretion of ions (e.g., potassium) also exhibits diurnal variation. Fig. 1.3 illustrates circadian rhythms of several physiologic variables for persons living on a standard day-active schedule.

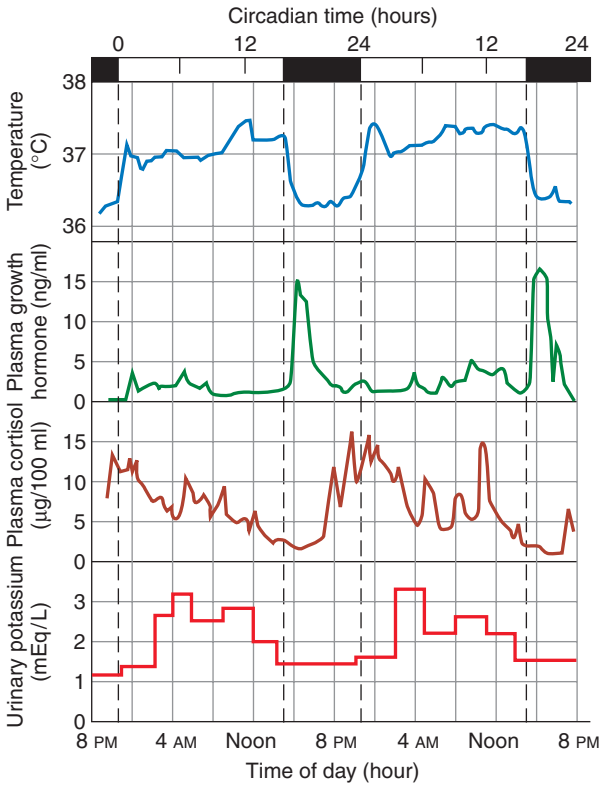


FIG 1.3 Circadian rhythms of several physiologic variables in a human subject depict the effect of light and dark. In an experiment with lights on (*open bars* at top) for 16 hours and off (*black bars* at top) for 8 hours, temperature readings and plasma growth hormone, plasma cortisol, and urinary potassium levels exhibit diurnal variation. (Redrawn from Vander AJ et al: *Human physiology,* ed 7, New York, 1998, McGraw-Hill.)

KEY POINTS

- Determining whether clinical findings are normal, abnormal, or a normal variation is an essential, but often difficult, process in evaluating for the presence or absence of disease.
- Normal ranges for laboratory tests are typically defined as the mean ±2 standard deviations; thus 5% of the normal population may fall outside the normal range despite the absence of disease. Laboratory tests must be evaluated in concert with clinical information.
- The predictive value of a clinical test is the extent to which it can differentiate between the presence and absence of disease in an individual. Tests with high sensitivity and specificity generally have better predictive value.
- Variations in physiologic processes may be a result of factors other than disease or illness. Age, gender, genetic and ethnic background, geographic area, and time of day may influence various physiologic parameters.
- Trends and changes in a particular individual are more reliable than single observations.

PATTERNS OF DISEASE IN POPULATIONS

Concepts of Epidemiology

Differences among *individuals* are, of course, very important in determining the diseases to which they are susceptible and their reactions to the diseases once contracted. But epidemiology, or the study of *patterns of disease* involving aggregates of people (Fig. 1.4), provides yet another important dimension. Information may be gained by examining the occurrence, incidence, prevalence, transmission, and distribution of diseases in large groups of people or populations.

FIG 1.4 A, The aggregate focus in disease: influence of crowds on disease transmission. Crowd gathered at a public market in Russia. **B,** Crowds gathered to purchase goods at a public market in Guangzhou, China. (Photographed by L-E Copstead.)

Endemic, Pandemic, and Epidemic Diseases

A disease that is native to a local region is called an *endemic disease.* If the disease is disseminated to many individuals at the same time, the situation is called an *epidemic. Pandemics* are epidemics that affect large geographic regions, perhaps spreading worldwide. Because of the speed and availability of human travel around the world, pandemics are more common than they once were.

Almost every flu season, a new strain of influenza virus quickly spreads from one continent to another. The 2014–2015 Ebola epidemic in West Africa provides an excellent example of an epidemic that spread worldwide, becoming *pandemic.* The first case was reported in Guinea in March 2014, and the disease spread in the neighboring countries of Liberia and Sierra Leone. Over the span of a year, the Ebola pandemic caused more than 10 times as many cases of Ebola than the combined total of all those reported in previous Ebola outbreaks. As the outbreak became more widespread, travel-associated cases appeared in Nigeria, Mali, Senegal, and countries outside Africa, including the United States.

Aggregate Factors

Principal factors affecting patterns of disease in human populations include the following: (1) age (i.e., time in the life cycle), (2) ethnic group, (3) gender, (4) socioeconomic factors and lifestyle considerations, and (5) geographic location.

Age. In one sense, life is entirely different during the 9 months of gestation. The structures and functions of tissues are different: they are primarily dedicated to differentiation, development, and growth. Certainly the environment is different; the individual is protected from the light of day, provided with predigested food (even preoxygenated blood), suspended in a fluid buffer, and maintained at incubator temperature. This is fortunate because the developing embryo or fetus has relatively few homeostatic mechanisms to protect it from environmental change. (The factors that produce disease in utero are discussed in Chapter 6.) Diseases that arise during the postuterine period of life and affect the neonate include immaturity, respiratory failure, birth injuries, congenital malformations, nutritional problems, metabolic errors, and infections. These conditions are discussed in separate chapters.

Accidents, including poisoning, take their toll in childhood. Infections in children reflect their increased susceptibility to agents of disease. Consideration of other childhood diseases is addressed in each chapter as appropriate and given separate consideration throughout the text. The study of childhood processes and of changes that occur in this period of life is the domain of pediatrics; specific diseases that occur during maturity (ages 15 to 60) are emphasized in this text.

The changes in function that occur during the early years of life are termed *developmental processes.* Those that occur during maturity and postmaturity (age 60 and beyond) are called *aging processes.* The study of aging processes and other changes that occur during this period of life is called *gerontology.* The effects of aging on selected body systems are so important physiologically that they also receive separate consideration throughout the text. The immune, cardiac, respiratory, musculoskeletal, neurologic, special sensory, endocrine, gastrointestinal, and integumentary systems are among those affected by the process of aging.

Ethnic group. It is difficult to differentiate precisely between the effects of ethnicity on patterns of disease and the socioeconomic factors, religious practices, customs, and geographic considerations with which ethnicity is inseparably bound. For example, carcinoma of the penis is virtually unknown among Jews and Muslims who practice circumcision at an early age (avoiding the carcinogenic stimulus that arises from accumulation of smegma about the glans penis).

However, comparisons reveal significant differences in the occurrence of certain disease states in ethnic groups that seem to be more closely related to genetic predisposition than to environmental factors. For example, sickle cell anemia has a much higher rate of occurrence in African populations, whereas pernicious anemia occurs more frequently among Scandinavians and is rare among black populations worldwide.

The study of racial and ethnic group variation in disease states is the domain of medical anthropology. Disease-specific differences that relate to racial or ethnic group differences are a developing research frontier. In clinical practice, recognition of diversity in disease risk by racial or ethnic group is useful in disease diagnosis, prevention, and management. Ethnic group–specific differences, where important, are presented in individual chapters.

Gender. Particular diseases of the genital system obviously show important differences between the sexes; men do not have endometriosis, nor do women have hyperplasia of the prostate, and carcinoma of the breast is more common in women than in men. Pyelonephritis is more common in young women than in men of comparable age (before they develop prostatic hyperplasia) because the external urethral orifice of women is more readily contaminated, and bacteria can more easily travel up a short urethra than a long one. Less obviously related to the reproductive system, the onset of severe atherosclerosis in women is delayed nearly 20 years or more over that in men, presumably because of the protective action of estrogenic hormone.

However, some gender-specific factors defy explanation. For example, systemic lupus erythematosus is much more common in women. Toxic goiter and hypothyroidism are also more common in women. Rheumatoid arthritis is more common in women, but osteoarthritis affects men and women with equal frequency. Thromboangiitis obliterans (a chronic, recurring, inflammatory peripheral vascular disease) occurs more commonly in men. Gender differences in predisposition to cancer and other diseases, where relevant, are presented throughout the text.

Socioeconomic factors and lifestyle considerations. The environment and the political climate of countries determine how people live and the health problems that are likely to ensue. The importance of poverty, malnutrition, overcrowding, and exposure to adverse environmental conditions, such as extremes of temperature, is obvious. Volumes have been written about the effects of socioeconomic status on disease. Sociologists study the influence of these factors. Social class influences education and occupational choices.

Disease is related to occupational exposure to such agents as coal dust, noise, or extreme stress. Lifestyle considerations are closely related to socioeconomic factors. Many people living in the United States, for example, consume too much food, alcohol, and tobacco and do not exercise enough. Childhood obesity is a problem in the United States and is becoming a global problem as well. Arteriosclerosis; cancer; diseases of the kidney, liver, and lungs; and accidents cause most deaths in the United States. By contrast, people living in developing nations suffer and frequently die of undernutrition and infectious diseases.

However, infectious disease is not limited to developing countries. The Centers for Disease Control and Prevention (CDC) estimates that 2 million people annually acquire infections while hospitalized, and 90,000 people die as a result of those infections. More than 70% of hospital-acquired infections have become resistant to at least one of the drugs commonly used to manage them, largely attributable to the overprescribing of antibiotics. *Staphylococcus,* the leading cause of hospital infections, is now resistant to 95% of first-choice antibiotics and 30% of second-choice antibiotics. Poor hand hygiene is considered the leading source for infections acquired during hospitalizations. Unfortunately, efforts to convince health care personnel to reduce transmission of infection through practices as simple as more frequent and thorough handwashing have met with only modest success.

The incidence of many parasitic diseases is closely tied to socioeconomic factors and lifestyle considerations. Worm infections, for example, are related to the use of human feces as fertilizer. In some areas, such as parts of Asia, Africa, and tropical America, the frequency of schistosomiasis (a parasitic infestation by blood flukes) is directly related to the widespread use of irrigation ditches that harbor the intermediate snail host. There is adequate opportunity for transmission of schistosomiasis because children often play in these ditches, and families wash their clothes in ditch water (Fig. 1.5).

Trichinosis, a disease caused by the ingestion of *Trichinella spiralis,* occurs almost entirely from eating inadequately cooked, infected pork. People who are fond of raw meat and inadequately cooked sausage are at highest risk.

Education is often very effective in changing lifestyle patterns that contribute to disease. In Tokyo, for example, mass public education about minimizing the use of sodium—a common ingredient in most traditional Japanese cooking—has been effective in changing dietary practices.

Examples of educational efforts directed at lifestyle modification in the United States are numerous. Antidrug, antismoking, and profitness messages fill the media and are prevalent on the Internet. Choosing

FIG 1.5 Risk factors for schistosomiasis include the widespread use of irrigation ditches that harbor the intermediate snail host. (Photographed in China by L-E Copstead.)

healthy alternatives over unhealthy ones is made easier through positive peer pressure and support groups.

Geographic location. Patterns of disease vary greatly by geographic location. Certainly there is considerable overlap with ethnicity, socioeconomic factors, and lifestyle choices, but physical environment also is an important aspect. Obviously, frostbite in Antarctica and dehydration in the Sahara are examples of disorders that are more prevalent in specific geographic settings. However, important patterns of disease occur within individual countries. For example, the incidence and type of malnutrition vary tremendously by geographic region.

Many diseases have a geographic pattern for reasons that are clear. For example, malaria, an acute and sometimes chronic infectious disease resulting from the presence of protozoan parasites within red blood cells, is transmitted to humans by the bite of an infected female *Anopheles* mosquito. The *Anopheles* mosquito can live only in certain regions of the world (Fig. 1.6).

Fungal diseases are both more common and more serious in hot, humid regions. But some infectious diseases are highly limited geographically for reasons that are not well understood. For example, bartonellosis, which is also called *Carrión disease,* is found only in Peru, Ecuador, Chile, and Colombia. This disease resembles malaria superficially in that the minute rickettsia-like organisms invade and destroy erythrocytes. Humans are infected by the bite of the sand fly. Although conditions in other parts of the world should be favorable for this disease, it remains limited geographically.

Ebola was first discovered in 1976 near the Ebola River in what is now the Democratic Republic of the Congo. Since then, outbreaks have appeared sporadically in Africa. The natural reservoir host of Ebola virus remains unknown. However, on the basis of evidence and the nature of similar viruses, researchers believe that the virus is animal-borne and that bats are the most likely reservoir. Four of the five virus strains occur in an animal host native to Africa.

Taking a world view, there is widespread recognition of the importance of geographic factors in influencing human disease. For example, health care providers caring for Ebola patients and family and friends in close contact with Ebola patients are at the highest risk of getting sick because they may come in contact with infected blood or body fluids. Ebola also can be spread through direct contact with objects (like clothes, bedding, needles, syringes/sharps, or medical equipment) that have been contaminated with infected body fluids. Additionally, people can become sick with Ebola after coming in contact with infected wildlife. In Africa, Ebola may spread as a result of handling bushmeat (wild animals hunted for food) and contact with infected bats.

The World Health Organization (WHO) and the National Institutes of Health (NIH) have been deeply concerned with geographic problems in disease. Consult WHO and NIH home pages on the World Wide Web for additional information. (Web locations are provided on the Evolve website.)

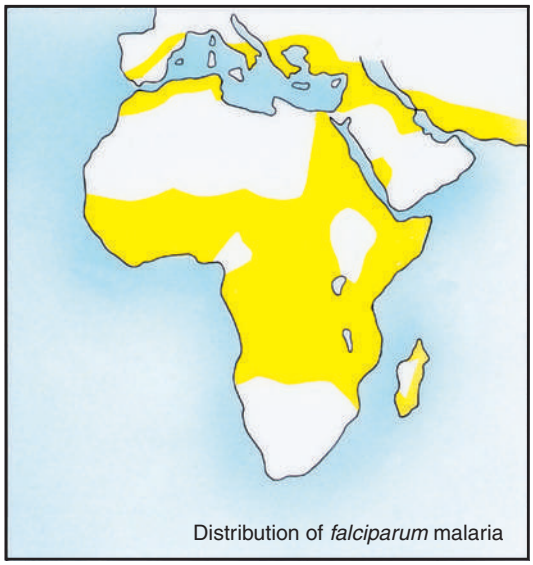


FIG 1.6 Geographic distribution of malaria. (From Centers for Disease and Control Prevention. https://www.cdc.gov/malaria/about/distribution.html. April 27, 2017.)

Levels of Prevention

The goal of health care should encompass much more than the prevention of illness. What is needed instead is some notion of positive health or physical "wholeness" that extends beyond the absence of ill health. The WHO defines health as complete physical, mental, and social well-being and not merely the absence of disease or infirmity. For some individuals, health implies the ability to do what they regard as worthwhile and to conduct their lives as they want. Aging and ill health are not synonymous, and many elders enjoy excellent health, even in the face of chronic disease (Fig. 1.7).

Epidemiologists suggest that treatment implications fall into categories called *levels of prevention.* There are three levels of prevention: primary, secondary, and tertiary. Primary prevention is prevention of disease by altering susceptibility or reducing exposure for susceptible individuals. Secondary prevention (applicable in early disease, i.e., preclinical and clinical stages) is the early detection, screening, and management of the disease. Tertiary prevention (appropriate in the stage of advanced disease or disability) includes rehabilitative and supportive care and attempts to alleviate disability and restore effective functioning.

Primary prevention. Prolongation of life has resulted largely from decreased mortality from infectious disease. Primary prevention in terms of improved nutrition, economy, housing, and sanitation for those living in developed countries is also responsible for increased longevity. Certain childhood diseases—measles, poliomyelitis, pertussis (whooping cough), and neonatal tetanus—are decreasing in prevalence, due to a rapid increase in coverage by immunization programs. More than 120 million children younger than age 5 in India were immunized against poliomyelitis in a single day in 1996. Globally, coverage of children immunized against six major childhood diseases increased from 5% in 1974 to 80% in 1995. In 1985 Rotary International launched the PolioPlus program to protect children worldwide from the cruel and fatal consequences of polio. In 1988 the World Health Assembly challenged the world to eradicate polio. Since that time, Rotary International's efforts and those of partner agencies, including the WHO, the United Nations Children's Fund, the CDC, and governments around the world, have achieved a significant reduction in the number of polio cases worldwide. Although the partners had pledged to wipe out the polio virus by 2018, that would have meant that transmission of the virus would have to be stopped by the end of 2015, because eradication may be certified by the WHO only 3 years after the last case.

As of September 2015, 41 cases were reported: 32 in Pakistan and 9 in Afghanistan. Some of the strategies used in Nigeria, which ended polio as of 2015, are now being implemented in Pakistan. But violence against vaccinators remains a serious obstacle.

The prevalence of cardiovascular diseases in developed countries (except those in Eastern Europe) is diminishing, thanks to the spread of health education and promotion. Infant and child death rates and the overall death rate are continuing to decrease globally.

High school health education programs are other examples of primary prevention efforts. Primary prevention also includes adherence to safety precautions, such as wearing seat belts, observing the posted speed limit

FIG 1.7 Healthy aging: elders exercising in an aerobics class **(A)** and painting **(B)** illustrate the concept that aging and disease are not synonymous. The artist, a healthy woman in her mid-70s, is also a breast cancer survivor. (Photographed by Therese A. Capal, Rockville, MD.)

on highways, and taking precautions in the use of chemicals and machinery. Violent crimes involving dangerous weapons must be stopped to achieve primary prevention of the traumatic or fatal injuries they cause.

Environmental pollutants, such as benzene, mold, asbestos, and lead, poison the body's organs. Some experts fear the emergence of an epidemic of cancer attributable to the carcinogenic chemicals afflicting the environment. Public health measures to ensure clean food, air, and water prevent many diseases, including cancer. As air, water, and soil quality is improved, the risk of exposure to harmful carcinogens is minimized.

Secondary prevention. Yearly physical examinations and routine screening are examples of secondary prevention that lead to the early diagnosis of disease and, in some cases, cures. The routine use of Pap smears has led to a decline in the incidence of invasive cancer of the uterine cervix. Also, more women are examining their own breasts monthly for cancer; thus earlier diagnoses are achieved.

Prenatal diagnosis of certain genetic diseases is possible. New diagnostic laboratory techniques provide definitive information for the genetic counseling of parents. This information can aid in predicting chances of involvement or noninvolvement of offspring for a given genetic disorder.

Amniocentesis is a diagnostic technique that consists of removing a small amount of fluid from the amniotic sac that surrounds the fetus and analyzing the cells and chemicals in the fluid. Blood samples can also be obtained from the fetus by amniocentesis; the amniotic fluid and fetal blood are then studied to determine defects in enzymes, to ascertain gender, and to measure substances associated with defects in the spinal cord and brain.

Tertiary prevention. Once a disease becomes established, treatment—within the context of traditional Western medicine—generally falls into one of the following two major categories: medical (including such measures as physical therapy, pharmacotherapy, psychotherapy, radiation therapy, chemotherapy, immunotherapy, and experimental gene therapy) and surgical. Numerous other subspecialties of medicine and surgery also have evolved to focus on a given organ or technique. In a clinical setting, a large array of professional caregivers provides rehabilitative and supportive tertiary prevention to the affected individual. Every professional brings the perspective of his or her discipline to the caregiving situation. Each makes clinical judgments about the patient's needs and problems and decides which goals and intervention strategies are most beneficial.

KEY POINTS

- Epidemiology is the study of patterns of disease in human populations.
- Diseases may be endemic, epidemic, or pandemic depending on location and the number of people affected.
- Aggregate factors such as age, ethnicity, gender, lifestyle, socioeconomic status, and geographic location are epidemiologic variables that influence the occurrence and transmission of disease in populations.
- Understanding the epidemiologic aspects of a disease is essential for effective prevention and treatment.

SUMMARY

Most people recognize what it is to be healthy and would define disease or illness as a change from or absence of that state. Under closer scrutiny, the concept of health is difficult to describe in simple, succinct terms. Correspondingly, the concepts of disease and illness also are complex. Environment, genetic constitution, socioeconomic status, lifestyle, and previous physical health all affect the timing and ultimate expression of disease in individuals.

Because humans exhibit considerable diversity, healthy structure and function are not precisely the same in any two individuals. By discovering common and expected patterns of responses to abnormalities, general prediction of etiology, pathogenesis, clinical manifestations, and targeted levels of prevention and intervention becomes possible.

RESOURCES

General Concepts of Pathophysiology

Dorland W: *The American illustrated medical dictionary*, St Louis, MO, 2015, Elsevier.

Graham K, et al: Alcohol-related negative consequences among drinkers around the world. *Addiction* 106(8):1391–1405, 2011.

National Institutes of Health, Office of AIDS Research. Available at http://www.niaid.nih.gov/topics/hivaids/understanding/howhivcausesaids/pages/hivcausesaids.aspx (Accessed 9 January 2015).

Pickering LK: Syphilis. In *Red book: report of the committee on infectious diseases*, ed 30, Elk Grove Village, IL, 2015, American Academy of Pediatrics, pp 755–768.

Venes D, Taber CW: *Taber's cyclopedic medical dictionary*, ed 22, Philadelphia, 2013, F.A. Davis, 1752.

Zyga S, Kolovos P: Cardiovascular Disease and Chronic Inflammation in End Stage Kidney Disease. *Int J Caring Sci* 6(1):29–36, 2013.

Normal Variations and Statistical Analysis

Balthazart J: Minireview: hormones and human sexual orientation. *Endocrinology* 152(8):2937–2947, 2011.

Blair ML: Sex-based differences in physiology: what should we teach in the medical curriculum? *Adv Physiol Educ* 31(1):23–25, 2007.

Buckley TM, Schatzberg AF: On the interactions of the hypothalamic-pituitary-adrenal (HPA) axis and sleep: normal HPA axis activity and circadian rhythm, exemplary sleep disorders. *J Clin Endocrinol Metab* 90(5):3106–3114, 2005.

Clinical and Laboratory Standards Institute (CLSI): *Defining, establishing, and verifying reference intervals in the clinical laboratory: approved guidelines—third edition*, Wayne, PA, 2015, Author. NCCLS document EP28-A3c (Formerly C28-A3c).

Hopfl G, Ogunshola O, Gassmann M: Hypoxia and high altitude: the molecular response. *Adv Exp Med Biol* 543:89–115, 2003.

Lindahl TL, et al: INR calibration of Owren-type prothrombin time based on the relationship between PT% and INR utilizing normal plasma samples. *Thromb Haemost* 91(6):1223–1231, 2004.

Linton AD, Lach HW: *Matteson & McConnell's gerontological nursing: concepts and practice*, ed 3, Philadelphia, 2006, Saunders.

Wu AHB: *Tietz clinical guide to laboratory tests*, ed 4, Philadelphia, 2006, Saunders.

Patterns of Disease and Epidemiology

Centers for Disease Control and Prevention: *Ebola (ebola virus disease)*. Available at http: https://www.cdc.gov/vhf/ebola/index.html. (Accessed 30 May 2017).

Daynard RA: Lessons from tobacco control for the obesity control movement. *J Public Health Policy* 24(3-4):274–290, 2003.

Division of Cancer Epidemiology & Genetics: *Cancer, organic solvents and other industrial chemicals*, Bethesda, MD, National Cancer Institute, U.S. National Institutes of Health. (Accessed 9 January 2015).

Facts about polio. Available at http://www.who.int/topics/poliomyelitis/en/Accessed (Accessed 9 June 2015).

Jekel JF: *Epidemiology, biostatistics, preventive medicine, and public health,* 4e, 2013, Elsevier.

Klippel JH: *Primer on the rheumatic diseases*, ed 13, Atlanta, GA, 2008, Arthritis Foundation, pp 86–93, 184–190.

Laden F, et al: Reduction in fine particulate air pollution and mortality: extended follow-up of the Harvard Six Cities study. *Am J Respir Crit Care Med* 173:667–672, 2006.

McCabe BW, et al: Practice of health-promoting behaviors by nursing home residents. *West J Nurs Res* 27(12):1000–1016, 2005.

Puechal X, Fiessinger JN: Thromboangiitis obliterans or Buerger's disease: challenges for the rheumatologist. *Rheumatology (Oxford)* 46:192–199, 2007.

Sartin J, et al: Medical management issues surrounding community-acquired pneumonia in adults. *Gundersen Lutheran Med Found J* 1(2):6–9, 2003.

Surks MI, et al: Subclinical thyroid disease: scientific review and guidelines for diagnosis and management. *JAMA* 291:228–238, 2004.

Trends in tuberculosis incidence—United States, 2006. *JAMA* 297:1765–1767, 2007.

Wadden TA, et al: Randomized trial of lifestyle modification and pharmacotherapy for obesity. *N Engl J Med* 353(20):2111–2120, 2005.

Walsh SJ, Rau LM: Autoimmune diseases: a leading cause of death among young and middle-aged women in the United States. *Am J Public Health* 90(9):1463–1466, 2000.

Wesolowski A, Metcalf CJ, Eagle N, et al: Quantifying seasonal population fluxes driving rubella transmission dynamics using mobile phone data. *Proc Natl Acad Sci USA* 2015.

Wizemann TM, Pardue M-L, editors: *Exploring the biological contributions to human health: does sex matter?* Washington, DC, 2001, Institute of Medicine Board of Health Sciences Policy.

World Health Organization: *Tropical diseases research*. Available at http://www.who.int/en/. (Accessed 9 January 2015).

World Health Organization: *World Malaria Report:* Available at http://www.who.int/en/. (Accessed 9 June 2015).

World Health Organization: Available at http://www.who.int/en/. (Accessed 9 June 2015).

2

Homeostasis, Allostasis, and Adaptive Responses to Stressors

Debra A. Jansen and Roberta J. Emerson

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- What is the relationship between homeostasis and allostasis?
- How do the sympathetic nervous system and neuroendocrine system respond to stress?
- What are the key features of Selye's General Adaptation Syndrome?
- What factors affect the stress response?
- How does allostatic overload contribute to the development of disease?

CHAPTER OUTLINE

Homeostasis and Allostasis, 12
Homeostasis, 12
Allostasis, 13
Stress as a Concept, 13
The General Adaptation Syndrome and Allostasis, 14
Alarm Stage, 14
Resistance or Adaptation Stage, 14
Exhaustion Stage, 16
Stressors, Gender and Developmental Influences, and Risk Factors, 16
Neurohormonal Mediators of Stress and Adaptation, 17
Catecholamines: Norepinephrine and Epinephrine, 17
Adrenocortical Steroids: Cortisol and Aldosterone, 18
Endorphins, Enkephalins, and Immune Cytokines, 19
Sex Hormones: Estrogen, Testosterone, and Dehydroepiandrosterone, 19
Growth Hormone, Prolactin, and Oxytocin, 19
Adaptation, Coping, and Illness, 20
Adaptation, Coping, and Resilience, 20
Allostatic Overload and Illness, 21

Humans, like all living organisms, must be able to respond and adapt to alterations in the environment. Changes in the external environment, such as moving outside from a warm house on a cold winter day, demand physiologic adjustments in the body's internal environment beyond the simple addition of layers of clothing to the outside of the body. Variations in the internal environment, such as a fever caused by infection, also necessitate physiologic responses to return the body's temperature to the normal range. The human organism maintains a variety of highly complex interactions with both internal and external environments. These interactions facilitate ongoing compensatory changes designed to support the organism physically and psychologically. Compensatory changes are necessary because they allow the perpetuation of both the individual and the species. Researchers, however, have found that the body's efforts to adapt to prolonged and repeated or extraordinarily demanding environmental changes may be associated with many physical and psychological health problems. This chapter explores the historical and current perspectives of homeostasis, allostasis, and stress responses and their relationship to health and illness.

HOMEOSTASIS AND ALLOSTASIS

Homeostasis

The word *homeostasis* is derived from the Greek words *homeo,* or similar, and *stasis,* or standing still, and means maintaining internal conditions in a stable state by keeping parameters relatively the same. Homeostasis often is conceptualized as a state of being in which all systems are in balance around a particular ideal "set-point." From this perspective, bodily changes formerly seen as conflicting or detrimental are understood as adaptive or compensatory to the maintenance of homeostasis within the body as a whole. Homeostasis reflects a tendency to stabilize an organism's functional systems, despite changes both internally and externally. Deviations from homeostasis resulting from these changes require elaborate systems to support the return of balance to the body. Over the past several decades, the definition of homeostasis has been criticized as being inadequate in encompassing the entire process of maintaining a stable state in complex organisms. But the fact remains that homeostatic concepts are an essential starting point for an exploration of stress, adaptation, and disease.

Claude Bernard, a nineteenth-century French physiologist, is credited with describing the basic premise of homeostasis. He believed that the various vital physiologic mechanisms of the body had as their goal the maintenance of a uniform and constant internal environment, or *milieu intérieur,* for the body. The stability of the internal environment was deemed necessary for the survival of the person, independent or free of the external environment. Disease occurred when the body did not respond appropriately to maintain internal stability when threatened by perceived or actual events. Building on Bernard's work, Walter B. Cannon created a concept that he referred to as *homeostasis* in his 1932 book *The Wisdom of the Body.* Homeostasis, according to Cannon, was a process in which each of the body's biochemical or physiologic variables (e.g., body temperature; oxygen, sodium, calcium, and glucose levels; and pH) was maintained within a narrow set-point range. Negative feedback loops sensed and corrected deviations from the set-point ranges for the variables, thereby supporting the survival of the individual, despite threats from the external or internal environments. These environmental threats could range from temperature extremes and water loss or gain, to "savage creatures" and bacterial infections. Box 2.1 provides examples of homeostatic systems designed to support the life of the person in the most basic sense.

Allostasis

The original concept of homeostasis, with the principle that the body attempts to achieve balance around a single optimal level or set-point for a given physiologic variable, has been challenged in recent decades. The innate complexity of biological organisms requires that set-points be readjusted for different circumstances (i.e., diverse situations necessitate different homeostatic set-points), including the anticipation of increased demands. For example, respiratory rate needs to increase when vigorously exercising or when ill with pneumonia to obtain more oxygen. At the same time, when responding to an internal or external environmental challenge (i.e., a stressor), *multiple* physiologic parameters may have to raise or lower their levels or actions to meet the demands posed by the challenge. Useful changes in one body system, though, may be detrimental to another, especially if prolonged; these changes, however, may ultimately be needed to support the survival of the organism as a whole at that particular point in time. For instance, in shock, when the life of the organism is at risk, blood flow to essential organs (brain and heart) is maintained by reducing perfusion to the kidneys, skin, and gastrointestinal tract. Simply stated, the body is not concerned about digesting dinner or making urine when it is trying to divert resources to a struggling brain and heart.

In 1988 Sterling and Eyer introduced the concept of allostasis in recognition of the complexity and variable levels of activity necessary to reestablish or maintain homeostasis. They described allostasis as the ability to successfully adapt to challenges. To survive, "an organism must *vary* all the parameters of its internal milieu and match them appropriately to environmental demands." Like homeostasis, allostasis is a derivation of the Greek words *allo,* meaning variable or different, and *stasis,* meaning standing still. Therefore this term accentuates the role of allostatic systems in maintaining the organism's stability by varying or changing. Allostasis is a dynamic process that supports and helps the body achieve a steady-state. In essence, the organism's overall stability is accomplished through change.

Allostasis involves intricate regulatory processes orchestrated by the brain. Through these processes, the body's parameters are continuously reevaluated and readjusted to match resources to the needs dictated by the situation. These parameter readjustments (e.g., of heart rate, blood pressure, or glucose levels) entail altering multiple set-points such that the person may be functioning at reduced or elevated levels or rates for numerous physiologic variables. Thus an individual may have different set-points for different circumstances (e.g., when resting versus running or when healthy versus sick). Allostasis comes into play in the complexity of social interactions and in our responses to adverse childhood experiences and low socioeconomic status (SES), exposure to environmental pollutants, and the demands of everyday modern living and working conditions, as well as in critical illness. The concept has garnered broad support in both the physical and the behavioral sciences. It seems especially applicable to subsequent discussions of adaptation and disease.

BOX 2.1 Examples of Homeostatic Systems

Baroreceptor response to acute changes in blood pressure
Vasopressin/antidiuretic hormone release from the posterior pituitary in response to changes in serum osmolality
Hypothalamic-mediated responses to changes in body core temperature
Central chemoreceptor responses to changes in $Paco_2$
Parathyroid gland response to changes in serum calcium level

KEY POINTS

- Contemporary concepts of homeostasis have a long history, reaching back to the ancient Greeks.
- Homeostasis is a state of equilibrium, of balance within the organism.
- Homeostatic responses refer to systems whose purpose is specifically to normalize selected physiologic variables.
- Allostasis is the overall process of adaptive change necessary to maintain survival and well-being.
- Allostasis may involve altering multiple physiologic variables to match the resources of the body to environmental demands. It helps the body achieve homeostasis.

STRESS AS A CONCEPT

Referring to stress as an "ambiguous" term is an understatement. Its ubiquitous use in everyday parlance is matched by its frequent presence in the health and psychology literature. Stress often is interpreted as a physical, chemical, or emotional factor that produces tension in the body or the mind ("He's experiencing a lot of stress"). But it also can mean the actual physical and mental state of tension ("I feel stressed"). Others use the term *stress* in relation to the response by the body to internal and external demands. Stress can be defined as a real or perceived threat to the balance of homeostasis. The neuroendocrinologist Robert Sapolsky more specifically distinguishes between the stress terminology and defines a stressor as anything that throws the body out of allostatic balance, whereas the stress response is the body's effort to try to restore the balance. To that end, stress is a natural outgrowth of the concept of homeostasis, but is even more applicable to the dynamic concept of allostasis. Sapolsky's definition also underscores an important point: The stress response by the body is meant to be helpful, at least in the short term; however, it becomes damaging when repeatedly activated or when it does not cease.

As early as the 1920s, Walter Cannon used the term *stress* in relation to humans and medicine. Hans Selye, however, often is erroneously credited with being the first person to borrow the term from the fields of engineering and physics and apply it to the human condition. In the

1930s Selye was experimenting with assorted ovarian and placental hormonal preparations and other tissue extracts and toxic agents. He was injecting these into rats when he serendipitously uncovered a biological basis for stress. Selye was expecting to find different physiologic responses in the rats, depending on which of the various substances was injected; however, to his surprise and disappointment, the same three changes occurred each time. In every animal tested, the cortex of the adrenal gland enlarged, lymphatic organs (thymus, spleen, and lymph nodes) shrank, and bleeding peptic ulcers developed in the stomach and duodenum. When Selye experimented with other noxious stimuli, such as exposing the rats to temperature extremes, surgery, or forced exercise, the same three changes occurred. Any kind of harmful physical stimuli he used produced the same observed physiologic changes. Selye termed the harmful stimuli or causative agents stressors and concluded that the changes observed represented a nonspecific response by the body to any noxious stimulus or demand, a general "stress" response. Because so many different agents caused the same changes, Selye called this process a general adaptation syndrome (GAS) with three components: an *alarm reaction,* a *stage of resistance,* and a *stage of exhaustion.* According to Selye, when confronted by stressors during daily life, individuals move through the first two stages repeatedly and eventually become adapted and "used to" the stressors.

Selye's original conceptualization of the stress response and GAS has been criticized as being too simplistic for the complexities of humans. In particular, evidence suggests the body does not produce the same responses to all types of stressors. Depending on the type and severity of stressor, different patterns of hormone release occur, with more of some substances and less of others being produced and at different speeds and for varying lengths of time. Moreover, Selye's early work in the 1930s concentrated on stimuli of a physical or biological nature. Beginning in the 1970s, researchers began to realize that perception of these stimuli was important to individuals' responses to stress and that responses could be physiologic, as Selye described, as well as behavioral in nature. When stress is generated by extreme psychological or environmental demands, balance is disrupted, and allostatic reactions are initiated to restore balance. The discussion that follows presents the GAS as a reflection of the responses to these diverse stimuli and incorporates much of the knowledge acquired since Selye's early pioneering work.

The General Adaptation Syndrome and Allostasis

Components of the GAS can be subdivided into three unique, largely physiologic stages (Table 2.1). Examining the stages separately is the best way to understand the entire GAS. The specific chemicals involved are among those seen today as integral to the broader view of allostatic responses to stress in the maintenance of homeostasis. All will be discussed later in the chapter.

Alarm Stage

The alarm stage has been called the *fight-or-flight response*, derived from Cannon's work, because it provides a surge of energy and physical alterations to either evade or confront danger (Fig. 2.1). This stage begins when the hypothalamus, as it monitors the internal and external environment, senses a need to activate the GAS in response to a stimulus, a stressor placing the balance of homeostasis at risk. The stressor might be physical or emotional, positive or negative—arguing with a friend, having an upper respiratory tract infection, running to catch a bus, or winning the lottery. The hypothalamus then secretes corticotropin-releasing hormone (CRH) to activate the sympathetic nervous system (SNS), which in turn also stimulates the adrenal medulla (the inner portion of the adrenal gland) to release the catecholamines—norepinephrine and epinephrine. The increased levels of catecholamines enable the body to rapidly take action to fight or flee the stressor. This series of events is part of the sympathetic-adrenal-medullary system, originally referred to as the fight-or-flight response by Walter Cannon. Additionally, the hypothalamus secretes CRH to stimulate the anterior pituitary gland to release adrenocorticotropic hormone (ACTH). ACTH then causes the adrenal cortex (the outer portion of the adrenal gland) to release substantial amounts of the glucocorticoids, specifically cortisol, eliciting its diverse responses. This cascade of effects is termed the *hypothalamic-pituitary-adrenal (HPA)* axis. Once the pituitary gland is activated, the alarm stage progresses to the stage of resistance. This coordinated systemic response to stress is illustrated in Fig. 2.2.

TABLE 2.1 Stages of the General Adaptation Syndrome

Alarm	Resistance	Exhaustion
Increased secretion of glucocorticoids and responses	Eventual normalization of glucocorticoid secretion	Increased glucocorticoid secretion followed by significant dysregulation
Increased sympathetic nervous system activity	Eventual normalization of sympathetic nervous system activity	Diseases of adaptation
Increased secretion of epinephrine (and some norepinephrine) from adrenal medulla	Eventual normalization of epinephrine and norepinephrine secretion from adrenal medulla	Loss of resistance to stressor; possible death of organism
Fight-or-flight manifestations	Resolution of fight-or-flight manifestations	
Reduced resistance to stressors	Increased resistance (adaptation) to stressor	

Allostasis is essentially the activation of these stress responses to evoke changes that return the organism to homeostasis. Mediators of allostasis include the aforementioned hormones and neurotransmitters of the HPA axis and the sympathetic-adrenal-medullary system (e.g., cortisol, epinephrine, and norepinephrine), various other hormones presented later in this chapter, and also cytokines from the immune system. The alarm stage of the stress response with the release of its various substances is meant to be helpful to the organism in overcoming the stressor.

Resistance or Adaptation Stage

If the alarm stage were to persist, the body would soon suffer undue wear and tear and become subject to permanent damage and even death. To survive, the body must move beyond the alarm stage to a stage of resistance (also called *adaptation*) supportive of the allostatic return to a state of homeostasis. As the body moves into the stage of resistance, the sympathetic nervous system and adrenal medulla and cortex are functioning at full force to mobilize resources to manage the stressor. The resources include glucose, free fatty acids, and amino acids. Concentrations of these chemicals are elevated through the effects of cortisol and the catecholamines (i.e., epinephrine and norepinephrine). These resources are used for energy and as building blocks, especially the amino acids, for the later growth and repair of the organism after

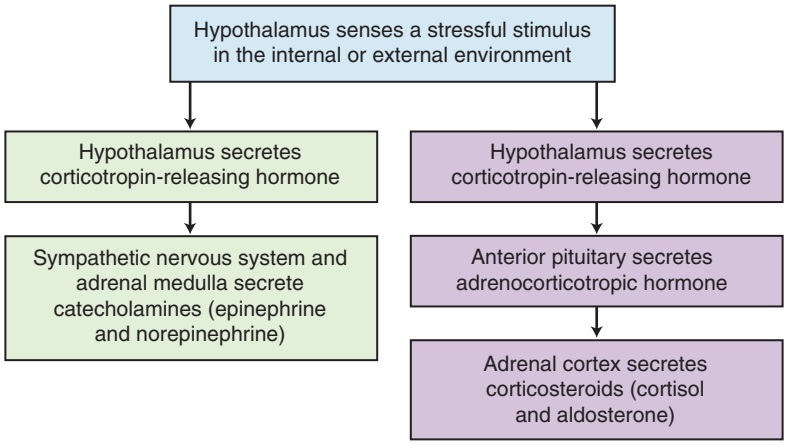


FIG 2.1 Steps of Selye's alarm stage of the general adaptation syndrome. (Modified from McKenry L et al: *Mosby's pharmacology in nursing,* ed 22, St Louis, 2006, Mosby.)

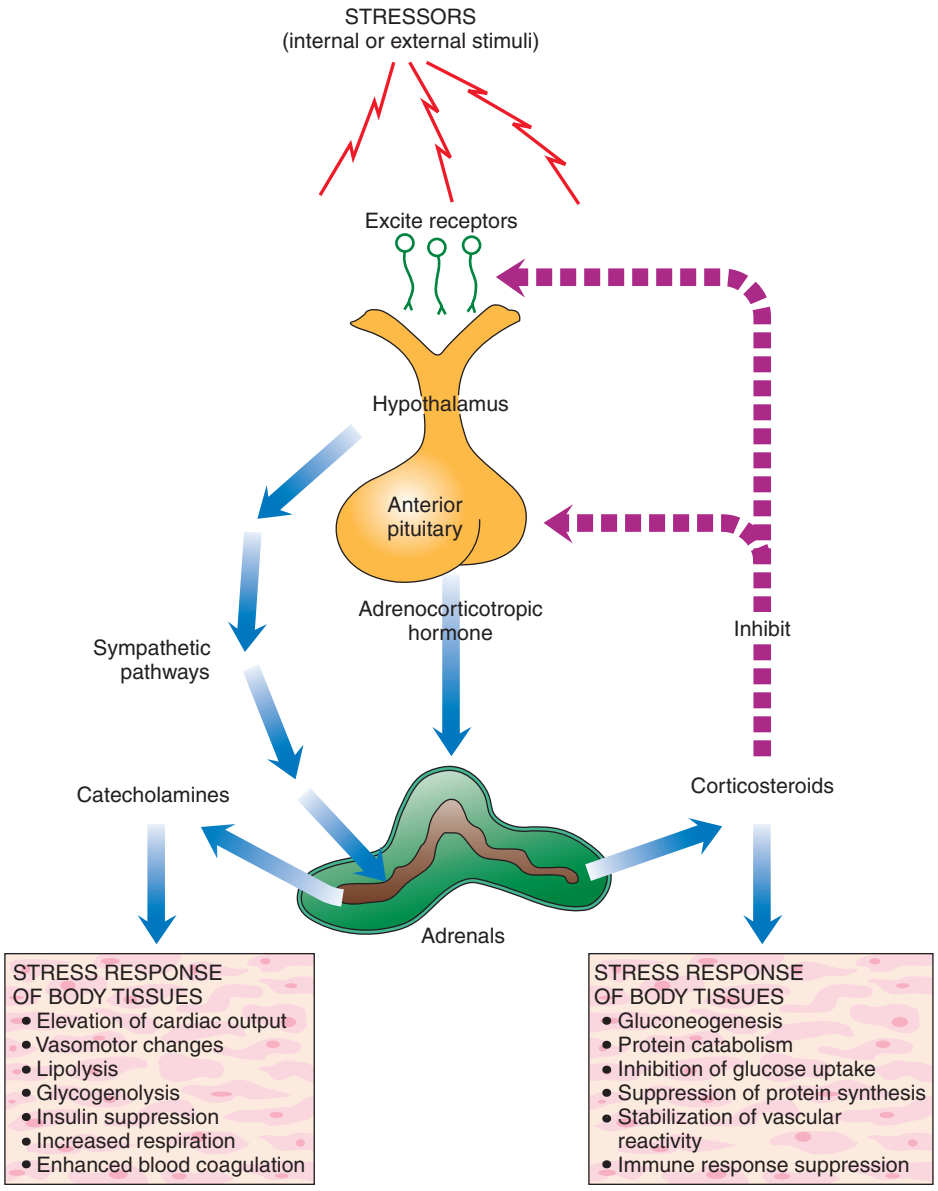


FIG 2.2 Neuroendocrine interactions in response to a stressor. Receptors are excited by stressful stimuli and relay the information to the hypothalamus. The hypothalamus signals the adrenal cortex (by way of the anterior pituitary) and the sympathetic pathways (by way of the autonomic nervous system). The stress response is then mediated by the catecholamines (i.e., epinephrine and norepinephrine) and by the glucocorticoids (predominantly cortisol).

the stress abates. If the stressor is adequately addressed and resolved, the organism returns to its steady-state, having reattained allostatic balance. This process described by Selye is clearly a part of the more recently described process of allostasis. However, with the current understanding of allostasis, it is possible that the organism may have to function at a new baseline steady-state for different physiologic variables, either higher or lower than the previous set-points to achieve adaptation. For instance, the normal partial pressure of carbon dioxide (P_{CO_2}) in the blood is 35 to 45 mm Hg, and the normal oxygen saturation is greater than 94% in a healthy individual. For someone with chronic obstructive pulmonary disease, a new normal P_{CO_2} value might be 50 to 60 mm Hg and the oxygen saturation may be 88% to 90%, while still maintaining a homeostatically normal serum pH.

Exhaustion Stage

Exhaustion occurs when the body is no longer able to effect a return to homeostasis after prolonged exposure to noxious agents. Selye postulated that when energy resources are completely depleted, death occurs because the organism is no longer able to adapt. He speculated that individuals are born with a given amount of adaptation energy. However, when these adaptive energy stores are depleted, no other resource exists to facilitate recovery. Diseases of adaptation such as hypertension and heart disease occur when the body is continuously taxed by stressors. It is now understood that exhaustion and stress-related disease do not necessarily occur because resources are depleted; instead, they can occur because the actual stress response itself, with all its various biological mediators, can be harmful when repeatedly activated.

Concepts related to allostasis help with understanding the damaging effects of stress. The HPA axis, the sympathetic-adrenal-medullary system, and other systems (including the immune system) work to help the person adapt to and defend against stressors. Wear and tear on the body and on the brain occurs when these body systems are chronically overactivated or underactivated in their attempts to support an allostatic return to homeostasis. The accumulation of all the various mediators produced by the systems is damaging to tissues over time. This wear and tear on the body and brain is called *allostatic load*. Allostatic load is basically due to the typical demands that are part of daily life as well as unpredictable events. However, with chronic, unremitting, or excessive demands, allostatic load can become an overload. This allostatic overload reflects the "cost" to the body's organs and tissues for an allostatic response that is excessive or ineffectively regulated and unable to deactivate.

Stressors, Gender and Developmental Influences, and Risk Factors

Stressors are agents or conditions that are capable of producing stress and endangering homeostasis. They initiate stress response systems to return to a state of allostatic balance. Every day the human organism encounters stressors. These may be external to the individual (e.g., air pollution, radiation, a motor vehicle accident) or internal (e.g., low blood glucose level or a threat to self-esteem). Common general stressors are physical (e.g., extreme hot or cold air temperature), chemical (e.g., auto exhaust), biological (e.g., bacteria and viruses), social (e.g., overcrowding, difficult relationships, information overload), cultural (e.g., behavioral norms), or psychological (e.g., feelings of hopelessness). Stressors of an emotional or mental origin may be present or anticipated, or may involve the recollection of prior traumatic events. Less commonly noted but extremely powerful stressors are psychosocial experiences over which a person may have little or no personal control. Racial and socioeconomic stressors as well as childhood abuse can produce many of the manifestations of stress described in this chapter.

Stressors vary in their scope, intensity, and duration. A stressor of less intensity can still have a significant impact if it persists for some time. A glass of water held at arm's length poses little stress initially, but as minutes turn into hours the stress on the body escalates. Even events associated with happiness may serve as stressors—holidays, childbirth, and vacations. Stressors of all types challenge human adaptation.

The identification of specific stressors in isolation provides little insight into today's complex global society. A given stressor, depending on its type and duration, may have a particular pattern of hormone and neuropeptide release. Researchers have explored innumerable factors that can indirectly increase or decrease the impact of stressors. It is now generally well accepted that individual personal characteristics, as well as the psychological and circumstantial context of the situation, developmental state of the person, and availability of coping resources, allow for a great deal of variation in the way humans perceive and respond to stressors, and thus the type of stress response produced. The activation of both the sympathetic-adrenal-medullary system response and the HPA axis occurs with a wide variety of physical, mental, and psychosocial stressors. The HPA axis with its glucocorticoid response, however, seems to be notably prominent and dysregulated in cases of depression and posttraumatic stress disorder (PTSD), and is also active when a person's sense of self is negatively evaluated or the person lacks a sense of control. On the other hand, the sympathetic system is particularly active with anxiety and vigilant states. Furthermore, personality characteristics have been found to be associated with variations in cortisol release and sympathetic-adrenal-medullary system activation in the stress response. Indeed, the effect of personality on the stress response differs with the situational context. An individual's appraisal of a situation and subsequent reaction is influenced by one's past experiences and conditioning, cultural influences, social expectations, the availability of social support, and a person's genetic and epigenetic profile and gender.

Beginning in the early 1970s researchers started to examine gender differences related to stress, and recent research has continued to expand what is known about these differences. For example, one study in the 1980s examined the differences in performance and stress responses of men and women under controlled laboratory conditions. When subjected to a stressful task, there was a 50% to 100% increase in epinephrine release in men, whereas there was little, if any, increase noted in women, who were also found to perform as well as or better than their male counterparts. Women did have an elevation in epinephrine release in a more real-life stress situation (i.e., an academic examination), but these elevations remained well below those of men.

Although some researchers consider these stress response differences between men and women, at least in part, to be related to gender-associated roles and psychological factors, other researchers also attribute these variances to the effects of the sex hormones on the stress response. Data from a recent National Health Interview Survey of adults revealed that in every age group, more women than men reported having serious psychological distress. Women are known to have higher prevalence rates of stress and anxiety-related disorders such as depression and PTSD, whereas men show more substance abuse disorders. Estradiol and estrogen may account for some of the gender differences, as the incidence of affective mood disorders appears to become similar for men and women after the age of 55, after menopause. Estradiol may heighten stress reactivity. Accordingly, stressors and traumatic events may have a greater influence if they occur when menstrual cycle estrogen levels are higher for premenopausal women.

To date, data regarding gender differences in physiologic responses to stressors are conflicting, with some researchers reporting no differences and others indicating higher or lower levels of cortisol or ACTH in

men versus women. The responses seem to depend on the types of stressors assessed. For instance, for adult men, higher cortisol levels were found in studies involving cognitive, harassment, mathematical, and verbal stressors and with social stress testing, whereas higher cortisol levels were found in women compared with men in response to a social interaction stressor. Researchers believe very slight differences in brain architecture due to the effects of sex hormones may result in men and women interpreting and responding to stressors with varied approaches.

Developmental stage of life and age also appear to effect the response to stressors. Children exposed to physical and/or psychosocial maltreatment have been shown to have smaller cortices in the prefrontal region of the brain, higher cortisol levels, and elevated levels of inflammation in comparison to other children. Many of these changes persist into adulthood. Additionally, infants whose mothers were depressed or anxious during pregnancy displayed higher cortisol levels in response to stressors. These studies demonstrate the vulnerability to stressors of brain development extending from the prenatal period through adolescence.

Risk factors alone are not inherently stressors, but rather conditions or situations that increase the likelihood of encountering or experiencing a stressor. Using a mobile phone to send text messages while driving is a risk factor for having a motor vehicle accident; running in the dark is a risk factor for falling; and inadequate immunization is a risk factor for certain infectious illnesses. Low SES is a risk factor for exposure to lead and air pollutants in the home and neighborhood. Risk factors also include genetic predisposition and epigenetic factors, as well as adverse early life experiences. By being aware of risk factors, it is possible to decrease the probability of exposure to certain stressors and the associated threat to homeostasis.

KEY POINTS

- Stress is a real or perceived threat to the balance of homeostasis. The stress response is designed to restore balance.
- Selye's general adaptation syndrome (GAS) theory reflects the view of a nonspecific physiologic response to stress. It incorporates three stages indicating the changes in the body's systemic response: alarm, resistance, and exhaustion.
- Stressors are agents or conditions capable of producing stress.
- The body's response to stressors is meant to be helpful, at least initially, in terms of mobilizing resources to help manage stressors.
- Response to a stressor depends on its magnitude and the meaning that the stressor has for an individual. Perception of the stressfulness of a stressor depends on genetic constitution, gender, past experiences and conditioning, and social and cultural influences. Stressors may be external or internal. They may be physical, chemical, biological, sociocultural, or psychological.
- Individuals may be more vulnerable to the effects of stressors at certain times including developmental stage of life, the availability of resources, and the effects of other previous or concurrent stressors.
- Risk factors are conditions or situations that increase the likelihood of encountering or experiencing a stressor.

NEUROHORMONAL MEDIATORS OF STRESS AND ADAPTATION

Numerous hormones and signaling molecules are involved in the daily maintenance of homeostasis through allostatic processes. These mediators are briefly described here, and their roles in allostasis, adaptation, and disease are discussed in later parts of this chapter. A key idea to the understanding of homeostasis is that once the challenges contributing to allostatic load have been resolved, levels of these chemicals should return to their baselines. However, in cases of allostatic overload, pathologies of a physiologic, psychological, or behavioral nature may result.

Catecholamines: Norepinephrine and Epinephrine

Cannon identified that the body's response to threats resulted in the activation of the adrenal medulla and sympathetic nervous system and therefore the release of the catecholamine neurotransmitters. He deemed this the "sympathico-adrenal system" and believed it was ultimately responsible for what he termed the *fight-or-flight* reaction. The purpose of the fight-or-flight reaction was the maintenance of the physical and psychological integrity of the organism. The catecholamine neurotransmitters—epinephrine and norepinephrine—play integral roles in allostasis.

Release of catecholamines is initiated through the activation of the hypothalamus gland, a collection of nerve centers situated near the third ventricle close to the base of the brain (see Chapter 39). The cerebral cortex and limbic system (including the hippocampus and amygdala, important for memory and emotions) receive information regarding stressors and determine whether or not something is potentially harmful to the organism (i.e., whether it is stressful). They relay the information to the hypothalamus. (It should be noted, though, that the stress response, depending on the type of stressor, may occur to some extent even in comatose and sedated individuals.) In response to these stressors, the hypothalamus prompts the release of norepinephrine from the sympathetic branch of the autonomic nervous system and locus ceruleus (group of neurons in the brainstem pons) and epinephrine and some norepinephrine from the adrenal medulla. Norepinephrine is released by sympathetic neurons directly into the synaptic clefts near the effector organs and tissues. Additionally, preganglionic fibers from the sympathetic nervous system neurons synapse at the adrenal medulla, stimulating the release of epinephrine and, to a lesser extent, norepinephrine. The adrenal catecholamines are released into the bloodstream and travel to effector organs and tissues (endocrine). These circulating adrenal catecholamines have essentially the same effects as sympathetic nerve stimulation and are often seen as an extension of the sympathetic nervous system. The responses on the part of the sympathetic nervous system and the adrenal medulla may differ according to the stimulus. During situations such as exposure to cold temperatures the sympathetic nervous system response with norepinephrine production dominates. Emotional distress or acute hypoglycemia, however, causes a greater response from the adrenal medulla, with increased production of epinephrine.

The effects of catecholamines are profound. They affect cardiovascular function, control fluid volume by activating the renin–angiotensin–aldosterone mechanism, have a role in inflammation and immunity, and affect metabolism; and they are associated with attentiveness, arousal, and memory formation in the central nervous system (CNS). Norepinephrine is the primary constrictor of smooth muscle in blood vessels. It therefore regulates blood flow through tissues and its distribution through the organs, as well as, importantly, maintenance of blood pressure. It also reduces gastric secretion, inhibits insulin secretion, and innervates the iris and ciliary muscles of the eyes, thereby dilating the pupils and increasing night and far vision. Epinephrine enhances myocardial contractility and increases heart rate and venous return to the heart, thus increasing cardiac output. It additionally relaxes bronchial smooth muscle, thereby dilating the airways to enable better oxygenation. Epinephrine also increases glycogenolysis, the release of glucose from the liver and inhibits insulin secretion, further elevating blood glucose levels. In the brain, the increased blood flow and availability of glucose lead to augmentations in mental attention and alertness. Epinephrine and norepinephrine also exert immune system effects by influencing

the production of cytokines by immune and adipose cells. The effects of the catecholamines are summarized briefly in Table 2.2. For more detail, see Chapter 43.

Adrenocortical Steroids: Cortisol and Aldosterone

Among the most versatile hormones in the human body, glucocorticoids have regulatory roles in maintaining fluid volume, metabolism, immunity, inflammatory responses, and brain function (Table 2.3). Glucocorticoids are lipid-soluble hormones, allowing them to pass through cell membranes to bind with receptors in the cytosol or nucleus and initiate changes in cellular activities. Practically every body tissue has intracellular glucocorticoid receptors. As opposed to the catecholamines, the onset of their effects is slower, but the duration of action is longer.

TABLE 2.2 Brief Summary of Effects of Catecholamines on Tissues and Organs of the Body

Tissue/Organ	Catecholamine Effect
Heart	Increases rate
	Increases speed of impulse conduction
	Increases contractility
Respiratory tract	Relaxes bronchial smooth muscle to dilate airway
Vascular smooth muscle	Constricts to reduce perfusion
Skin, mesenteric bed, kidneys	
Skeletal muscle, lungs, heart	Dilates to increase perfusion
Peripheral vasculature	Constricts to increase blood pressure
Gastrointestinal tract	Decreases peristalsis
	Contracts sphincters
	Decreases gastric acid secretion
Eyes	Contracts radial muscle to dilate iris and pupil
	Relaxes ciliary muscle for far vision
Liver	Glycogenolysis and gluconeogenesis for increased glucose levels and thus energy
Central nervous system	Promotes arousal, attention, and vigilance

TABLE 2.3 Major Effects of Glucocorticoids in the Stress Response

Metabolism	Catabolism of muscle, fat, lymphoid tissue, skin, and bone
	Liver gluconeogenesis
	Opposes insulin in transport of glucose into cells
	Increased appetite
Fluid balance	Sodium and water retention
Inflammation and infection	Suppressed inflammatory response
	Increased neutrophil release
	Decreased new antibody release
	Decreased T-lymphocyte production and function
	Decreased production of eosinophils, basophils, and monocytes
Support catecholamines	Increased epinephrine synthesis
	Enhanced vasoconstriction

The glucocorticoids are so named because of their significant role in glucose metabolism. The primary glucocorticoid, cortisol, is secreted by the adrenal cortex in response to ACTH from the anterior pituitary. Release of ACTH is itself affected by another releasing hormone, CRH, from the hypothalamus. Negative feedback loops help to maintain cortisol level within a normal range. Cortisol binds to receptors on the hypothalamus and anterior pituitary gland to suppress CRH and ACTH release in a negative feedback loop.

The actions of the HPA axis and catecholamines may synergize or antagonize each other. Catecholamines facilitate the release of ACTH, thereby helping to maintain the function of the HPA axis and release of cortisol. Glucocorticoids promote adrenal medulla synthesis of epinephrine through control of the major enzyme phenylethanolamine *N*-methyltransferase (PNMT). Glucocorticoids also support the actions of the catecholamines in the maintenance of normal blood pressure and cardiac output. In skeletal muscle, catecholamines antagonize the catabolic glucocorticoid effects by impeding the breakdown of somatic protein. Together, the catecholamines and glucocorticoids facilitate the brain's development of memory, which is especially important when hazardous circumstances have occurred.

The metabolic effects of cortisol are significant. Cortisol affects protein metabolism. It has an *anabolic* effect leading to increased rates of protein synthesis in the liver. However, it has a *catabolic* effect in muscle, lymphoid, and adipose tissues and on skin and bone. This protein breakdown produces increased levels of circulating amino acids. The resulting pool of amino acids from catabolized proteins ensures their availability for the liver. Cortisol then stimulates gluconeogenesis in the liver and a sixfold to tenfold increase in the rate of amino acid conversion to ketoacids and glucose. The catabolism of adipose tissue releases free fatty acids and glycerol that also can be used for gluconeogenesis and to create ketoacids for fuel. Gluconeogenesis ensures an adequate supply of glucose for body tissues in general, but nerve cells have priority. Cortisol may act to preserve available glucose for brain nerve cell use by limiting the uptake and oxidation of glucose by other cells in the body. Cortisol also promotes appetite and food-seeking behaviors.

Glucocorticoids are known for their significant role in the control of the immune response. They suppress the acute-phase response to infection and inflammation, helping to curtail overactivity. This is accomplished by inhibiting the production of select immune cytokines (signaling molecules), by increasing the production of other cytokines, and in some cases by directly inhibiting the proliferation and activation of specific immune system cells. At the same time, when the acute stress of tissue injury or infection occurs, the resulting release of glucocorticoids and catecholamines assists the movement of the necessary immune cells to the affected location. However, with prolonged stress and chronic elevation in the levels of glucocorticoids, desensitization and down-regulation (decrease) of glucocorticoid receptors may occur on some immune cells, eventually resulting in fewer antiinflammatory effects over time. In fact, continued stress can even result in proinflammatory effects. Thus the relationship of the immune system to stress is quite multifaceted, and our understanding of it is evolving.

Aldosterone is the primary mineralocorticoid steroid hormone secreted by the adrenal cortex. Stimulation of the sympathetic nervous system activates the renin–angiotensin system, and the release of aldosterone is the final chemical outcome. The specific stressor of fluid volume depletion also activates the release of renin, similarly initiating the renin–angiotensin system. The primary effect of aldosterone, once bound to receptors in the kidneys' distal tubules and collecting ducts, is reabsorption of sodium and an increase in the excretion of potassium. Because of osmotic force, water tends to follow sodium; therefore enhanced reabsorption of sodium leads to increased extracellular fluid volume and increased blood pressure. Endogenous glucocorticoids have

a small amount of mineralocorticoid effect, but the greatest effect on circulating volume is through aldosterone. Additionally, angiotensin II, whose formation stimulates aldosterone release, is a potent vasoconstrictor. This chemical mediator provides support for the catecholamine-induced increase in blood pressure.

Endorphins, Enkephalins, and Immune Cytokines

Stress naturally activates the inhibition of pain through the release of small peptides called *endorphins* and *enkephalins.* First discovered in 1975, endorphins and enkephalins are endogenous opioids that are produced within the CNS and released in response to stressors, by certain foods (most notably chocolate), by laughter, and from massage or acupuncture. The term *endorphin* comes from *endogenous* and *morphine.* Like the opiate drug morphine, endorphins raise the pain threshold (reduce pain) and produce sedation and euphoria. During acute stress, the endogenous opioids also may attenuate the stress response, especially activation of the SNS by CRH. Some immune cells (B and T lymphocytes, granulocytes, macrophages, and monocytes) also produce several types of endorphins that are released in response to stressors, CRH, antiinflammatory cytokines, and catecholamines. Opioid receptors have been identified on immune cells, and when activated they modulate both immune cell proliferation and immune cell activity. On the other hand, in the presence of acute or chronic stress, activated immune cells (mast cells, neutrophils, macrophages, and T lymphocytes) can release proinflammatory cytokines that enhance pain. Pain is a classic manifestation of the inflammatory response (Chapter 9). Thus the central and peripheral nervous systems and the immune system maintain complex patterns of interactions with pain pathways as part of the allostatic mechanism to return to homeostasis.

Another example of the interaction between stress, the nervous system, and the immune system is interleukin-1 (IL-1), one of the cytokines secreted by macrophages and other immune cells. It is capable of affecting the production of CRH by the hypothalamus. Leukocytes are also capable of producing some of the other hormones, such as ACTH, that are involved in the signaling system. Some researchers propose that stressors of relatively short duration (e.g., less than 2 hours) could augment facets of immune function, including the emigration of immune cells from the lymphoid tissues (e.g., the spleen) to the skin and peripheral components of the vascular system. On the other hand, numerous studies over the years have shown that severe and persistent psychological stress can down-regulate or suppress immune functioning through innumerable and elaborate mechanisms. Immune system suppression caused by severe or persistent stress represents a direct link between stress and illness. Expanded understanding of the interrelationships between the nervous, endocrine, and immune systems holds great promise in the identification of new therapeutic interventions.

Sex Hormones: Estrogen, Testosterone, and Dehydroepiandrosterone

As noted previously, women and men often have different stress responses, and this may be partially attributable to the influences of the sex hormones and their interactions with the mediators of allostasis. Excessive stress appears, in general, to inhibit female reproduction and affect premenstrual symptoms. Cortisol exerts inhibiting effects on the female reproductive system by suppressing release of gonadotropin-releasing hormone, luteinizing hormone, estradiol, and progesterone; whereas estradiol normally down-regulates glucocorticoid receptor binding in the brain and alters regulatory feedback control. Androgens, such as testosterone and dehydroepiandrosterone (DHEA), also may inhibit the effects of glucocorticoids. Androgens oppose the catabolic effects of glucocorticoids on bone and the impact of glucocorticoids on lymphoid tissues, inflammatory cytokines, and leukocytes. DHEA interacts with numerous neurotransmitters in the brain, counteracting the depressive tendencies often noted with glucocorticoids. Testosterone also appears to have antidepressant and antianxiety effects as it elevates mood. Numerous stressful stimuli, such as illness, surgery, strenuous physical exercise, heart failure, and stressful academic programs, result in significant reductions in circulating testosterone levels. In combination with another hormone, vasopressin, testosterone enhances blood pressure and heart rate reactivity and augments the "fight-or-flight" response. In contrast, the hormone oxytocin (whose impact is modulated by estrogen) and the endogenous opioids are thought to produce a calming effect during times of stress, resulting in the notion that women may have a "tend and befriend" response rather than a "fight-or-flight" response in some situations.

Growth Hormone, Prolactin, and Oxytocin

Growth hormone (somatotropin) is released from the anterior pituitary gland and affects protein, lipid, and carbohydrate metabolism. It has anabolic effects, increasing protein synthesis and bone and muscle mass growth. It also increases fat mobilization (lipolysis) while decreasing the rate of carbohydrate utilization by peripheral tissues. Growth hormone is normally secreted in a cyclic basal pattern, primarily at night, and changes with developmental stage. Growth hormone secretion is highest during adolescence and then gradually declines during adulthood. Serum levels of growth hormone also increase acutely after a variety of intensely stressful physical or psychological stimuli, such as strenuous exercise or extreme fear. Growth hormone appears to enhance immune function. However, continued activation of the stress response eventually results in the decreased secretion of growth hormone, accounting for stunted growth in children experiencing prolonged chronic stress.

Prolactin is similar in structure to growth hormone and is secreted from the anterior pituitary gland in response to stress, sexual activity, pregnancy, and suckling (even in men) and breast feeding. It suppresses ovulation. Numerous tissues have receptors for prolactin in addition to the breast, including the kidneys, liver, and adrenal glands. Lymphocytes also have prolactin receptors, suggesting a role for prolactin in immune regulation. A significant increase in the level of growth hormone or prolactin tends to require more intense stimuli than the stress that increases the concentrations of catecholamines and glucocorticoids.

Oxytocin is produced during childbirth, lactation, and sexual behavior (in both genders) and has been associated with promoting bonding and social attachment. Oxytocin is thought to moderate the stress response and have a calming effect, with reductions in HPA and sympathetic activation and reduced perceived anxiety. In an intriguing study, researchers found that preteenage girls who spoke with their mothers on the phone immediately after a laboratory social stressor had higher levels of urinary oxytocin and lower levels of salivary cortisol than did those who merely texted their mothers. Hearing the mother's voice appeared important to the release of the oxytocin and buffered the stress response. Oxytocin also may have some analgesic effects. It is synthesized by the hypothalamus and secreted by the posterior pituitary gland and other brain regions. Oxytocin is believed to have stronger effects in females in comparison to males because of the interaction of estrogen and oxytocin.

Through interactions of the primary stress hormones—catecholamines and glucocorticoids—as well as numerous other mediating influences, the allostatic process needed to sustain the human organism is achieved. In some cases, these stress-related hormones have similar and synergistic effects and in others they work in opposition. This state of counterbalancing helps to facilitate allostasis, ideally returning the human organism back toward homeostasis.

KEY POINTS

- Modern views of allostatic maintenance of homeostasis in the face of stress are primarily derived from an understanding of negative feedback, as well as the roles of the sympathetic nervous system catecholamines and the glucocorticoid cortisol.
- The primary role of the sympathetic nervous system (SNS) is appraisal of a stressful stimulus and release of norepinephrine. Norepinephrine released from sympathetic nerve endings increases heart rate and contractility, constricts blood vessels to decrease blood flow to less essential tissues and organs and raise blood pressure, reduces gastrointestinal motility and gastric acid secretion, dilates the pupils, and inhibits insulin secretion.
- Stress stimulates sympathetic activation of the adrenal medulla to release epinephrine. Epinephrine's actions are similar to norepinephrine and are particularly important for increasing cardiac performance (increased heart rate, contractility, and cardiac output), promoting the release of glucose from the liver, and enhancing bronchodilation.
- Cortisol, from the adrenal cortex, has widespread effects on numerous tissues that are both synergistic and antagonistic with catecholamines and has an antiinflammatory role.
- Aldosterone promotes fluid volume expansion and increases blood pressure.
- Endorphins and enkephalins are released by the CNS in response to painful stressors, leading to decreased perception of pain and increased sedation and euphoria. Immune cells in the periphery also contribute to pain modulation.
- Immune cells respond to the hormones released by the HPA axis and sympathetic nervous system. They also release cytokines that in turn affect the functioning of these stress systems.
- Sex hormones and differential release of growth hormone, prolactin, and oxytocin produce effects on the stress response that may differ between genders.

ADAPTATION, COPING, AND ILLNESS

Although much has been learned about the dynamic biological systems and human/environmental interactions involved, stress is personal in that individual stress responses change with time and circumstances. Indeed, the effects of stress on each individual are influenced by genetics, SES, environmental context, perception, developmental history, prior susceptibilities, preexisting health status, and individual coping abilities. Clearly, the maintenance of homeostasis requires the human organism to routinely initiate allostatic responses to the stressors of daily life, as well as to the less frequent severe assaults on the integrity of the body and the mind, responsible for allostatic load. Systemically, allostasis may be seen as beginning with some degree of the alarm stage (fight-or-flight activation), and ideally moving to an effective resolution through adaptation, ultimately culminating in a return to homeostasis. The prolonged effects of allostatic overload—the long-term wear-and-tear costs of adaptation efforts—provide a conceptual foundation for examining the long-term consequences of stress on health. What Selye called "diseases of adaptation" are the outcomes of allostatic overload.

Adaptation, Coping, and Resilience

Adaptation broadly refers to the biopsychosocial process of adjusting physiology, morphology, and behavior in response to new or altered circumstances, internal and external in origin, in the physical and social environment. Related to adaptation is the term *resilience.* Resilience is the ability to withstand threats to stability and is a measure of the capacity to adapt to challenges. It has been described as the "ability to bend without breaking in the face of environmental or psychological perturbations."

Because allostasis is a process of attaining and maintaining stability through change that leads to a state of adaptation, the terms *adaptation* and *resilience* have been intertwined with allostasis. Resilience is the capacity to "adapt to stress and adversity," thereby preventing the development of disease and mitigating the effects of allostatic load. It means attaining a positive outcome when encountering adversity. Resilience can be affected by multiple factors, some of which include genetics; epigenetic changes resulting from early life exposure to traumatic events; the availability of social support and coping mechanisms; a sense of optimism and control; prior learning; and lifestyle factors such as physical activity (exercise), diet and nutrition, and adequate sleep. Resilience is influenced by the chemical allostatic mediators that promote adaptation such as cortisol, and some neural proteins such as brain-derived neurotrophic factor (BDNF), which may be necessary for brain plasticity.

Maladaptation, a less frequently used term, refers to ineffective, inadequate, or inappropriate change in response to new or altered circumstances. *Coping* is another term used and is most often seen as a behavioral adaptive response to a stressor. Coping mechanisms are typically culturally based, and consequently vary with the individual within the parameters of what is acceptable to the given culture. The coping behavior is usually dictated by the specific stressor; thus it fluctuates with the circumstances, but individuals typically embrace a specific repertoire of coping behaviors. These behavioral adaptations allow an individual or a group to withstand successfully the stressful experience or the stress response generated by the experience. A coping strategy can be considered effective or functional if it helps resolve either the situation or the feelings. In some cases, such as exercise, mindfulness meditation, and seeking social support, the coping method can promote health. A coping strategy is considered ineffective or dysfunctional if it does not achieve the desired goal. Coping that achieves unintended goals is considered dysfunctional or maladaptive. Adaptation may result in the adoption of less-than-desirable coping behaviors, such as excessive eating or alcohol consumption, smoking, or other types of substance abuse and avoidance behaviors. Unfortunately, these dysfunctional coping behaviors can ultimately be damaging to overall health. Smoking and overeating contribute to atherosclerosis, the underlying pathophysiology of coronary artery disease and a risk factor for hypertension. Excess weight accumulated through overeating is a contributing factor for type 2 diabetes mellitus and metabolic syndrome. Although coping is customarily interpreted as behavioral adaptation only, the terms *coping* and *adapting* often are used interchangeably.

Perception and expectations of the stressor can affect its interpretation, and therefore the behavior evoked by it. Perceptions can be related to uncertainty about the meaning of the stressor. Consider the stressor of undue noise. The "bang" of a car backfiring could also be the sound of a gun being fired. Depending on the environment and circumstances, one or the other etiology would be more expected, dictating different adaptive responses. The term *distress* describes the experience of perceiving an inability to cope with a physiological or psychological stressor. This distress further activates the release of catecholamines from the adrenal gland medulla and glucocorticoids through the HPA axis, escalating levels of circulating mediators, and may exacerbate existing allostatic load and preexisting pathophysiologic conditions. For instance, the person with asthma who is experiencing an episode of acute shortness of breath is likely to become even more short of breath when discovering an inhaler is not readily available.

Adaptation to a stressor can occur in several ways. Loud noise is a known stressor. Yet people who live close to busy airports often reach

a point at which they barely notice the noise of airplanes flying over their homes. They become *habituated* to the stressor (loud noise). One important way to habituate to a stressor is to manipulate, or "train," the hypothalamus to react less forcefully to a perceived threat or stressor. Repeatedly ignoring a specific stressor prevents the inappropriate triggering of the GAS. The result is a more acceptable level of stress response. Techniques that accomplish this desensitization change the predominant brain waves of the individual from beta to alpha waves, which are slower and more normal. Biofeedback, visualization, and meditation are examples of therapies that use this principle. Practicing these techniques for 20 to 30 minutes daily can enhance the ability to alter how a stressor is perceived and modulate the stress response. These techniques have documented efficacy in enhancing immune function. Desensitization methods have been found to be beneficial for common stress-related conditions, such as migraine headache, chronic back pain, and hypertension.

Allostatic Overload and Illness

When adaptation mechanisms are inadequate or the total amount of allostatic load is excessive, overwhelming allostasis capacities, the result is allostatic overload. There are several ways in which allostatic load can accumulate in an individual: (1) repeated exposures to multiple stressors, (2) inability to habituate or adapt to the stressor, (3) unnecessarily prolonged stress response or stress response that continues after the stressor is removed, and (4) inadequate response to the stressor that causes other stress response mediators to attempt to compensate. Homeostasis, the steady-state that previously existed, cannot be attained. Instead, allostatic overload occurs, and the resulting maladaptation can be reflected in a range of pathophysiologic states that span the traditional boundaries of health care, from psychiatric and endocrine disorders to inflammatory disease.

Hair loss, emotional tension, burnout, mouth sores, insomnia, asthma, heart palpitations, neuromuscular movement disorders (tics), tension headaches, muscle contraction backaches, digestive disorders, and irritable bladder are just a few of the common disorders that can be caused by or worsened by stress. Reproductive disorders such as menstrual irregularity in women and male impotence also have been linked with the effects of allostatic overload. Box 2.2 summarizes some of the physiologic and psychological effects of excessive stress. Fig. 2.3 depicts the multiple body organs and systems in which the effects of insufficient or overactive stress responses may be seen.

The chemical mediators of stress have a strong physiologic basis for the role they play in contributing to illness. Cortisol being released from the adrenal cortex supports Selye's stage of resistance or adaptation, but may also be accountable for pathologic changes. The same can be said of the catecholamines and the other chemical mediators (e.g., immune cytokines). Because these bloodborne chemicals have such broad effects systemically, the impact of excessive or inadequate amounts is understandably wide reaching. In some cases, the relationships have been well substantiated by research; in others, they are hypothesized based upon knowledge of the effects of these chemicals.

The relationship between excessive catecholamine levels and what historically have been called "stress-related" illnesses has often been associated with cardiovascular pathologies such as hypertension, stroke, and myocardial infarction. Abdominal fat cells are well supplied with cortisol receptors, and excessive secretion of cortisol results in the collection of fat in this area. When this fat is released into the bloodstream, the resulting increase in the levels of circulating free fatty acids plays a role in cardiovascular risk. Additionally, repeated or prolonged elevation of blood pressure, especially in combination with the metabolic effects of elevated cortisol and catecholamine levels, promotes the development of atherosclerosis and, ultimately, many cardiovascular pathologies. Not only do catecholamines contribute to the development of atherosclerosis and hypertension, they also increase the risk of developing cardiac dysrhythmias and sudden cardiac death, and even stress-induced cardiomyopathy. They increase platelet activity, resulting in clot formation, and elevate serum lipid levels, significant factors in the pathogenesis of myocardial infarction. A growing body of evidence further suggests that inflammation may mediate a link between stress and cardiovascular disease. Stress has been associated with the production of proinflammatory cytokines such as IL-1, IL-6, and tumor necrosis factor (TNF). These cytokines can trigger the production of C-reactive protein (CRP), a cytokine associated with cardiovascular disease.

The field of psychoneuroimmunology has provided substantive evidence of the roles of the stress hormones in the brain. In the central nervous system, specifically the brain, the mediators of adaptation facilitate learning, memory, and neuroendocrine and autonomic regulation. This heightened memory, at least in the short term, allows the individual to be more aware of the potential stressor in the future. Chronic overactivity or underactivity, however, may result in atrophy of some nerve cells (especially in the hippocampus), impairing memory, whereas others have been found to hypertrophy (especially in the amygdala) and undergo remodeling, resulting in an increase in fear, anxiety, and other mood disorders. In essence, allostatic overload results in altered and impaired cognitive function. Some evidence suggests that inflammation associated with allostatic overload may play a role in learning and memory impairment. For instance, elevated levels of proinflammatory cytokines, such as IL-1 and IL-6, markers of

BOX 2.2 Physical and Behavioral Indicators of High Stress

Physical Indicators

Elevated blood pressure
Increased muscle tension
Elevated pulse rate
Increased respiration
Sweaty palms
Cold extremities (hands and feet)
Fatigue
Tension headache
Upset stomach: nausea, vomiting, diarrhea
Change in appetite
Change in weight
Increased blood catecholamine level
Hyperglycemia
Restlessness
Insomnia

Behavioral and Emotional Indicators

Anxiety (nonspecific fears)
Depression
Increased use of mind-altering substances (e.g., alcohol, chemical substances)
Change in eating, sleeping, or activity pattern
Mental exhaustion
Feelings of inadequacy; loss of self-esteem
Increased irritability
Loss of motivation
Decreased productivity
Inability to make good judgments
Inability to concentrate
Increased absenteeism and illness
Increased proneness to accidents

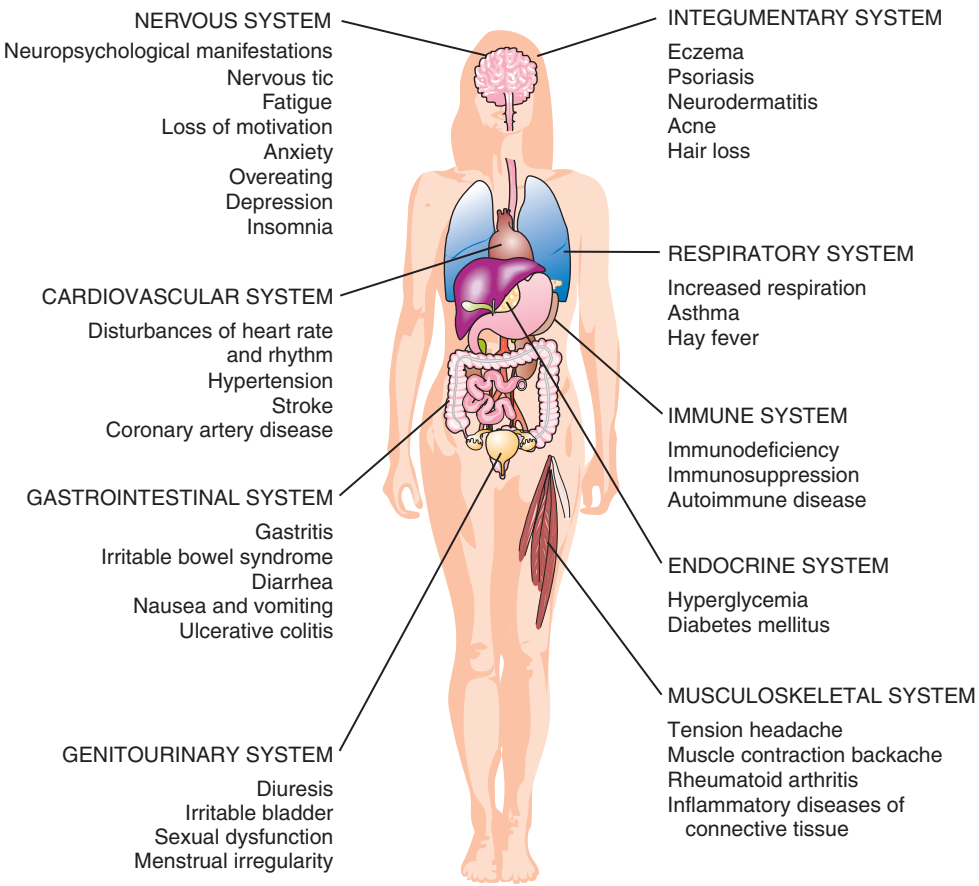


FIG 2.3 Effects of allostatic overload on body organs and systems.

inflammation associated with stress, have been linked with memory loss, cognitive decline, and the development of Alzheimer disease. Emerging research indicates that males and females may show different patterns of brain structure remodeling in response to chronic stress. Estrogen may exert neuroprotective effects.

Stress hormones have been found to be elevated and dysregulated in major depressive illness. Increased cortisol secretion, decreased testosterone levels in men and women, and increased levels of growth hormone and proinflammatory cytokines have been documented in major depressive illness. Depression is common with chronic diseases such as fibromyalgia and multiple sclerosis, and the elevated cortisol levels associated with allostatic overload may be significant to the progression of some of the diseases. Another condition, PTSD, also appears to be associated with heightened sympathetic-adrenal-medullary responses as well as alterations in the HPA axis. Evidence suggests cortisol and norepinephrine help promote long-term memory consolidation and retention of traumatic and fearful events; however, administration of beta-blockers such as propranolol that interfere with the effects of norepinephrine has been shown to reduce the incidence of PTSD symptoms in many cases.

Allostatic mediators activate and maintain energy reserves, which are initially meant to be helpful in managing stressors. Nonetheless, obesity (especially abdominal obesity), diabetes, atherosclerosis, metabolic syndrome, osteoporosis and bone demineralization, and other diseases, as well as accelerated aging, are associated with their chronic activation. The food-seeking behavior initiated by cortisol is beneficial in the short term, but when cortisol levels are increased by chronic stress of either a physiologic or a psychological origin, this adaptation gone awry results in obesity. Obesity is a risk factor for decreased effectiveness of glucose transport into the cells (insulin resistance), the pathophysiologic basis for type 2 diabetes. Elevated cortisol levels also directly increase insulin resistance. Additionally, obesity is associated with the production of proinflammatory cytokines such as TNF-α, IL-1, and IL-6, which also have been connected to diabetes.

In acute stress, activation of the immune system allows for the coordinated defense of the body from damage. At first, leukocytosis and immune function, including phagocytosis and antibody production, may be enhanced to protect the body from foreign invaders (e.g., bacteria and viruses), but then is followed quickly by immunosuppression. Chronic activation of the stress mediators produces immunosuppression and increases the risk of infection and has been implicated in the development of autoimmune diseases and some cancers. Such overactivation also prolongs existing infections and the development of secondary infections. Research supports the hypothesis that physical and emotional stress and dysfunctional coping mechanisms impair both antibody and T-cell–mediated responses to viruses and antiviral and antibacterial vaccines. Stressors of more than 1 month's duration have been found to be the greatest predictors of the development of colds. Cumulated evidence, in both human and animal models, supports the premise that stress-induced dysregulation of the cellular and humoral arms of the immune system increases risk of infectious disease. Stress has been found by numerous studies to accelerate the progression of HIV infection.

Immune dysregulation can also include the excessive production of cytokines that promote the inflammatory response. Both physical and psychological stressors have been found to accomplish this, sensitizing the overall inflammatory response so that subsequent activations are markedly increased. Early life stress such as low SES and childhood

abuse and neglect have been associated with signs of inflammation, including elevated levels of IL-6, TNF-α, and CRP later in adulthood and older age. Inflammation may partially explain the relationship between childhood stress and some later-life illness conditions, including cardiovascular disease, type 2 diabetes, and some forms of cancers. Many other diseases also are associated with chronic inflammation: Alzheimer disease, chronic inflammatory bowel disease, depression, PTSD, osteoporosis, rheumatoid arthritis, asthma, and periodontal disease, to name but a few. Wound healing also is impaired by multiple mediators of stress in excessive amounts.

Epigenetics is emerging as a new area of study, with epigenetic changes providing a link between some types of stressors and disease pathologies. Epigenetics refers to modifications in gene activity that do not entail altering the basic DNA sequence. These modifications may involve chemical tags or markers on the DNA that can turn a gene on or off. Examples of the mechanisms include DNA methylation, histone modification, and microribonucleic acid (miRNA) gene expression regulation. For instance, adverse childhood events have been associated with DNA methylation in adults and consequent production of IL-6. It is possible that epigenetic changes acquired as a result of early life adverse experiences such as famine or poor maternal attachment and care may make the person susceptible later as an adult to respond to stressors in ways that are adaptive in a threatening environment but maladaptive in a more friendly situation. As an example, a person prone to responding to stressors with a heightened sense of vigilance and arousal may find this response helpful in the dangerous situation, but it may lead to anxiety and depression when repeatedly used in a safer environment.

Another new area of stress research attracting attention pertains to telomeres and telomerase. Telomeres are the tail ends of chromosomes that get shaved down with repeated cell division; and thus older cells tend to have shorter telomeres than younger ones. These cells with shortened telomeres are more susceptible to death. Telomeres are both markers and mechanisms of biological aging and may serve as a means of measuring a person's total accumulated exposure to stressors. Chronic stress related to caregiving, lower SES, and number of childhood adversities has been linked with shorter telomere length. Depression and anxiety and several other diseases (e.g., cardiovascular disease, diabetes, and cancer) also have been associated with shortened telomeres. This research suggests a mechanism by which stress may contribute to cell death and disease, because telomere shortening may be connected to some extent to elevated cortisol, catecholamine, and inflammatory cytokine levels produced as part of the stress response. On the other hand, telomerase is an enzyme capable of lengthening telomeres and is inversely related to perceptions of stress. In one study, 30 men and women took part in a 3-month meditation retreat program aimed at reducing psychological distress. By the end of the study period, the participants had significantly higher telomerase activity levels in comparison to wait-list controls. In another study, adherence to comprehensive lifestyle changes involving stress management, aerobic exercise, nutrition, and social support were associated with greater telomere length at 5 years after study initiation in men with prostate cancer in comparison to a control group. More research is needed to understand the relationships among telomeres, telomerase, stress, stress-related diseases, and coping methods, as well as aging and longevity.

KEY POINTS

- Adaptation, or allostasis, is a network of biopsychosocial processes of responding to a stressor with the goal of reestablishing homeostasis. Coping mechanisms are usually seen as behavioral adaptations to stress, but are often used interchangeably with adaptation; they can contribute to resilience.
- The wear-and-tear effect of adaptation on the body and mind is the allostatic load. It occurs as mediators produced by the stress response systems accumulate and contribute to tissue damage over time. Allostatic load reflects the cumulative costs of adaptation.
- A number of disorders are thought to be related to excessive stress or inappropriate stress responses—allostatic overload. These are a result of the dysregulation and excessive use of the mechanisms and mediators involved in the stress response.

SUMMARY

Homeostasis is the state of balance of the body's biopsychosocial systems. Stressors evoke a stress response and initiate adaptive efforts, an allostatic process, designed to return to this steady-state. The response to stressors is affected by a wide variety of factors. Recently there has been an increase in knowledge regarding the complex interactions of the HPA axis, the SNS, the immune system, genetics and epigenetics, and the chemical mediators of the stress response.

Excessive or prolonged stress and overactivity or underactivity of associated chemical mediators produce disproportionate responses in the body, a condition of allostatic overload known as *stress-induced illness*. As humans strive to adapt to the constant changes of modern life, the study of stress and stress-related disease has become vital to public health and contributes to the development of increasingly sophisticated models of health and illness.

RESOURCES

Homeostasis and Allostasis

Baffy G, Loscalzo J: Complexity and network dynamics in physiological adaptation. *Physiol Behav* 131:49–56, 2014.

Beckie TM: A systematic review of allostatic load, health, and health disparities. *Biol Res Nurs* 14:311–346, 2012.

Brody GH, Lei MK, Chen E, Miller GE: Neighborhood poverty and allostatic load in African American youth. *Pediatrics* 134:1362–1368, 2014.

Cannon WB: *The wisdom of the body*, New York, 1932, W. W. Norton & Company.

Cuesta JM, Singer M: The stress response and critical illness: a review. *Crit Care Med* 40:3283–3289, 2012.

Danese A, McEwen BS: Adverse childhood experiences, allostasis, allostatic load, and age-related disease. *Physiol Behav* 106:29–39, 2012.

Goldstein DS: Concepts of scientific integrative medicine applied to the physiology and pathophysiology of catecholamine systems. *Compr Physiol* 3:1569–1610, 2013.

Karatsoreos IN, McEwen BS: Annual research review: the neurobiology and physiology of resilience and adaptation across the life course. *J Child Psychol Psychiatry* 54:337–347, 2013.

Lovallo WR: *Stress & health*, ed 3, Los Angeles, 2016, Sage.

McEwen BS: Interacting mediators of allostasis and allostatic load: towards an understanding of resilience in aging. *Metabolism* 52:10–16, 2003.
McEwen BS: Biomarkers for assessing population and individual health and disease related to stress and adaptation. *Metabolism* 64:S2–S10, 2015.
McEwen BS, Gray J, Nasca C: Recognizing resilience: learning from the effects of stress on the brain. *Neurobiol Stress* 1:1–11, 2015.
Picard M, Juster R, McEwen BS: Mitochondrial allostatic load puts the 'gluc' back in glucocorticoids. *Nat Rev Endocrinol* 10:303–310, 2014.
Ramsay DS, Woods SC: Clarifying the roles of homeostasis and allostasis in physiological regulation. *Psychol Rev* 121:225–247, 2014.
Sapolsky RM: *Why zebras don't get ulcers*, ed 3, New York, 2004, Henry Holt.
Sinha R, Jastreboff AM: Stress as a common risk factor for obesity and addiction. *Biol Psychiatry* 73:827–835, 2013.
Sterling P: Allostasis: a model of predictive regulation. *Physiol Behav* 106:5–15, 2012.
Sterling P: Homeostasis vs allostasis: implications for brain function and mental disorders. *JAMA Psychiatry* 71:1192–1193, 2014.
Sterling P, Eyer J: Allostasis: a new paradigm to explain arousal pathology. In Fisher S, Reason J, editors: *Handbook of life stress, cognition, and health*, New York, 1988, John Wiley & Sons, pp 629–649.
Tomasdottir MO, Sigurdsson JA, Petursson H, et al: Self reported childhood difficulties, adult multimorbidity and allostatic load. A cross-sectional analysis of the Norwegian HUNT study. *PLoS ONE* 10:e0130591, 2015.

Neurohormones and Concepts of Stress

Altemus M, Sarvaiya N, Neill Epperson C: Sex differences in anxiety and depression clinical perspectives. *Front Neuroendocrinol* 35:320–330, 2014.
Bangasser DA, Valentino RJ: Sex differences in stress-related psychiatric disorders: neurobiological perspectives. *Front Neuroendocrinol* 35:303–319, 2014.
Bystritsky A, Kronemyer D: Stress and anxiety: counterpart elements of the stress/anxiety complex. *Psychiatr Clin North Am* 37:489–518, 2014.
Choi JC, Chung MI, Lee YD: Modulation of pain sensation by stress-related testosterone and cortisol. *Anaesthesia* 67:1146–1151, 2012.
Cohen S, Janicki-Deverts D, Doyle WJ, et al: Chronic stress, glucocorticoid receptor resistance, inflammation, and disease risk. *Proc Natl Acad Sci USA* 109:5995–5999, 2012.
Costanza M, Binart N, Steinman L, Pedotti R: Prolactin: a versatile regulator of inflammation and autoimmune pathology. *Autoimmun Rev* 14:223–230, 2015.
Davidson MW: The endorphin collection, Florida State University. Available at http://micro.magnet.fsu.edu/micro/gallery/endorphin/endorphins.html.
Dhabhar FS: Pscyhological stress and immunoprotection versus immunopathology in the skin. *Clin Dermatol* 31:18–30, 2013.
Dieleman GC, Huizink AC, Tulen JH, et al: Alterations in HPA-axis and autonomic nervous system functioning in childhood anxiety disorders point to a chronic stress hypothesis. *Psychoneuroendocrinology* 51:135–150, 2015.
Doamekpor LA, Dinwiddie GY: Allostatic load in foreign-born and US-born blacks: evidence from the 2001-2010 national health and nutrition examination survey. *Am J Public Health* 105:591–597, 2015.
Doom JR, Gunnar MR: Stress physiology and developmental psychopathology: past, present, and future. *Dev Psychopathol* 25:1359–1373, 2013.
Evans JF, Ragolia L: Systemic and local ACTH produced during inflammatory states promotes osteochondrogenic mesenchymal cell differentiation contributing to the pathologic progression of calcified atherosclerosis. *Med Hypotheses* 79:823–826, 2012.
Fagundes CP, Way B: Early-life stress and adult inflammation. *Curr Dir Psychol Sci* 23:277–283, 2014.
Faron-Gorecka A, Kusmider M, Solich J, et al: Involvement of prolactin and somatostatin in depression and the mechanism of action of antidepressant drugs. *Pharmacol Rep* 65:1640–1646, 2013.
Frankenhaeuser M: The sympathetic-adrenal and the pituitary-adrenal response to challenge: comparison between the sexes. In Dembroski TM, Smidt TH, Blumchen G, editors: *Biobehavioral bases of coronary heart disease*, New York, 1983, Karger, pp 91–105.
Groer M: Allostasis: a model for women's health. In Kendall-Tacket K, editor: *The psychoneuroimmunology of chronic disease*, Washington, DC, 2010, American Psychological Association, pp 183–218.
Hannibal KE, Bishop MD: Chronic stress, cortisol dysfunction, and pain: a psychoneuroendocrine rationale for stress management in pain rehabilitation. *Phys Ther* 94:1816–1825, 2014.
Hui LY, Wang YW, Zhou FL, et al: Association between MKP-1, BDNF, and gonadal hormones with depression on perimenopausal women. *J Womens Health* 00:1–7, 2015.
Hunter RG, McEwen BS: Stress and anxiety across the lifespan: structural plasticity and epigenetic regulation. *Epigenomics* 5:177–194, 2013.
Ingawale DK, Mandlik SK, Patel SS: An emphasis on molecular mechanisms of anti-inflammatory effects and glucocorticoid resistance. *J Complement Integr Med* 12:1–13, 2015.
Jobin J, Wrosch C, Scheier MF: Associations between dispositional optimism and diurnal cortisol in a community sample: when stress is perceived as higher than normal. *Health Psychol* 33:382–391, 2014.
Karatsoreos IN, McEwen BS: Resilience and vulnerability: a neurobiological perspective. *F1000Prime Rep* 5:13, 2013.
Kaur K, Maas D: Hypothalamic-pituitary axis. In Benjamin IJ, editor: *Andreoli and Carpenter's Cecil essentials of medicine*, ed 9, Philadelphia, 2016, Elsevier, pp 624–632.
Lavretsky H, Newhouse PA: Stress, inflammation, and aging. *Am J Geiatr Psychiatry* 20:729–733, 2012.
Lee Y, IM EO: Stress and premenstrual symptoms in reproductive-aged women. *Health Care Women Int* 19:1–25, 2015.
McEwen BS: Brain on stress: how the social environment gets under the skin. *Proc Natl Acad Sci USA* 109:17180–17185, 2012.
McEwen BS: Sex, stress and the brain: interactive actions of hormones on the developing and adult brain. *Climacteric* 17:18–25, 2014.
McEwen BS, Morrison JH: The brain on stress: vulnerability and plasticity of the prefrontal cortex over the life course. *Neuron* 79:16–29, 2013.
Molitch ME: Anterior pituitary. In Goldman L, Schafer AI, editors: *Goldman's Cecil medicine*, ed 25, Philadelphia, 2016, Elsevier, pp 1480–1494.
Naber M, Frassle S, Rutishauser U, Einhauser W: Pupil size signals novelty and predicts later retrieval success for declarative memories of natural scenes. *J Vis* 13:11, 2013.
Ninkovic J, Roy S: Role of the mu-opioid receptor in opioid modulation of immune function. *Amino Acids* 45:9–24, 2013.
Opacka-Juffry J, Mohiyeddini C: Experience of stress in childhood negatively correlates with plasma oxytocin concentration in adult men. *Stress* 15:1–10, 2012.
Panagiotakopoulos L, Neigh GN: Development of the HPA axis: where and when do sex differences manifest? *Front Neuroendocrinol* 35:285–302, 2014.
Pereira Suarez AL, Lopez-Rincon G, Martinez Neri PA, Estrada-Chavez C: Prolactin in inflammatory response. *Adv Exp Med Biol* 846:243–264, 2015.
Pongratz G, Straub RH: The sympathetic nervous response in inflammation. *Arthritis Res Ther* 16:504, 2014.
Reul JMHM, Collins A, Saliba RS, et al: Glucocorticoids, epigenetic control and stress resilience. *Neurobiol Stress* 1:44–59, 2015.
Richardson CM, Rice KG, Devine DP: Perfectionism, emotion regulation, and the cortisol stress response. *J Couns Psychol* 61:110–118, 2014.
Rohleder N: Stimulation of systemic low-grade inflammation by psychosocial stress. *Psychosom Med* 76:181–189, 2014.
Roney JR, Simmons ZL: Elevated psychological stress predicts reduced estradiol concentrations in young women. *Adapt Human Behav Physiol* 1:30–40, 2015.
Seltzer LJ, Prososki AR, Ziegler TE, Pollak SD: Instant messages vs. speech: hormones and why we still need to hear each other. *Evol Hum Behav* 33:42–45, 2012.
Selye H: *The stress of life*, New York, 1984, McGraw-Hill.
Skoluda N, Strahler J, Schlotz W, et al: Intra-individual psychological and physiological responses to acute laboratory stressors of different intensity. *Psychoneuroendocrinology* 51:227–236, 2015.
Szabo S, Tache Y, Somogyi A: The legacy of Hans Selye and the origins of stress research: a retrospective 75 years after his landmark brief "letter" to the editor of *Nature*. *Stress* 15:472–478, 2012.

Taylor SE, Master SL: Social responses to stress: the tend-and-befriend model. In Contrada RJ, Baum A, editors: *The handbook of stress science*, New York, 2011, Springer, pp 101–109.

Valentino RJ, Van Bockstaele E: Endogenous opioids: the downside of opposing stress. *Neurobiol Stress* 1:23–32, 2015.

Wei J, Yuen EY, Liu W, et al: Estrogen protects against the detrimental effects of repeated stress on glutamatergic transmission and cognition. *Mol Psychiatry* 19:588–598, 2014.

Weissman JF, Pratt LA, Miller EA, Parker JD: Serious psychological distress among adults: United States, 2009-2013. *NCHS Data Brief* (203):1–8, 2015.

Adaptation and Coping

Aschbacher K, Kornfeld S, Picard M, et al: Chronic stress increases vulnerability to diet-related abdominal fat, oxidative stress, and metabolic risk. *Psychoneuroendocrinology* 46:14–22, 2014.

Aschbacher K, O'Donovan A, Wolkowitz OM, et al: Good stress, bad stress and oxidative stress: insights from anticipatory cortisol reactivity. *Psychoneuroendocrinology* 38:1698–1708, 2013.

Cao Y, Lu L, Liu M, et al: Impact of epigenetics in the 84. management of cardiovascular disease: a review. *Eur Rev Med Pharmacol Sci* 18:3097–3104, 2014.

Debono M, Prema A, Hughes TJ, et al: Visceral fat accumulation and postdexamethasone serum cortisol levels in patients with adrenal incidentaloma. *J Clin Endocrinol Metab* 98:2383–2391, 2013.

Eng JWL, Kokolus KM, Reed CB, et al: A nervous tumor microenvironment: the impact of adrenergic stress on cancer cells, immunosuppression, and immunotherapeutic response. *Cancer Immunol Immunother* 63:1115–1128, 2014.

Epel ES, Lithgow GJ: Stress biology and aging mechanisms: toward understanding the deep connection between adaptation to stress and longevity. *J Gerontol A Biol Sci Med Sci* 69:S10–S16, 2014.

Esser N, Legrand-Poels S, Piette J, et al: Inflammation as a link between obesity, metabolic syndrome and type 2 diabetes. *Diabetes Res Clin Pract* 105:141–150, 2014.

Jacobs TL, Epel ES, Lin J, et al: Intensive meditation training, immune cell telomerase activity, and psychological mediators. *Psychoneuroendocrinology* 36:664–681, 2011.

Lindgren ME, Fagundes CP, Alfano CM, et al: Beta-blockers may reduce intrusive thoughts in newly diagnosed cancer patients. *Psychooncology* 22:1889–1894, 2013.

McEwen BS, Gianaros PJ: Stress- and allostasis-induced brain plasticity. *Annu Rev Med* 62:431–445, 2011.

McEwen BS, Karatsoreos IN: Sleep deprivation and circadian disruption: stress, allostasis, and allostatic load. *Sleep Med Clin* 10:1–10, 2015.

McVicar A, Ravalier JM, Greenwood C: Biology of stress revisited: intracellular mechanisms and the conceptualization of stress. *Stress Health* 30:272–279, 2014.

Nasca C, Bigio B, Zelli D, et al: Mind the gap: glucocorticoids modulate hippocampal glutamate tone underlying individual differences in stress susceptibility. *Mol Psychiatry* 20:755–763, 2015.

Oglodek E, Szota A, Just M, et al: The role of the neuroendocrine and immune systems in the pathogenesis of depression. *Pharmacol Rep* 66:776–781, 2014.

Okereke OI, Prescott J, Wong JY, et al: High phobic anxiety is related to lower leukocyte telomere length in women. *PLoS ONE* 7:e40516, 2012.

Ornish D, Lin J, Chan JM, et al: Effect of comprehensive lifestyle changes on telomerase activity and telomere length in men with biopsy-proven low-risk prostate cancer: 5-year follow-up of a descriptive pilot study. *Lancet Oncol* 14:1112–1120, 2013.

Ridker PM, Luscher TF: Anti-inflammatory therapies for cardiovascular disease. *Eur Heart J* 35:1782–1791, 2014.

Starkweather AR, Alhaeeri AA, Montpetit A, et al: An integrative review of factors associated with telomere length and implications for biobehavioral research. *Nurs Res* 63:36–50, 2014.

Steiger A, Dresler M, Kluge M, Schussler P: Pathology of sleep, hormones and depression. *Pharmacopsychiatry* 46:S30–S35, 2013.

Steptoe A, Kivimaki M: Stress and cardiovascular disease: an update on current knowledge. *Annu Rev Public Health* 34:337–354, 2013.

Tait EM, Laditka SB, Laditka JN, et al: Use of complementary and alternative medicine for physical performance, energy, immune function, and general health among older women and men in the United States. *J Women Aging* 24:23–43, 2012.

Tomiyama AJ, O'Donovan A, Lin J, et al: Does cellular aging relate to patterns of allostasis? An examination of basal and stress reactive HPA axis activity and telomere length. *Physiol Behav* 106:40–45, 2012.

Wang DS, Zurek AA, Lecker I, et al: Memory deficits induced by inflammation are regulated by α5-subunit-containing GABAA receptors. *Cell Rep* 2:488–496, 2012.

Zschucke E, Renneberg B, Dimeo F, et al: The stress-buffering effect of acute exercise: evidence for HPA axis negative feedback. *Psychoneuroendocrinology* 51:414–425, 2015.

3

Cell Structure and Function

Jacquelyn L. Banasik and Brianne N. Banasik

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- What are the major cellular structures and their functions?
- How do cells acquire and use energy?
- How are substances transported across the cell membrane?
- Why is it that some cells can produce action potentials and others cannot?
- How do cells in a multicellular organism communicate with one another?
- What are the normal mechanisms of cellular growth control?

CHAPTER OUTLINE

Plasma Membrane, 27
Membrane Structure, 27
Lipid Bilayer, 27
Membrane Proteins, 29
Organization of Cellular Compartments, 30
Cytoskeleton, 30
Nucleus, 30
Endoplasmic Reticulum, 31
Golgi Apparatus, 32
Lysosomes and Peroxisomes, 33
Mitochondria, 34
Cellular Metabolism, 34
Glycolysis, 36
Citric Acid Cycle, 36
Oxidative Phosphorylation, 36
Functions of the Plasma Membrane, 39
Membrane Transport of Macromolecules, 39
Endocytosis and Exocytosis, 39
Membrane Transport of Small Molecules, 40
Active Transport Pumps, 41
Membrane Transport Carriers, 42
Membrane Channel Proteins, 44
Cellular Membrane Potentials, 45
Resting Membrane Potential, 45
Action Potential, 45
Intercellular Communication and Growth, 48
Cell Signaling Strategies, 48
Cell Surface Receptor–Mediated Responses, 49
Intracellular Receptor–Mediated Responses, 53
Regulation of Cellular Growth and Proliferation, 54

The cell is the fundamental unit of life. As more diseases are understood on the cellular and molecular levels, it appears that the cell is also the fundamental unit of disease. Detailed knowledge of cellular dysfunction has led to the development of more specific and appropriate prevention and treatment modalities for many disease processes. Thus an understanding of cellular mechanisms is essential for health care providers and fundamental to the discussions of pathophysiologic processes presented throughout the remainder of this text.

Cells are complex, membrane-bound units packed with a multitude of chemicals and macromolecules. They are able to replicate and thus form new cells and organisms. The first cells on Earth probably arose about 3.5 billion years ago. Over billions of years, the self-replicating molecules now known as *deoxyribonucleic acid (DNA)* and *ribonucleic acid (RNA)* are believed to have evolved by chance association and natural selection. Development of the cell membrane created a closed compartment that provided a selective advantage for the cell and

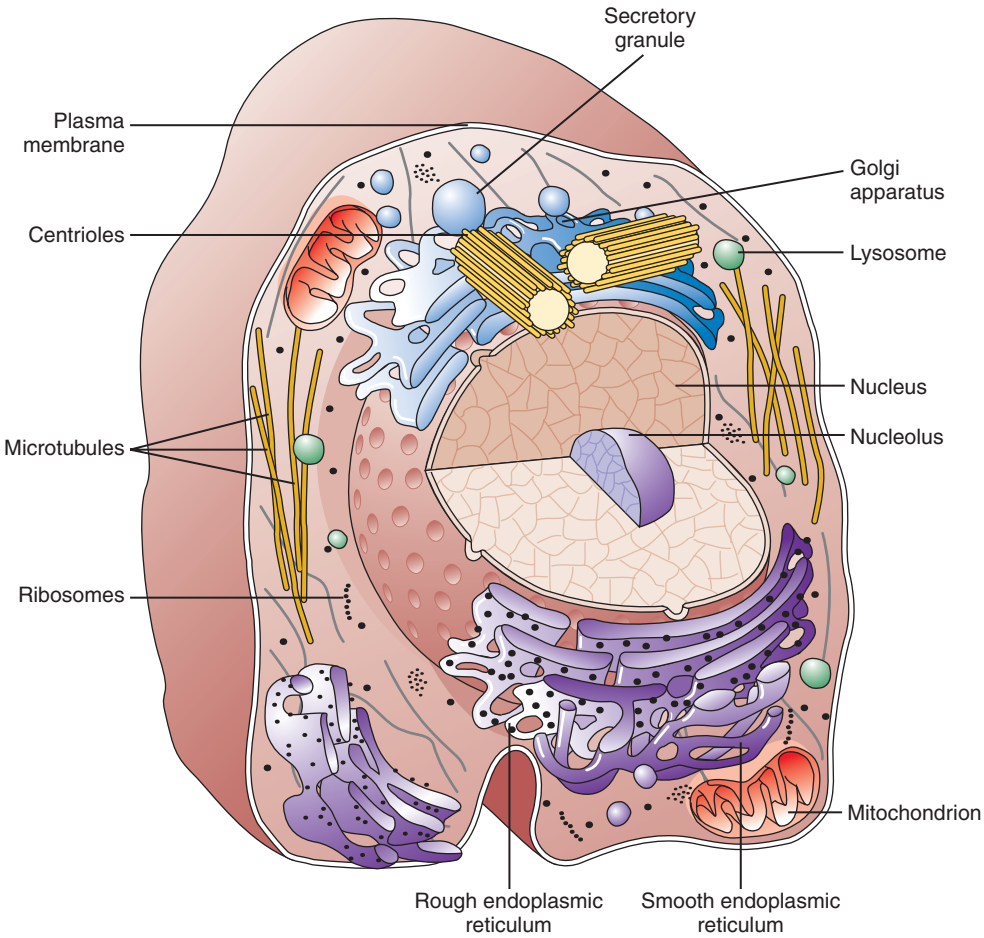


FIG 3.1 Structure of a typical eukaryotic cell showing intracellular organelles.

accomplished the first separation of life (inside) from nonlife (outside). In this protected environment, the early cells continued to evolve and develop. Today, a large number of different cell types exist, but many of the basic biochemical mechanisms of these cells are remarkably similar. Scientists believe that all modern cells, from bacteria to human neurons, evolved from common primordial cells. It is therefore possible to unlock many of the secrets of human cellular physiology by studying easily grown and rapidly proliferating cells, such as yeasts and bacteria.

Much of our knowledge of cell physiology has derived from study of the class of cells known as *prokaryotic,* which includes bacteria and archaea. Prokaryotic cells are smaller and simpler than eukaryotic cells, having no defined nucleus or cytoplasmic organelles. Fungi, plants, and animals belong to the eukaryotic class of cells, which possess a membrane-bound nucleus and a host of cytoplasmic organelles (Fig. 3.1). In this chapter, the essentials of eukaryotic cell structure, physiology, metabolism, and communication are reviewed.

PLASMA MEMBRANE

Membrane Structure

All cells are enclosed by a barrier composed primarily of lipid and protein called the *plasma membrane* (plasmalemma). This cell membrane is a highly selective filter that shields internal cell contents from the external environment. The plasma membrane performs a variety of functions, including transport of nutrients and waste products; generation of membrane potentials; and recognition, communication, and growth regulation of cells. The cell membrane is a sensor of signals and enables the cell to respond and adapt to changes in its environment.

According to the fluid mosaic model first described in the 1960s by Singer and Nicolson, the plasma membrane is a dynamic assembly of lipid and protein molecules. Most of the lipids and proteins move about rapidly in the fluid structure of the membrane. As shown in Fig. 3.2, the lipid molecules are arranged in a double layer, or lipid bilayer, which is highly impermeable to most water-soluble molecules, including ions, glucose, and proteins. A variety of proteins embedded, or "dissolved," in the lipid bilayer perform most of the membrane's functions. Some membrane proteins are involved in the transport of specific molecules into and out of the cell; others function as enzymes or respond to external signals; and some serve as structural links that connect the plasma membrane to adjacent cells. The lipid structure of the plasma membrane is similar to the structure of the membrane that surrounds the cell's organelles (e.g., nucleus, mitochondria, endoplasmic reticulum, Golgi apparatus, lysosomes).

Lipid Bilayer

The bilayer structure of all biological membranes is related to the special properties of lipid molecules that cause them to spontaneously assemble into bilayers. The three major types of membrane lipids are cholesterol, phospholipids, and glycolipids. All three have a molecular structure that is amphiphilic; that is, they have a hydrophilic (water-loving) charged or polar end and a hydrophobic (water-fearing) nonpolar end. This amphiphilic nature causes the lipids to form bilayers in aqueous solution.

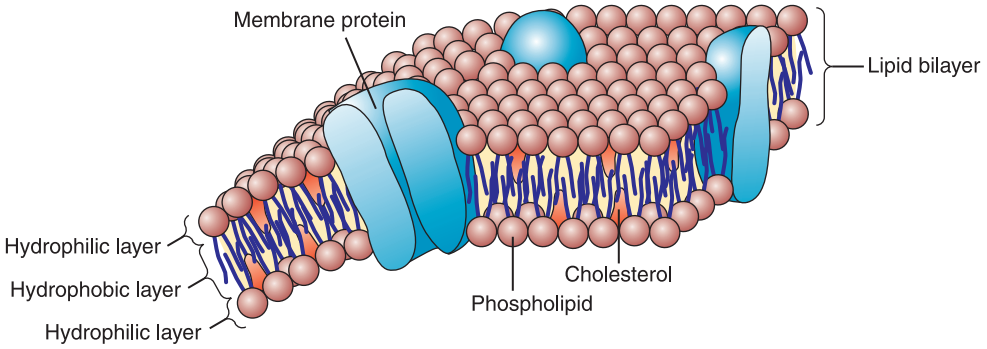


FIG 3.2 Section of the cell membrane showing the lipid bilayer structure and integral membrane proteins.

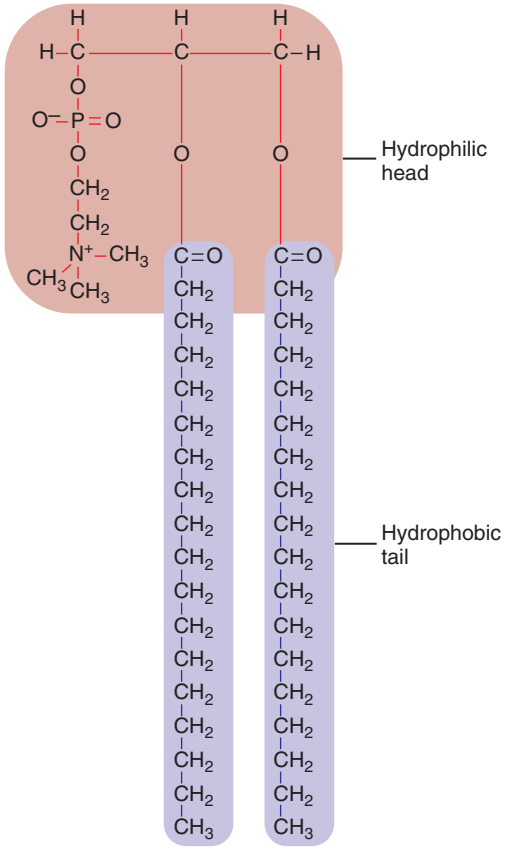


FIG 3.3 Schematic drawing of a typical membrane phospholipid molecule showing the amphiphilic nature of the structure.

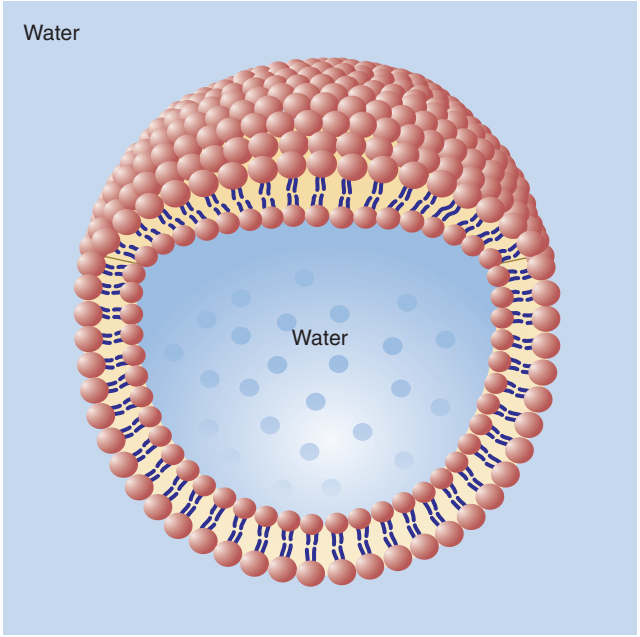


FIG 3.4 The amphiphilic nature of membrane lipids results in bilayer structures that tend to form spheres.

A typical phospholipid molecule is shown in Fig. 3.3. The hydrophobic nonpolar tails tend to associate with other hydrophobic nonpolar tail groups to avoid association with polar water molecules. The hydrophilic polar head groups preferentially interact with the surrounding aqueous environment. A bilayer, with tails sandwiched in the middle, allows both portions of the lipid molecules to be chemically "satisfied." In addition, the lipid bilayers tend to close on themselves, forming sealed, spherical compartments (Fig. 3.4). If the membrane is punctured or torn, it will spontaneously reseal itself to eliminate contact of the hydrophobic tails with water.

For the most part, individual lipid and protein molecules can diffuse freely and rapidly within the plane of the bilayer. The degree of membrane fluidity depends on the lipid composition. Saturated lipids have straight tails that can pack together and tend to stiffen the membrane, whereas lipids with bent, unsaturated hydrocarbon tails tend to increase fluidity. About 50% of the lipid in eukaryotic cell membranes is cholesterol, which serves to decrease membrane permeability and prevent leakage of small water-soluble molecules. In addition to affecting fluidity by the degree of saturation of tail groups, the phospholipids that inhabit the membrane differ in the size, shape, and charge of the polar head groups. Fig. 3.5 shows the structures of the four most prevalent membrane phospholipids: phosphatidylethanolamine, phosphatidylserine, phosphatidylcholine, and sphingomyelin. Some membrane-bound proteins require specific phospholipid head groups to function properly. Some lipids—sphingolipids and cholesterol in particular—may bind together transiently to form rafts in the sea of moving lipids. These rafts may surround and help organize membrane proteins into functional units. For example, a membrane receptor and its intracellular target proteins may associate together in a raft to facilitate transfer of information across the membrane.

Glycolipids contain one or more sugar (i.e., carbohydrate) molecules at the polar head region. Glycolipids and glycoproteins are found only in the outer half of the lipid bilayer, with the sugar groups exposed at the cell surface (Fig. 3.6). Membrane glycolipids are involved in cell recognition and cell-to-cell interactions.

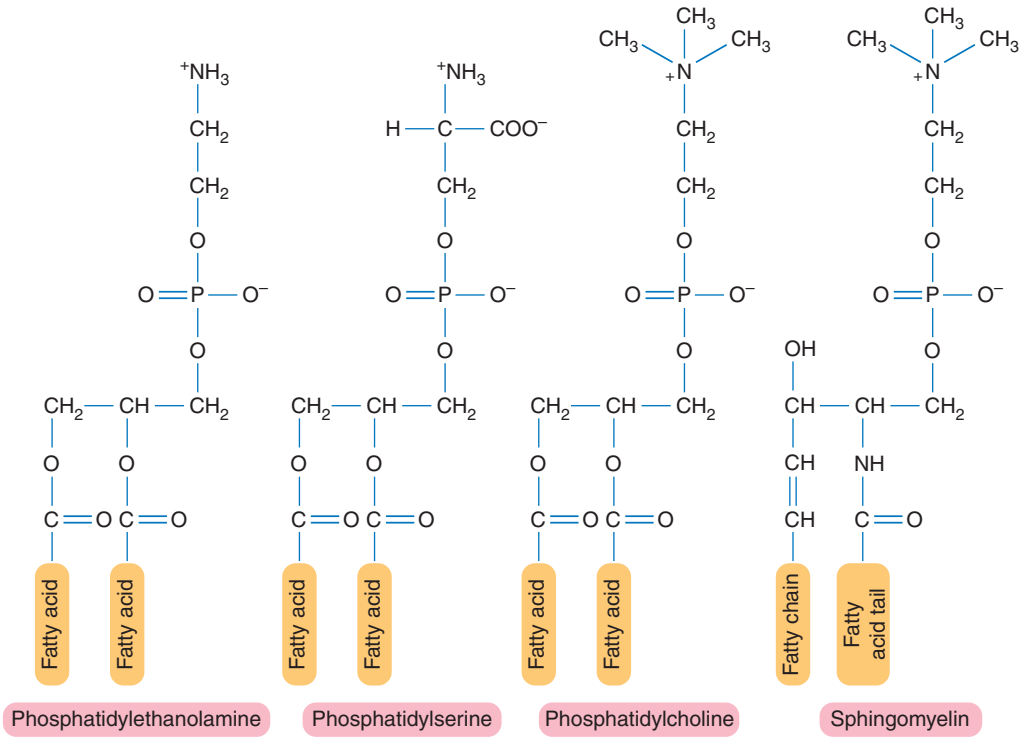


FIG 3.5 Chemical structures of the four most common membrane phospholipids.

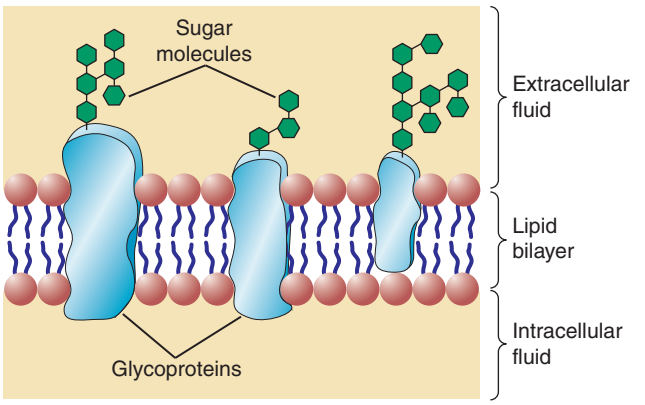


FIG 3.6 Portion of the cell membrane showing orientation of membrane glycoproteins toward the outer surface of the cell.

Membrane Proteins

Approximately 50% of the mass of a typical cell membrane is composed of protein. The specific types of membrane proteins vary according to cell type and environmental conditions. Some membrane proteins, called *transmembrane proteins,* extend across the membrane bilayer and are in contact with both the extracellular and the intracellular fluids. Transmembrane proteins serve a variety of functions, including transport of charged and polar molecules into and out of cells and transduction of extracellular signals into intracellular messages. Other peripheral membrane proteins are less tightly anchored to the membrane. The common structural orientations of membrane proteins are shown in Fig. 3.7. The amino acid structure of membrane proteins determines the way they are arranged in the membrane. Nonpolar amino acids tend to inhabit the hydrophobic middle of the membrane, whereas charged and polar amino acids protrude into the aqueous fluid or associate with polar lipid head groups. The three-dimensional structure of many membrane proteins is complex, with numerous twists and turns through the lipid bilayer (Fig. 3.8).

The type of membrane proteins in a particular cell depends on the cell's primary functions. For example, a kidney tubule cell has a large proportion of transmembrane proteins, which are needed to perform the kidney's function of electrolyte and nutrient reabsorption. In contrast, the human red blood cell (RBC) contains mainly peripheral proteins attached to the inner surface of the membrane. One of these proteins, spectrin, has a long, thin, flexible, rodlike shape that forms a supportive meshwork, or cytoskeleton, for the cell. It is this cytoskeleton that enables the RBC to withstand the membrane stress of being forced through small capillaries.

Although proteins and lipids are generally free to move within the plane of the cell membrane, many cells are able to confine certain proteins to specific areas. Using the example of the kidney tubule cell again, it is important for the cell to keep transport proteins on its luminal side to reabsorb filtered molecules (Fig. 3.9). This segregation of particular proteins is accomplished primarily by intercellular connections called tight junctions, which connect neighboring cells and function like a fence to confine proteins to an area of the membrane. Membrane proteins also can be immobilized by tethering them to cytoskeleton or extracellular matrix structures.

KEY POINTS

- The plasma membrane is composed of a lipid bilayer that is impermeable to most water-soluble molecules, including ions, glucose, and amino acids, but is permeable to lipid-soluble substances, such as oxygen and steroid hormones.
- Proteins embedded in the lipid bilayer execute most of the membrane's functions, including transport and signal transduction.

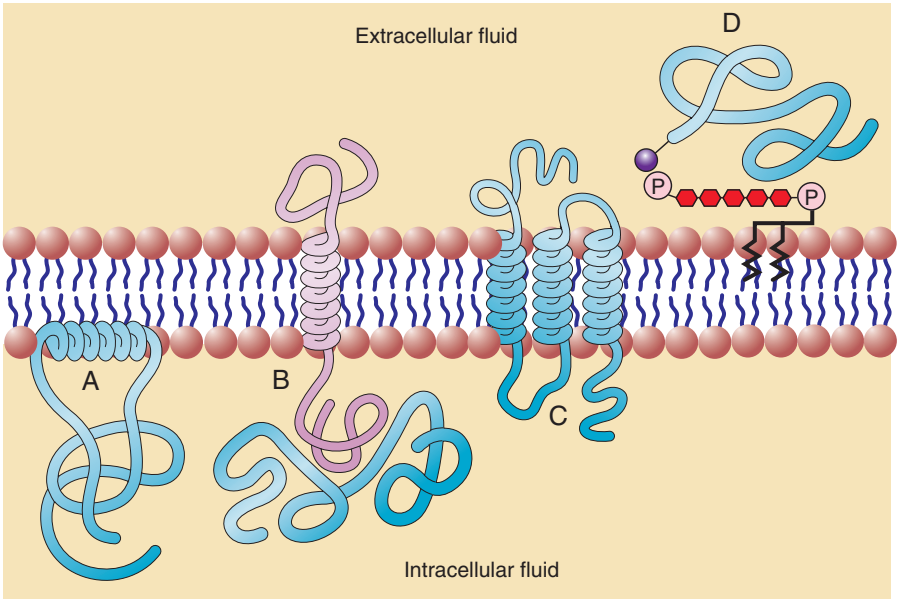


FIG 3.7 Structural orientation of some proteins in the cell membrane. **A,** Membrane-associated protein with noncovalent attachment to plasma lipids. **B,** Membrane protein with noncovalent attachment to another membrane protein. **C,** Transmembrane protein extending through the lipid bilayer. **D,** Covalently attached peripheral membrane protein.

ORGANIZATION OF CELLULAR COMPARTMENTS

Cytoskeleton

Eukaryotic cells have a variety of internal compartments, or organelles, that are membrane bound and carry out distinct cellular functions. The cell's organelles are not free to float around haphazardly in the cytoplasmic "soup"; rather, they are elaborately organized by a protein network called the *cytoskeleton* (Fig. 3.10). The cytoskeleton maintains the cell's shape, allows cell movement, and directs the trafficking of substances within the cell. Three principal types of protein filaments make up the cytoskeleton: actin filaments, microtubules, and intermediate filaments.

All three types of filaments consist of small proteins that can assemble (polymerize) into filaments of varying length. The filament structures are dynamic and can be rapidly disassembled and reassembled according to the changing needs of the cell. Actin filaments play a pivotal role in cell movement. As one might expect, muscle cells are packed with actin filaments, which allow the cell to perform its primary function of contraction. However, nonmuscle cells also possess actin filaments that are important for complex movements of the cell membrane, such as cell crawling and phagocytosis. Such movements of the cell membrane are mediated by dense networks of actin filaments that cluster just beneath the plasma membrane and interact with specific proteins embedded in it. Actin and some of the other cytoskeletal proteins make specific contacts with and through the plasma membrane and are involved in information transfer from the extracellular environment to signaling cascades within the cell.

Organization of the cytoplasm and its organelles is achieved primarily by microtubules. In animal cells, microtubules originate at the cell center, or centrosome, near the nucleus and radiate out toward the cell perimeter in fine lacelike threads. Microtubules guide the orderly transport of organelles and vesicles in the cytoplasm, as well as the equal distribution of chromosomes during cell division. Intermediate filaments, so named because their size is between that of microtubules and actin filaments, are strong, ropelike, fibrous proteins. A variety of intermediate filaments that differ from tissue to tissue have been identified. In addition to the three main groups of cytoskeletal filaments just described, a large number of accessory proteins are essential for cytoskeletal function. For example, the accessory protein myosin is needed to bind with actin to achieve motor functions. Different accessory proteins are present in different cell types.

Nucleus

The largest cytoplasmic organelle is the nucleus, which contains the genetic information for the cell in the form of DNA. The human genome contains nearly 20,000 genes that code for proteins, representing less than 1.3% of the total DNA structure composed of more than 6 billion base pairs. The nuclear contents are enclosed and protected by the nuclear envelope, which consists of two concentric membranes. The inner membrane forms an unbroken sphere around the DNA and contains protein-binding sites that help to organize the chromosomes inside. The outer nuclear membrane is continuous with the endoplasmic reticulum (ER) (see next section) and closely resembles it in structure and function (Fig. 3.11). The nucleus contains many proteins that help mediate its functions of genetic control and inheritance. These proteins, including histones, polymerases, and regulatory proteins, are manufactured in the cytosol and transported to the nucleus through holes in the membrane called *nuclear pores.* The nuclear pores are selective about which molecules are allowed access to the nuclear compartment, and in this way they protect the genetic material from enzymes and other molecules in the cytoplasm. The nuclear pores also mediate the export of products such as RNA and ribosomes that are synthesized in the nucleus but function in the cytosol. Ribosomes are manufactured in a specialized portion of the nucleus called the *nucleolus.* Nuclear pores are complexes of proteins that span across both the inner and the outer nuclear membrane, creating a pathway between the cytoplasm and the nuclear lamina (see Fig. 3.11).

A major function of the nucleus is to protect and preserve genetic information so that it can be replicated exactly and passed on during cell division. However, the nucleus is continuously functioning even when the cell is not actively dividing. The nuclear DNA controls the production of cellular enzymes, membrane receptors, structural proteins,

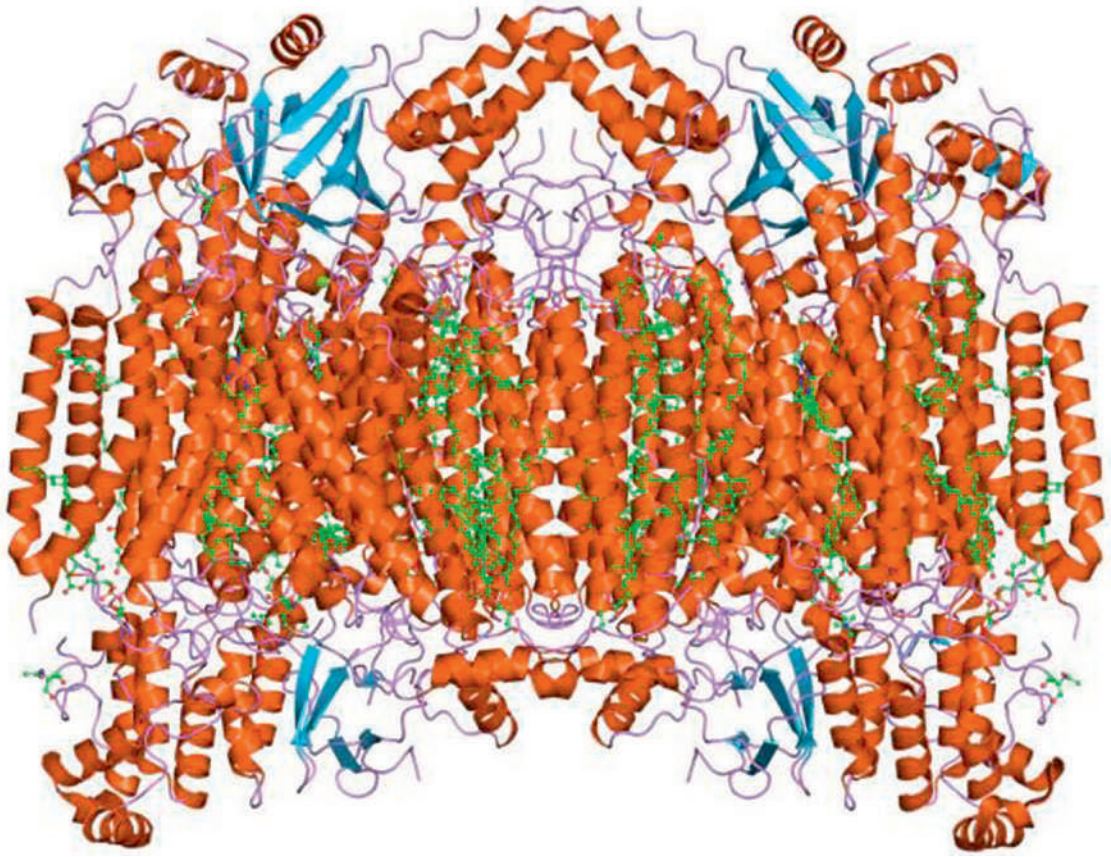


FIG 3.8 Transmembrane proteins have complex folded structures, often with several twists and turns through the membrane. Cytochrome C oxidase. (From The European Bioinformatics Institute. http://www.ebi.ac.uk/. March 20, 2017.)

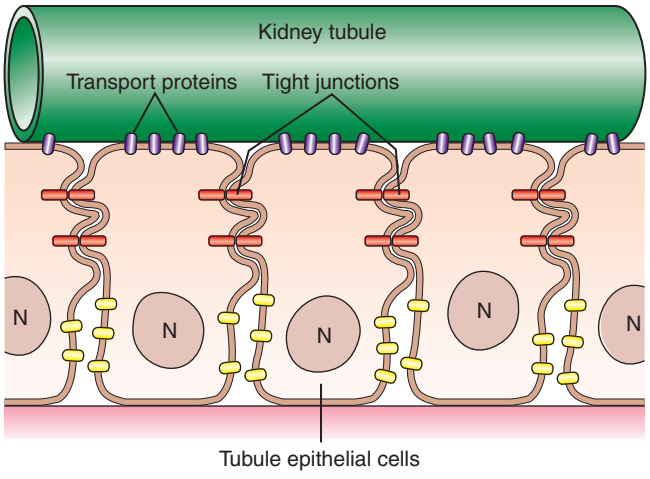


FIG 3.9 Transport proteins may be confined to a particular portion of the cell membrane by tight junctions. Segregation of transport proteins is important for the absorptive functions of the kidney epithelial cells. *N,* Nucleus.

and other proteins that define the cell's type and behavior. (The structure and function of DNA are discussed in Chapter 5.)

During mitosis, the complex structure of the nuclear membrane is retracted into the endoplasmic reticulum membrane. After cell division is complete, pieces of nuclear membrane reemerge to surround and gather the chromosomes and then fuse together to form a new nuclear membrane. Nuclear proteins and pore structures are then recruited back to their normal nuclear locations.

Endoplasmic Reticulum

The ER is a membrane network that extends throughout the cytoplasm and is present in all eukaryotic cells (Fig. 3.12). The ER is thought to have a single continuous membrane that separates the lumen of the ER from the cytosol. The ER plays a central role in the synthesis of membrane components, including proteins and lipids, for the plasma membrane and cellular organelles, as well as in the synthesis of products to be secreted from the cell. The ER is divided into rough and smooth types based on its appearance under the electron microscope. The rough ER is coated with ribosomes along its outer surface. Ribosomes are complexes of protein and RNA that are formed in the nucleus and transported to the cytoplasm. Their primary function is the synthesis of proteins (see Chapter 5). Depending on the destination of the protein to be created, ribosomes may float free in the cytosol or may bind to the ER membrane. Proteins synthesized by free-floating ribosomes are released within the cytosol of the cell. Proteins to be transported into the ER have a special sequence of amino acids that directs the ribosome responsible for its synthesis to the ER membrane. Special proteins called *signal recognition particles* bind to the leading sequence of the protein and then bind to a receptor on the ER membrane. As the ribosome adds amino acids to the growing protein chain, it is pushed into the lumen of the ER through a pore in the ER membrane called a *translocon.* After being processed in the ER and Golgi apparatus, the protein is

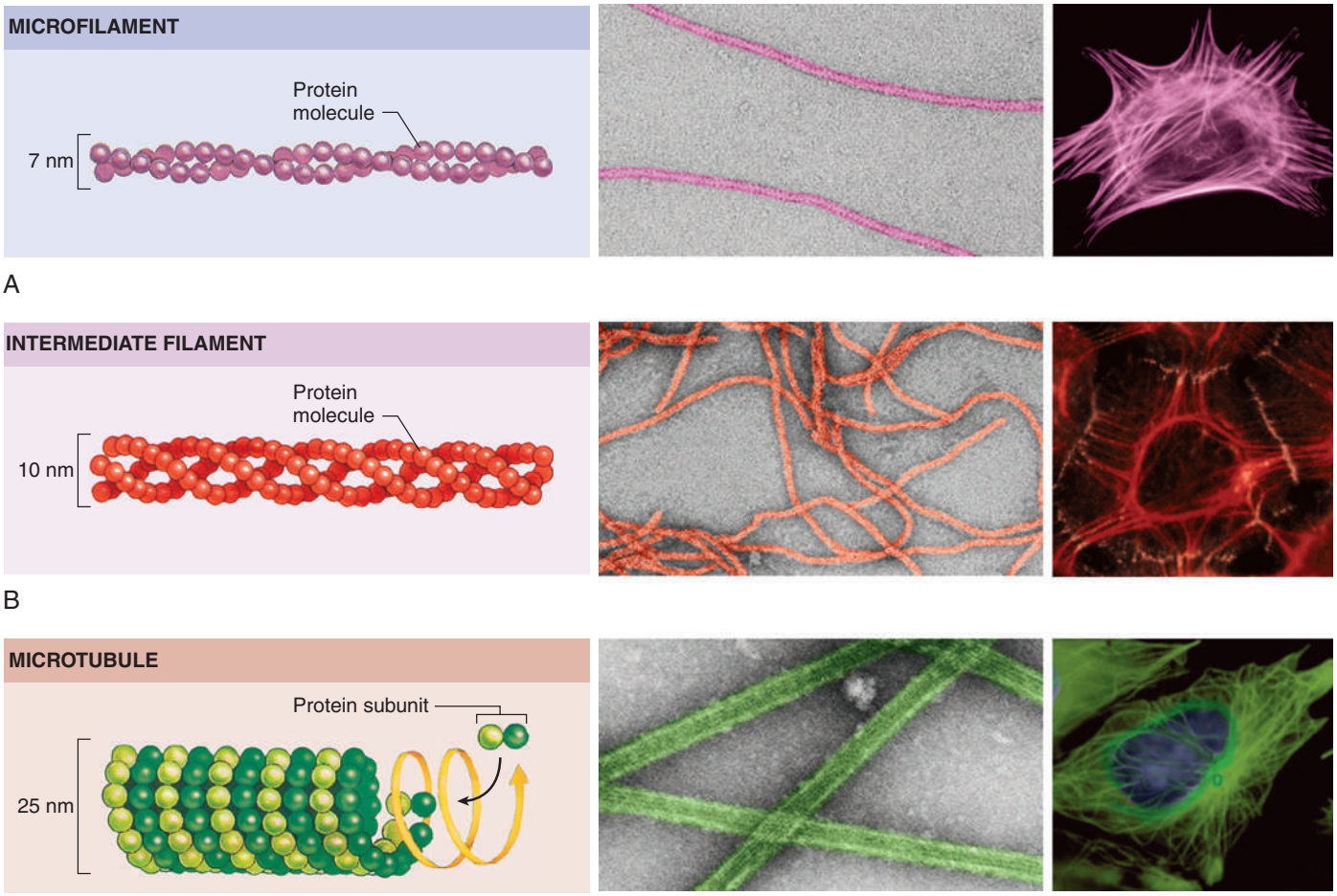


FIG 3.10 Schematic and micrographs of three major types of cytoskeletal proteins. **A,** Microfilaments shown are composed of actin proteins. **B,** Intermediate filaments are a large group of various types of proteins. **C,** Microtubules. (From Patton KT, Thibodeau GA: *Anatomy & physiology,* ed 9, St. Louis, 2016, Mosby, p. 89. Micrographs from Pollard T, Earnshaw W: *Cell biology, revised reprint, international edition,* Philadelphia, 2004, Saunders.)

eventually transported to the appropriate organelle or secreted at the cell surface. Free-floating and rough ER ribosomes are identical and interchangeable; their location depends on the amino acid structure of the protein they are producing at the time.

Regions of ER that lack ribosomes are called smooth ER. The smooth ER is involved in lipid metabolism. Most cells have very little smooth ER, but cells specializing in the production of steroid hormones or lipoproteins may have significant amounts of smooth ER. For example, the hepatocyte (liver cell) has abundant smooth ER–containing enzymes (P450) responsible for the manufacture of lipoproteins as well as the detoxification of harmful lipid-soluble compounds, such as alcohol. The cellular smooth ER can double in surface area within a few days if large quantities of drugs or toxins enter the circulation. Cells in the adrenal cortex and gonads that produce steroid hormones also have abundant smooth ER. In addition to synthetic functions, the ER sequesters calcium ions by pumping them from the cytoplasm. In response to specific signals, the ER releases calcium ions as part of important second-messenger cascades. Muscle cells have extensive smooth ER (sarcoplasmic reticulum) dedicated to the sequestration of calcium. When the cell is stimulated, the sarcoplasmic reticulum releases the calcium ions needed to accomplish muscle contraction.

Golgi Apparatus

The Golgi apparatus, or Golgi complex, is composed of a stack of smooth membrane-bound compartments resembling a stack of hollow plates or pita breads (see Fig. 3.12). These compartments, or *cisternae,* are organized in a series of at least three processing compartments. The first compartment (*cis* face) lies next to the ER and receives newly synthesized proteins and lipids by way of ER transport vesicles. These transport vesicles are outgrowths that bud off from the ER membrane and diffuse to the Golgi, where they bind and become part of the Golgi apparatus membrane. The proteins and lipids then move through the middle compartment (medial) to the final compartment (*trans* face), where they depart for their final destination. As the lipid and protein molecules pass through the sequence of Golgi compartments, they are modified by enzymes that attach or rearrange carbohydrate molecules. After specific arrangement of these carbohydrates has occurred, the lipids and proteins are packaged into Golgi transport vesicles (secretory vesicles). The particular configuration of carbohydrate molecules on the lipid or protein is believed to serve as an "address label," directing them to the correct destination within the cell. Golgi vesicles transport their contents primarily to the plasma membrane and to lysosomes.

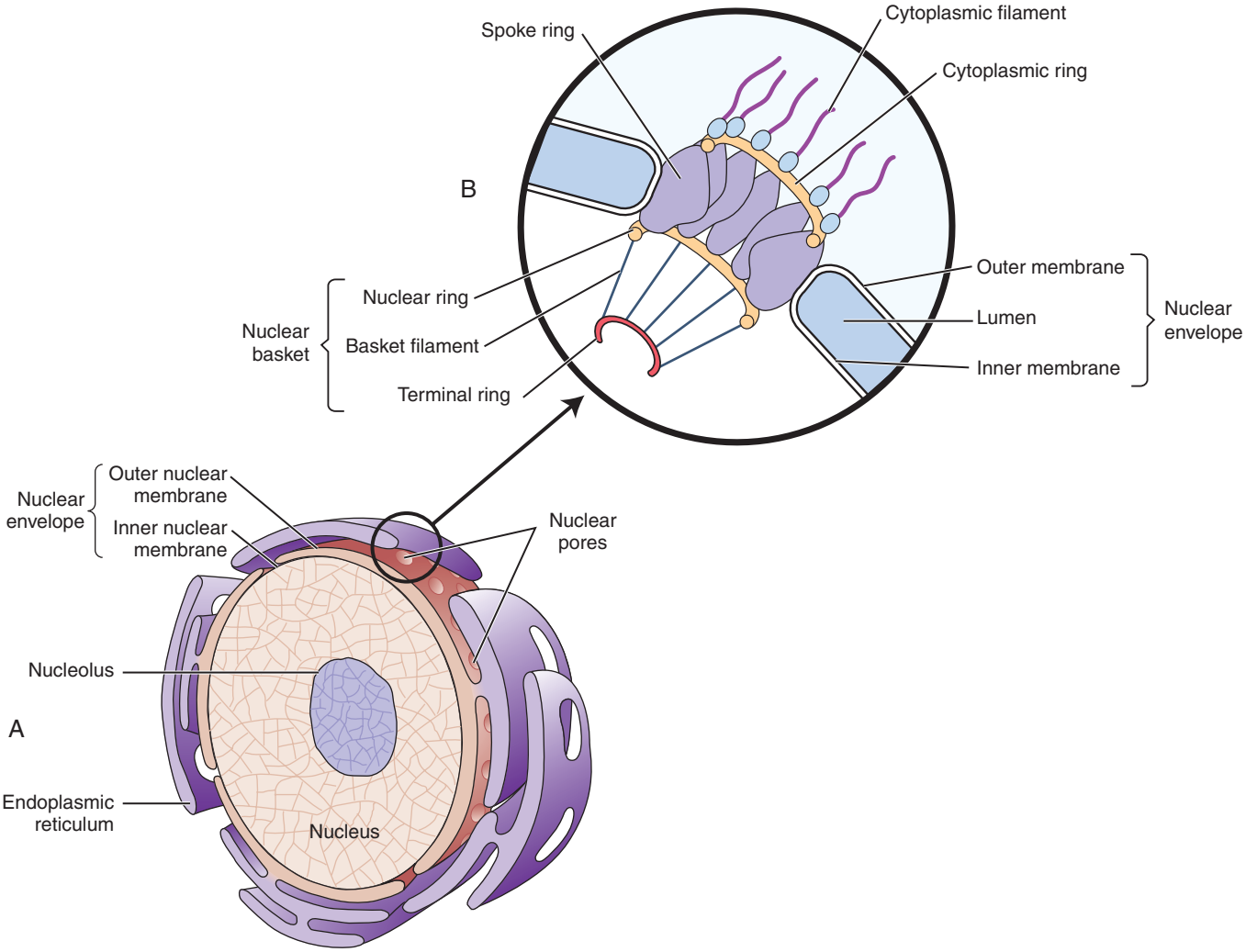


FIG 3.11 A, Structure of the double-membrane envelope that surrounds the cell nucleus. **B,** Detail of a nuclear pore.

Lysosomes and Peroxisomes

Transport of Golgi vesicles to the membrane-bound bags of digestive enzymes known as lysosomes has been well described and provides a model for Golgi sorting and transport to other destinations. Lysosomes are filled with more than 40 different acid hydrolases, which are capable of digesting organic molecules, including proteins, nucleotides, fats, and carbohydrates. Lysosomes obtain the materials they digest from three main pathways. The first is the pathway used to digest products absorbed by endocytosis. In this pathway, endocytotic vesicles bud off from the plasma membrane to fuse with endosomes. Endosomes mature into lysosomes as the Golgi delivers lysosomal enzymes to them; the pH inside the lysosome acidifies, and active digestion occurs. The second pathway is autophagy, whereby damaged and obsolete parts of the cell itself are destroyed. Unwanted cellular structures are enclosed by a membrane from the ER, which then fuses with the lysosome, leading to autodigestion of the cellular components. Autophagy also may occur during cell starvation or disuse, leading to a process called atrophy, in which the cells become smaller and more energy efficient. The third pathway providing materials to the lysosomes is present only in specialized phagocytic cells. White blood cells (WBCs), for example, are capable of ingesting large particles, which then form a phagosome capable of fusing with a lysosome. The final products of lysosomal digestion are simple molecules, such as amino acids, fatty acids, and carbohydrates, which can be used by the cell or secreted as cellular waste at the cell surface.

Discovery of the mechanism for sorting and transport of lysosomal enzymes was aided by studying patients suffering from lysosomal storage diseases. Patients with I-cell (inclusion cell) disease, for example, accumulate large amounts of debris in lysosomes, which appear as spots, or "inclusions," in the cells. These lysosomes lack nearly all of the hydrolases normally present and thus are unable to perform lysosomal digestion. However, all the hydrolases missing from the lysosomes can be found in the patient's bloodstream. The abnormality results from "mis-sorting" by the Golgi apparatus, which erroneously packages the enzymes for extracellular secretion rather than sending them to the lysosomes. Studies of this rare genetic disease resulted in the discovery that all lysosomal enzymes have a common marker, mannose-6-phosphate, which normally helps target the enzymes to the lysosomes. Persons with I-cell disease lack the enzyme needed for configuring this marker.

Peroxisomes (microbodies), like lysosomes, are membrane-bound bags of enzymes that perform degradative functions. They are particularly important in liver and kidney cells, where they detoxify various

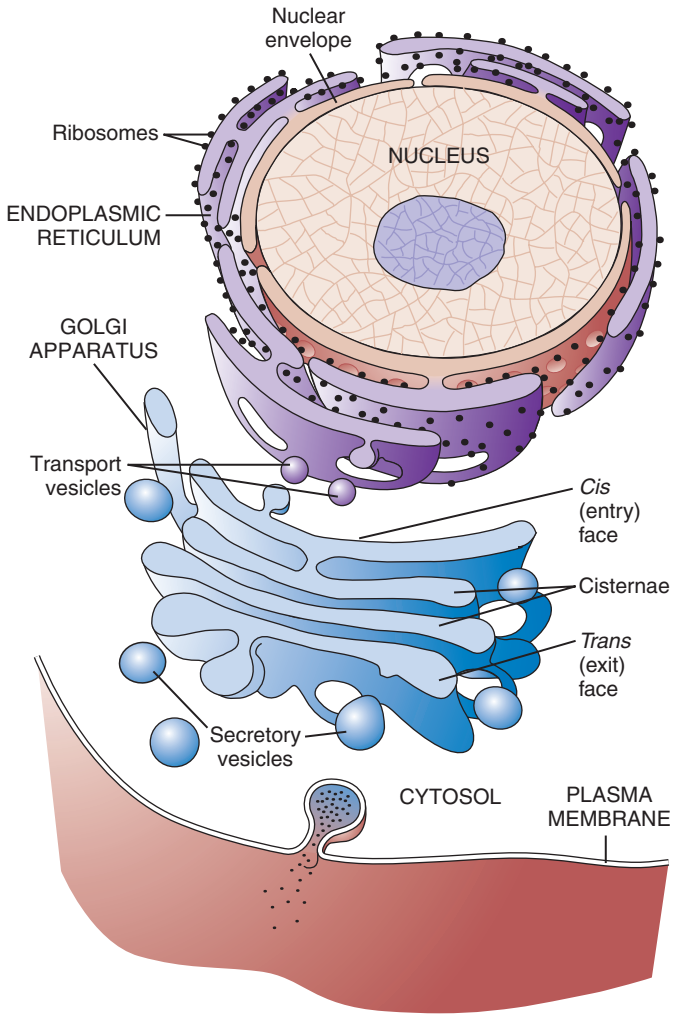


FIG 3.12 Schematic drawing of the endoplasmic reticulum and its relationship to the Golgi apparatus and nuclear envelope.

substances, such as alcohol. In contrast to lysosomes, which contain hydrolase enzymes, peroxisomes contain oxidative enzymes. These enzymes use molecular oxygen to break down organic substances by an oxidative reaction that produces hydrogen peroxide. The hydrogen peroxide is then used by another enzyme (catalase) to degrade other organic molecules, including formaldehyde and alcohol. Catalase also prevents accumulation of excess hydrogen peroxide in the cell by converting it to water and oxygen. Peroxisomes also oxidize fatty acids (β oxidation) to produce acetyl coenzyme A (acetyl CoA) that is used in cellular metabolism. Unlike lysosomes, which acquire their enzymes from Golgi vesicles, peroxisomes import enzymes directly from the cytoplasm.

Mitochondria

The mitochondria have been aptly called the "powerhouses of the cell" because they convert energy to forms that can be used to drive cellular reactions. A distinct feature of mitochondria is the large amount of membrane they contain. Each mitochondrion is bound by two specialized membranes. The inner membrane forms an enclosed space, called the *matrix,* which contains a concentrated mix of mitochondrial enzymes. The highly convoluted structure of the inner membrane with its numerous folds, called *cristae* (Fig. 3.13), provides a large surface area for the important membrane-bound enzymes of the respiratory chain. These enzymes are essential to the process of oxidative phosphorylation, which generates most of the cell's adenosine triphosphate (ATP). The outer membrane contains numerous porin transport proteins forming large aqueous channels that make the membrane porous, like a sieve. Fairly large molecules, including proteins up to 5000 daltons, can pass freely through the outer membrane such that the space between the outer and inner membranes is chemically similar to the cytosol. However, the inner membrane is quite impermeable, even to small molecules and ions. Specific protein transporters are required to shuttle the necessary molecules across the inner mitochondrial membrane.

Mitochondria are believed to have originated as bacteria that were engulfed by larger cells but that still retain some of their own DNA. Mitochondrial DNA codes for 22 transfer RNA molecules, 2 ribosomal RNAs that form mitochondrial ribosomes, and 13 proteins. During evolution the majority of mitochondrial genes were transferred to locations within the nuclear genome. Thus only a few of the mitochondrial enzymes are produced from DNA located in the mitochondria; the majority are transcribed from nuclear DNA. Nuclear genes are translated into protein in the cytoplasm and then transported to the mitochondria, whereas mitochondrial gene–derived proteins are made within the mitochondria. Several rare disorders are associated with mitochondrial gene defects (see Chapter 6). The number and location of mitochondria differ according to cell type and function. Cells with high energy needs, such as cardiac or skeletal muscle, have many mitochondria. These mitochondria may pack between adjacent muscle fibrils, such that ATP is delivered directly to the areas of unusually high energy consumption. Mitochondria undergo fission to produce more mitochondria in the cell. The details of mitochondrial energy conversion are discussed in the next section. Mitochondria also have an important role in programmed cell death, called *apoptosis,* which is discussed in Chapter 4.

KEY POINTS

- The cytoskeleton is made up of actin, microtubules, and intermediate filaments. These proteins regulate cell shape, movement, and the trafficking of intracellular molecules.
- The nucleus contains the genomic DNA. These nuclear genes code for the synthesis of proteins. There are nearly 20,000 protein-coding genes in the human genome.
- The endoplasmic reticulum (ER) and the Golgi apparatus function together to synthesize proteins and lipids for transport to lysosomes or to the plasma membrane.
- Lysosomes and peroxisomes are membrane-bound bags of digestive enzymes that degrade intracellular debris.
- Mitochondria contain enzymes necessary for oxidative phosphorylation to produce adenosine triphosphate (ATP). Mitochondria have their own small number of genes that code for some of the mitochondrial proteins.

CELLULAR METABOLISM

All living cells must continually perform essential cellular functions such as movement, ion transport, and synthesis of macromolecules. Many of these cellular activities are energetically unfavorable (i.e., they are unlikely to occur spontaneously). Unfavorable reactions can be driven by linking them to an energy source such as ATP, which is a molecule that contains high-energy phosphate bonds. In normal cells where the ATP concentration is high, approximately 11 to 13 kcal of

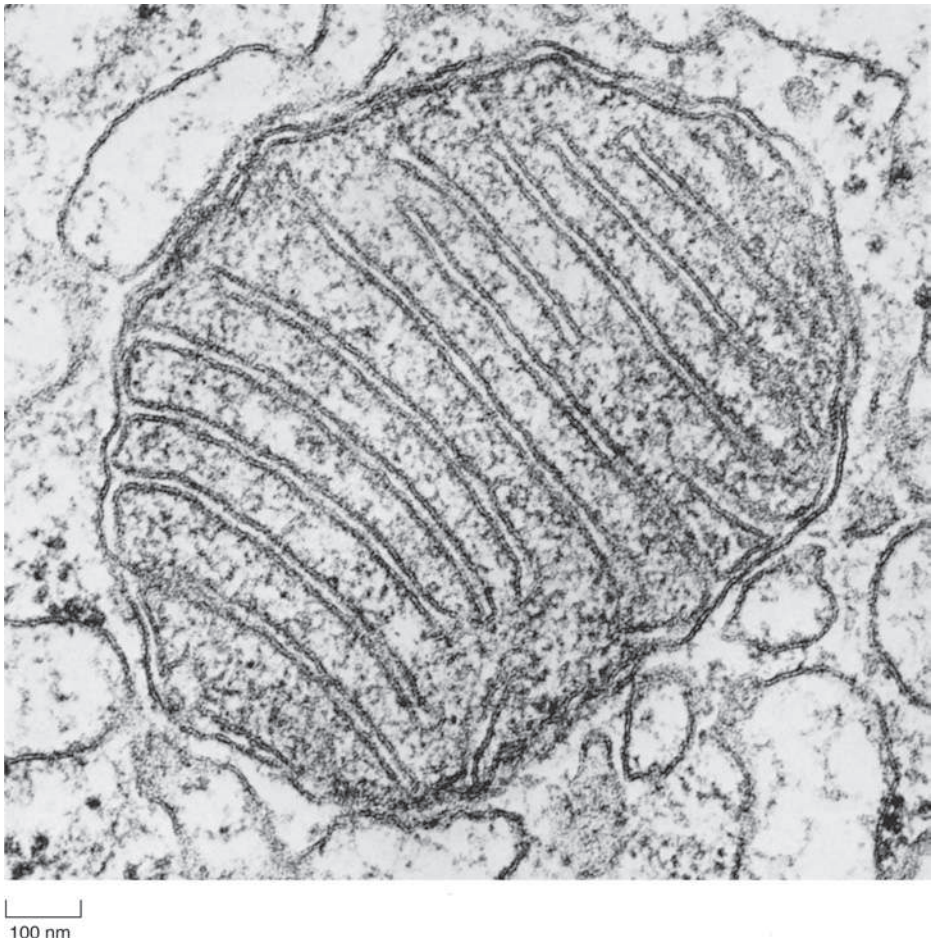


FIG 3.13 Electron micrograph **(A)** and schematic drawing **(B)** of the mitochondrial structure. The highly convoluted inner membrane provides a large surface area for membrane-bound metabolic enzymes. (A, From Alberts B et al, editors: *Molecular biology of the cell,* ed 6, New York, 2015, Garland Science, p 26. Micrograph courtesy Daniel S. Friend. All rights reserved. Used under license from The American Society for Cell Biology.)

energy per mole of ATP is liberated when one of the phosphate bonds is hydrolyzed (broken with the aid of water) in a chemical reaction. A variety of enzymes in the cell are able to capture the energy released from ATP hydrolysis and use it to break or make other chemical bonds. In this way, ATP serves as the "energy currency" of the cell. A specific amount of ATP is "spent" to "buy" a specific amount of work. Most cells contain only a small amount of ATP, sufficient to maintain cellular activities for just a few minutes. Because ATP cannot cross the plasma membrane, each cell must continuously synthesize its own ATP to meet its energy needs; ATP cannot be "borrowed" from other cells or "banked" in any significant quantity within a cell. It must be synthesized continually from the breakdown of glycogen and fat to meet the cell's energy needs.

An average adult has enough glycogen stores (primarily in liver and muscle) to supply about 1 day's needs, but enough fat to last for a month or more. After a meal, the excess glucose entering the cells is used to replenish glycogen stores or to synthesize fats for later use. Fat is stored primarily in adipose tissue and is released into the bloodstream for other cells to use when needed. When cellular glucose levels fall, glycogen and fats are broken down to provide glucose and fatty acyl molecules, respectively, which are ultimately metabolized to provide ATP. During starvation, body proteins can also be used for energy production by a process called *gluconeogenesis.*

Cellular metabolism is the biochemical process whereby foodstuffs are used to provide cellular energy and biomolecules. Cellular metabolism includes two separate and opposite phases: anabolism and catabolism. Anabolism refers to *energy-using* metabolic processes or pathways that result in the synthesis of complex molecules such as fats. Catabolism refers to the *energy-releasing* breakdown of nutrient sources such as glucose to provide ATP to the cell. Both of these processes require a long, complex series of enzymatic steps. The catabolic processes of cellular energy production are briefly discussed in the following sections. (See Chapter 42 for a detailed discussion of metabolism.)

Glycolysis

The catabolic process of energy production begins with the intestinal digestion of foodstuffs into small molecules: proteins into amino acids, polysaccharides into simple sugars (monosaccharides), and fats into fatty acids and glycerol. The second stage of catabolism occurs in the cytosol of the cell, where glucose molecules are further degraded by glycolysis into pyruvate (compounds with three carbon atoms). Glycolysis involves 10 enzymatic steps to break the six-carbon glucose molecule into a pair of three-carbon pyruvate molecules (Fig. 3.14). Glycolysis requires the use of two ATP molecules in the early stages but produces four ATP molecules in the later steps, for a net gain of two ATP molecules per glucose molecule. The production of ATP through glycolysis is relatively inefficient, and the pyruvate end products still contain substantial chemical energy that can be released by further catabolism in stage 3. However, glycolysis is an important provider of ATP under *anaerobic* conditions because oxygen is not required. Thus ATP production by glycolysis becomes important during conditions of reduced cellular oxygenation, which may accompany respiratory and cardiovascular disorders. The pyruvate that accumulates during prolonged anaerobic conditions is converted to lactate and excreted from the cell into the bloodstream. Lactic acidosis is a dangerous condition that may result from excessive lactate production attributable to severe or prolonged lack of oxygen (see Chapter 20). In addition to the two molecules of ATP and pyruvate, each glucose molecule produces two reduced nicotinamide adenine dinucleotide (NADH) molecules, which contain high-energy electrons that are transferred to the electron transport chain in the mitochondria. Cells that do not contain mitochondria, such as RBCs, must rely totally on glycolysis for ATP production.

Citric Acid Cycle

For most cells, glycolysis is only a prelude to the third stage of catabolism, which takes place in the mitochondria and results in the complete oxidation of glucose to its final end products: CO_2 and H_2O. The third stage begins with the citric acid cycle (also called the *Krebs cycle* or the *tricarboxylic acid cycle*) and ends with the production of ATP by oxidative phosphorylation. The purpose of the citric acid cycle is to break, by oxidation, the C–C and C–H bonds of the compounds produced in the second stage of catabolism. Pyruvate and fatty acids enter the mitochondrial matrix, where they are converted to acetyl CoA (Fig. 3.15). The pyruvate dehydrogenase complex cleaves pyruvate to form one CO_2, one NADH, and one acetyl CoA molecule. Fatty acids are cleaved by a process called *β oxidation* to form one NADH and one reduced flavin adenine dinucleotide ($FADH_2$, another type of electron carrier). No CO_2 is produced by β oxidation of fatty acids. Patients who have difficulty excreting CO_2 because of respiratory disease are sometimes given a high-fat, low-carbohydrate diet to take advantage of the lower CO_2 production that accompanies fat metabolism.

In the first reaction of the citric acid cycle, the two-carbon acetyl group is transferred from coenzyme A to a four-carbon oxaloacetate molecule. This results in the formation of the six-carbon molecule citrate for which the cycle is named. In a series of enzymatic oxidations, carbon atoms are cleaved off in the form of CO_2 (Fig. 3.16); this CO_2 is free to diffuse from the cell and be excreted by the lungs as a waste product. Two carbon atoms are removed to form two CO_2 molecules for each complete turn of the cycle. The extra oxygen molecules needed to create CO_2 are provided by the surrounding H_2O; therefore the citric acid cycle does not require molecular oxygen from respiration. However, the cycle will cease to function in the absence of oxygen because the carrier molecules, NADH and $FADH_2$, cannot unload their electrons onto the electron transport chain (which does require oxygen) and thus are unavailable to accept electrons from the citric acid cycle.

Although the citric acid cycle directly produces only one ATP molecule (in the form of guanosine triphosphate [GTP]) per cycle, it captures a great deal of energy in the form of activated hydride ions (H–). These high-energy ions combine with larger carrier molecules, which transport them to the electron transport chain in the mitochondrial membrane. Two important carrier molecules are nicotinamide adenine dinucleotide (NAD^+), which becomes NADH when reduced by H–, and flavin adenine dinucleotide (FAD), which becomes $FADH_2$ when reduced by H–. The energy carried by these molecules is ultimately used to produce ATP through a process called *oxidative phosphorylation.* One glucose molecule provides for two turns of the cycle and produces a net of two GTP, four CO_2, two $FADH_2$ and six NADH.

Oxidative Phosphorylation

Oxidative phosphorylation follows the processes of glycolysis and the citric acid cycle and results in the formation of ATP by the reaction of adenosine diphosphate (ADP) and inorganic phosphate (P_i): ADP + $P_i \rightarrow$ ATP. The energy to drive this unfavorable reaction is provided by the high-energy hydride ions (H–) derived from the citric acid cycle. This energy is not used to form ATP directly; a series of energy transfers through reduction-oxidation (redox) reactions is required. In eukaryotic cells, this series of energy transfers occurs along the electron transport chain on the inner mitochondrial membrane. The transport chain consists of three major enzyme complexes and two mobile electron carriers that shuttle electrons between the protein complexes

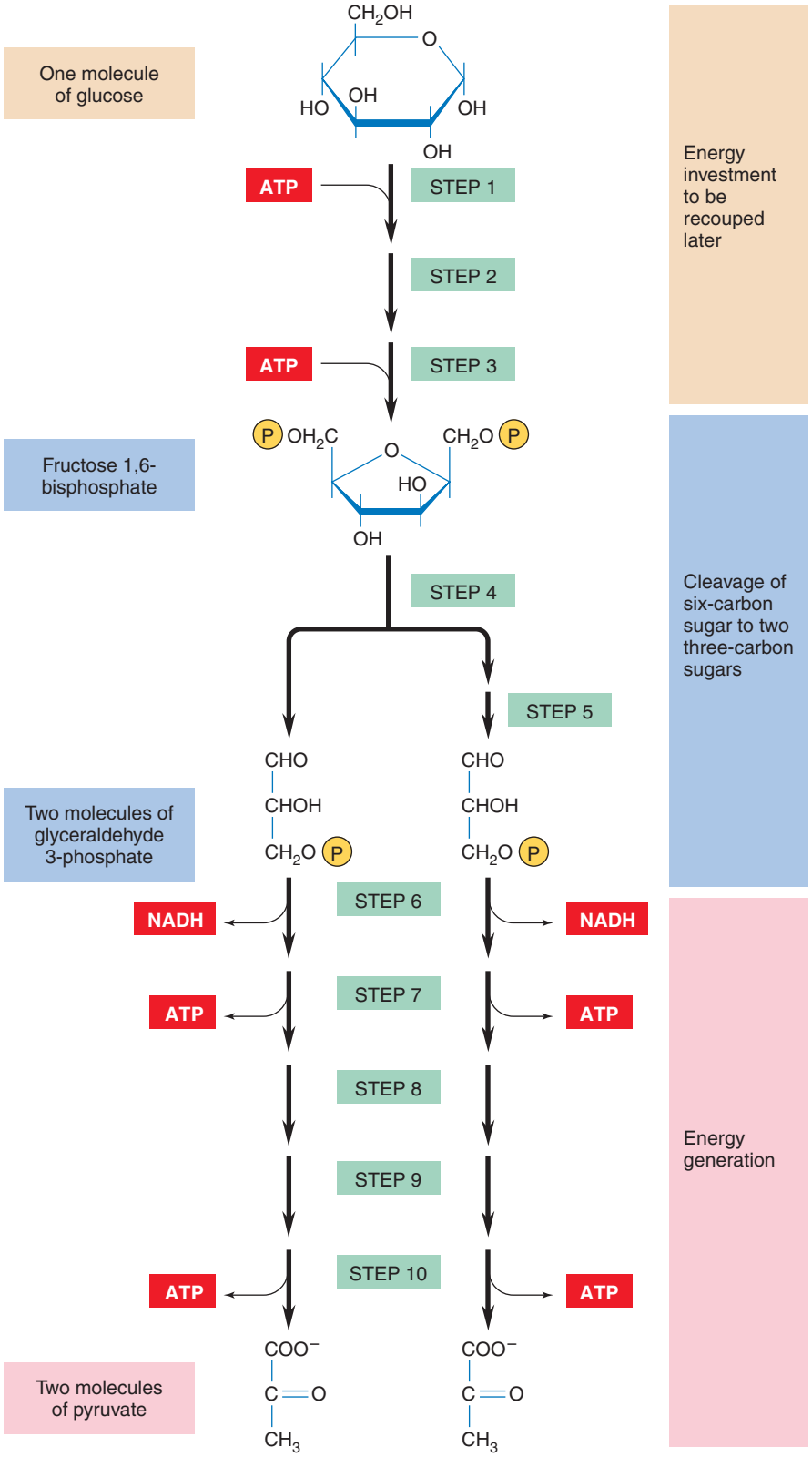


FIG 3.14 Ten enzymatic steps are required in glycolysis to break glucose into two 3-carbon pyruvate molecules. A net gain of two ATP molecules is achieved. (From Alberts B et al, editors: *Molecular biology of the cell,* ed 6, New York, 2015, Garland Science, p 26. Micrograph courtesy Daniel S. Friend. All rights reserved. Used under license from The American Society for Cell Biology.)

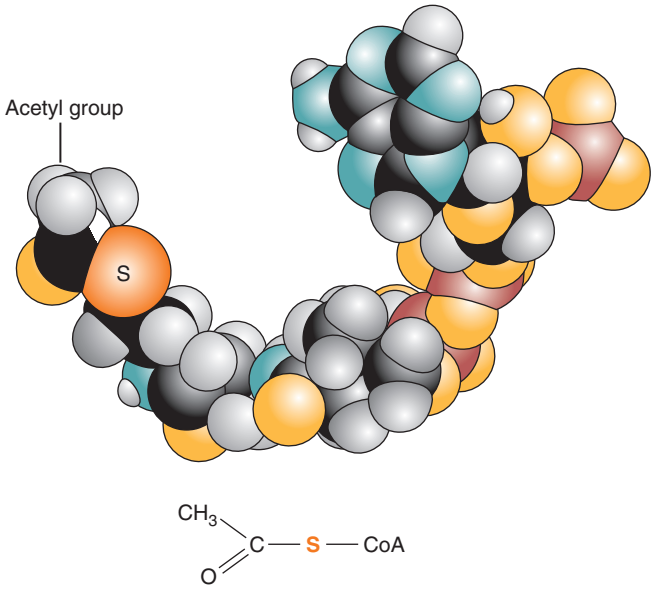


FIG 3.15 Space-filling model of acetyl CoA.

(Fig. 3.17). Respiratory chain proteins contain metal ions (iron, copper) that facilitate the transfer of electrons. The hydrogen molecules and their associated electrons are transported to the electron transport chain by the carrier molecules NADH and $FADH_2$. The path of electron flow is NADH → NADH dehydrogenase complex → ubiquinone → b-c_1 complex → cytochrome c → cytochrome oxidase complex. With each redox reaction the electrons pass from one complex to the next, and the free energy that is released is used to pump hydrogen ions (H^+) out of the mitochondrial matrix. At the very end of the transport chain, low-energy electrons are finally transferred to O_2 to form H_2O. Oxidative phosphorylation is called *aerobic* because of this oxygen-requiring step. The last enzyme in the chain, cytochrome oxidase, collects four electrons and then transfers all four at once to a molecule of O_2 to create two water molecules. If electrons are not transferred to oxygen in the correct ratio, then oxygen free radicals may be produced and damage the cell. Free radical generation is discussed in Chapter 4.

Thus far, little ATP synthesis has been accomplished. However, the enzymes of the transport chain have harnessed energy from the transported electrons in the form of a proton (H^+) gradient. Finally, the proton gradient is used to power the synthesis of ATP. A special enzyme in the inner mitochondrial membrane (ATP synthase) allows protons

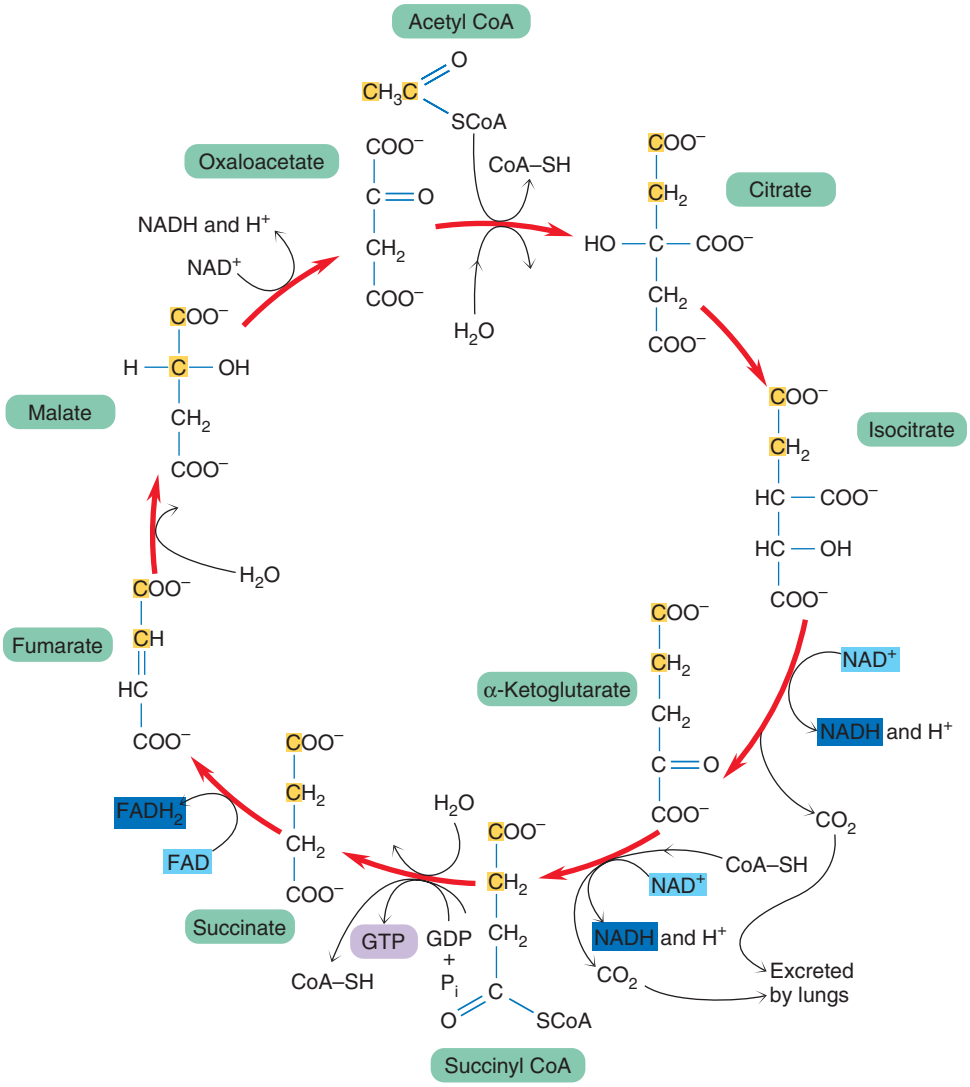


FIG 3.16 Chemical structures of the compounds of the citric acid cycle (Krebs cycle). In a series of enzymatic reactions, carbon atoms are cleaved to form CO_2 and high-energy hydride ions, which are carried by FAD and NAD.

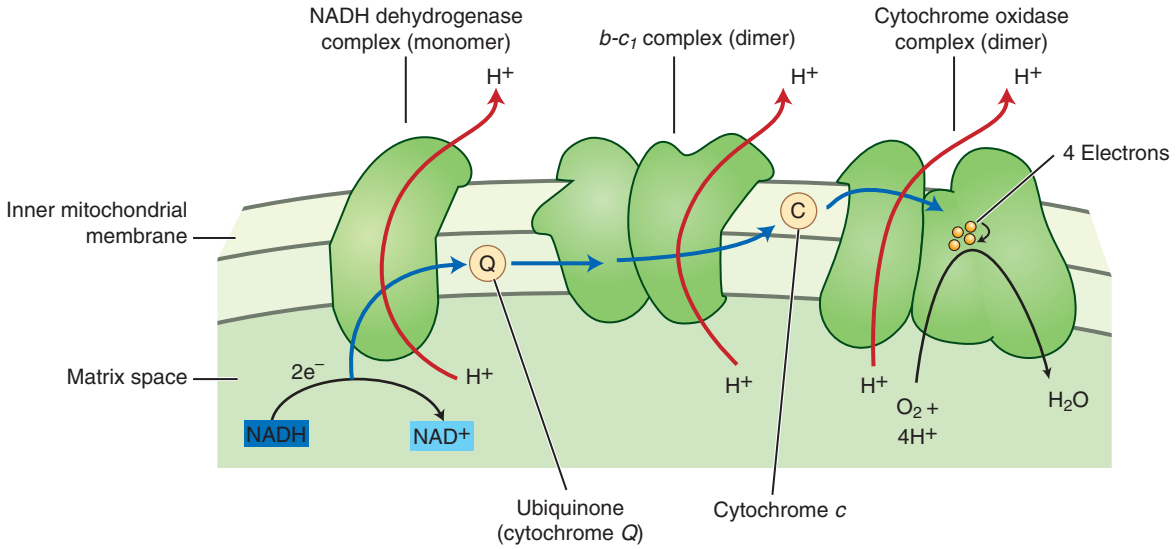


FIG 3.17 Representation of the electron transport chain located in the inner mitochondrial membrane. High-energy electrons are passed along the chain until they combine with oxygen to form water. The energy released at each electron transfer is used to pump H+ across the membrane.

to flow back into the mitochondria down their electrochemical gradient. The energy of the proton flow is used to drive ATP synthesis (Fig. 3.18). Under normal cellular conditions about 30 ATP molecules are formed from the complete oxidation of glucose into CO_2 and H_2O. Two of these are from glycolysis, two from the citric acid cycle (in the form of GTP), and the remainder from oxidative phosphorylation. The ATP formed within the mitochondria is transported to the cytosol by protein transporters in the mitochondrial membrane. The ATP is then available to drive a variety of energy-requiring reactions within the cell.

KEY POINTS

- Energy-requiring reactions within cells are driven by coupling to adenosine triphosphate (ATP) hydrolysis.
- ATP is not stored and must be continuously synthesized by each cell to meet the cell's energy needs.
- Glycolysis is an anaerobic process that produces two ATP molecules, two NADH molecules, and two pyruvate molecules per glucose molecule. Pyruvate enters the mitochondria and is converted to acetyl CoA with release of a CO_2 molecule. Pyruvate can also be converted to lactate when oxygen supply is insufficient for oxidative processes.
- The citric acid cycle in the mitochondrial matrix oxidizes the acetyl groups supplied by acetyl CoA to form large quantities of H– (hydride ions), which are carried to the respiratory chain by NADH and $FADH_2$.
- The respiratory chain enzymes capture the energy from electron transfer and use it to produce an H^+ (proton) gradient. Molecular oxygen is required at this stage (aerobic) to accept the electrons from the last enzyme in the transport chain.
- ATP is produced by ATP synthase, a protein in the mitochondrial membrane. ATP synthase produces ATP by capturing the energy of the proton gradient and using it to form a bond between ADP and inorganic phosphate (P_i). In total, about 30 ATP molecules are produced per glucose molecule.

FUNCTIONS OF THE PLASMA MEMBRANE

Membrane Transport of Macromolecules

Endocytosis and Exocytosis

The transport of large molecules, such as proteins and polysaccharides, across the plasma membrane cannot be accomplished by the membrane

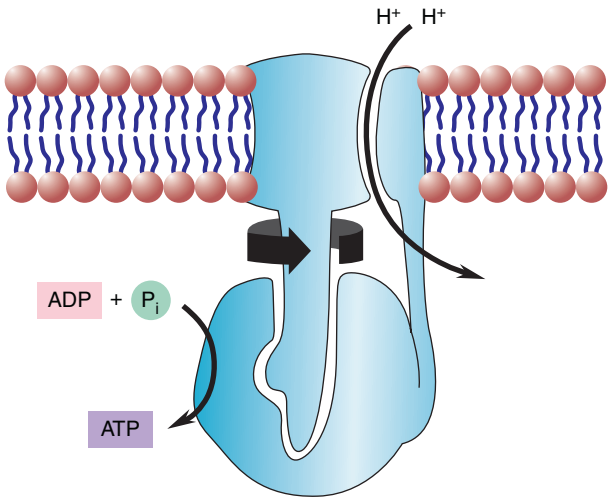


FIG 3.18 Inner mitochondrial ATP synthetase captures the potential energy of the H+ gradient in a manner similar to a turbine. The proton gradient drives the synthesis of ATP from adenosine diphosphate (ADP) and inorganic phosphate (Pi). A 360-degree rotation of the rotor requires 12 H+ ions and produces 3 ATP molecules.

transport proteins discussed earlier. Rather, macromolecules are ingested and secreted by the sequential formation and fusion of membrane-bound vesicles. *Endocytosis* refers to cellular ingestion of extracellular molecules. The process of cellular secretion is called *exocytosis*. There are two types of endocytosis, which are differentiated by the size of the particles ingested. Pinocytosis, or "cellular drinking," is the method of ingesting fluids and small particles and is common to most cell types. Phagocytosis, or "cellular eating," involves the ingestion of large particles, such as microorganisms, and is practiced mainly by specialized phagocytic WBCs. Endocytosis begins at the cell surface by the formation of an indentation, or "pit," in the plasma membrane, which is coated with special proteins, including clathrin *(coated pit)*. The indentation invaginates and then pinches off a portion of the membrane to become a vesicle (Fig. 3.19). Each vesicle thus formed is internalized, sheds its coat, and fuses with an endosome. The contents of these endocytic vesicles usually accumulate in lysosomes, where they are degraded.

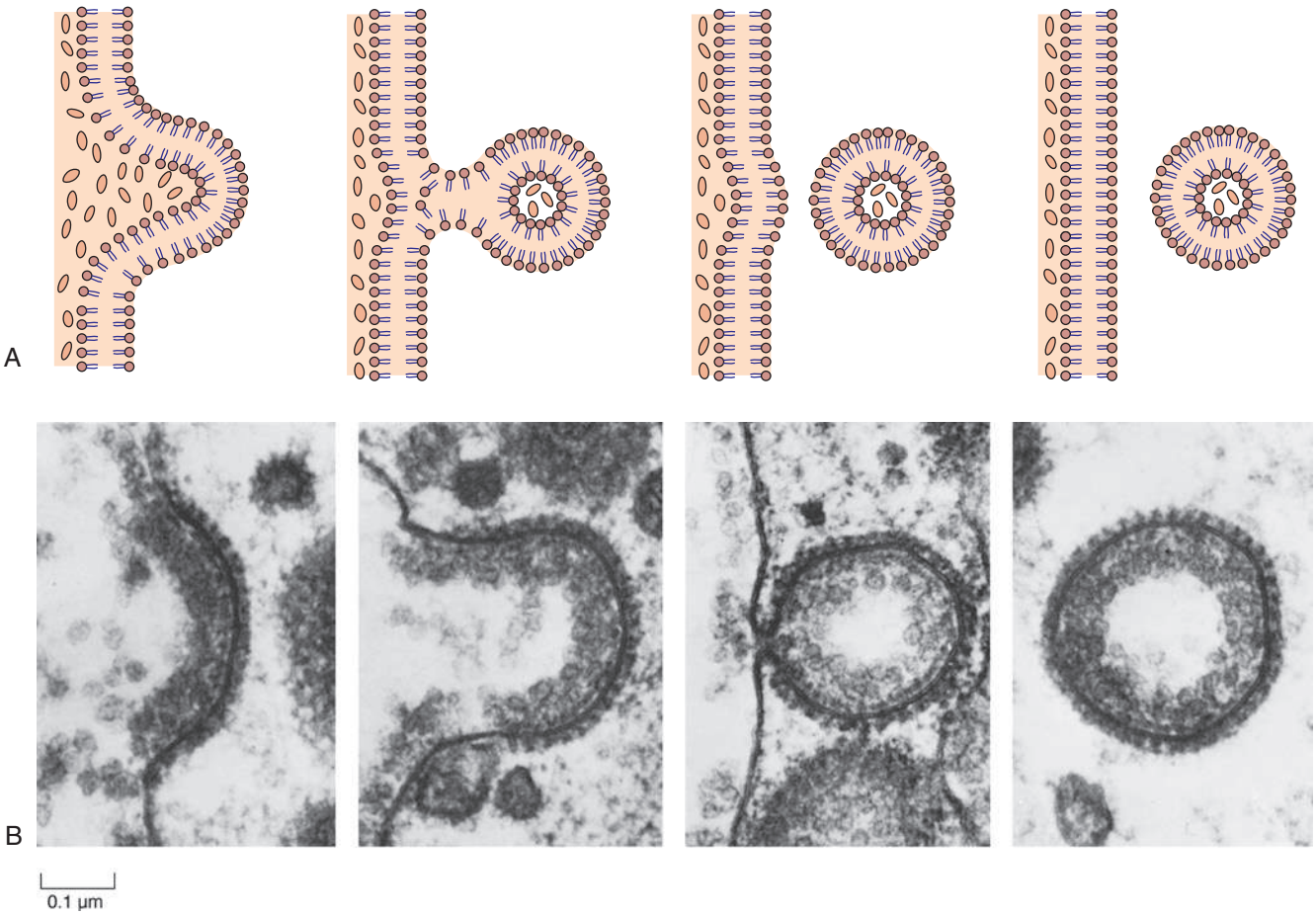


FIG 3.19 A, Representation of the steps of endocytosis. An invagination of the membrane occurs and pinches off to form a vesicle. Exocytosis progresses in essentially the reverse sequence. **B,** Electron micrograph showing the steps of endocytosis. (B, From Perry M, Gilbert A: Yolk transport in the ovarian follicle of the hen [*Gallus domesticus*]: lipoprotein-like particles at the periphery of the oocyte in the rapid growth phase, *J Cell Sci* 39:257–272, 1979.)

Endocytosis of certain macromolecules is regulated by specific receptors on the cell surface. These receptors bind the molecules (ligands) to be ingested and then cluster together in coated pits. The receptor–ligand complexes are internalized by the invagination process described previously. The vesicles generally fuse with endosomes where the ligand is removed from the receptor for processing by the cell. The receptor may be degraded in the lysosome or may be recycled to the cell surface to be used again. Receptor-mediated endocytosis allows the cell to be selective about the molecules ingested and to regulate the amount taken into the cell. The cell can produce greater numbers of cell surface receptors to ingest more ligand.

An example of receptor-mediated endocytosis is the cellular uptake of cholesterol. The process of cholesterol uptake by cells is shown in Fig. 3.20. Most cholesterol in the blood is transported by protein carriers called *low-density lipoproteins (LDLs)*. The cell can regulate the number of LDL receptors on its cell surface to increase or decrease the uptake of cholesterol. Once the LDL binds to its receptor, this complex is rapidly internalized in a coated pit. The coated vesicle thus formed sheds its coat and fuses with an endosome. In the endosome, the LDL receptor is retrieved and recycled to the cell surface to be reused. The LDL is transported to lysosomes and degraded to release free cholesterol, which the cell uses for synthesis of biomolecules such as steroid hormones.

Dangerously high blood cholesterol levels occur in some individuals who lack functional LDL receptors. These individuals inherit defective genes for making LDL receptor proteins (familial hyperlipidemia) and are incapable of taking up adequate amounts of LDL. Accumulation of LDL in the blood predisposes these individuals to development of atherosclerosis and heart disease (see Chapter 18).

Exocytosis is essentially the reverse of endocytosis. Substances to be secreted from the cell are packaged in membrane-bound vesicles and travel to the inner surface of the plasma membrane. There the vesicle membrane fuses with the plasma membrane, and the contents of the vesicle arrive at the cell surface. Some secreted molecules may remain embedded in the cell membrane, others may be incorporated into the extracellular matrix, and still others may enter the extracellular fluids and travel to distant sites. Many substances synthesized by the cell, including new membrane components, are constantly being packaged and secreted. This continuously operative and unregulated pathway is termed *constitutive*. In some specialized cells, selected proteins or small molecules are packaged in secretory vesicles, which remain in the cell until the cell is triggered to release them. These special secretory vesicles are typically regulated by stimulation of cell surface receptors. For example, the mast cell, a special type of WBC, releases large amounts of histamine when its cell surface receptors are activated (see Chapter 9).

Membrane Transport of Small Molecules

All cells must internalize essential nutrients, excrete wastes, and regulate intracellular ion concentrations. However, the lipid bilayer is extremely

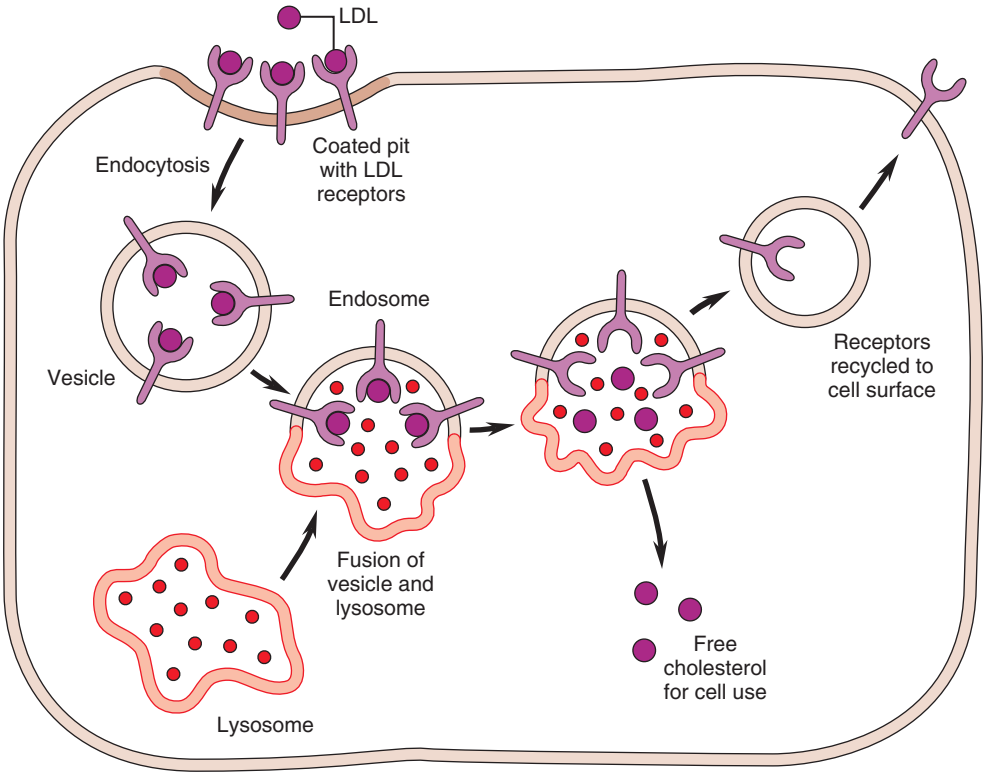


FIG 3.20 Steps in the process of receptor-mediated endocytosis of cholesterol. Cholesterol is carried in the blood by LDL. The uptake of LDL with its associated cholesterol, is mediated by a specific LDL-receptor protein on the cell surface. Once internalized, the cholesterol is removed from the LDL-receptor complex and used by the cell. The LDL receptors are sent back to the cell surface to bind more LDL.

impermeable to most polar and charged molecules. Transport of small water-soluble molecules is achieved by specialized transmembrane proteins called *transporter proteins.* Most membrane transporters are highly specific—a different transporter protein is required for each type of molecule to be transported. Only lipid-soluble molecules can permeate the lipid bilayer directly by simple diffusion.

Membrane transport proteins are of three basic kinds: ATP-driven pumps, carriers, and channel proteins. Channel proteins are the simplest of the three, forming a water-filled pore through the lipid bilayer. These pores are able to open and close to allow ions to pass through the membrane. The particular structure of the protein channel ensures that only ions of a certain size and charge can move through the membrane. Pumps and carrier proteins, however, bind to the solute to be transported and move it through the membrane by undergoing a structural, or conformational, change. Pumps and carriers have a transport maximum that is much lower than that of channels because they must bind to the molecules to be transported and then move them through the membrane. Pumps and carriers, which transport ions and nonelectrolyte molecules (e.g., glucose and amino acids), are also highly specific for the substances they transport.

Lipid-soluble particles can cross the lipid bilayer directly by simple diffusion through the hydrophobic lipid portion of the membrane. Polar or charged molecules must cross the membrane via protein channels or carriers. Transport through membrane proteins may be a passive or an active process. Passive transport through membrane proteins is called *facilitated diffusion.* Diffusion of ions occurs passively because of an electrochemical gradient. The electrochemical gradient exists because of differences in intracellular and extracellular charge and/or concentration of chemicals and is governed by laws of physics. Channel proteins only allow particles to move down their electrochemical or concentration gradients. Some carriers are passive, but others use the movement of one ion flowing down its concentration gradient (usually Na^+ moving into the cell) to move another substance uphill against its gradient. This process is called *secondary active transport* because ATP is not used directly; however, ATP is necessary to run the pumps that maintain the sodium gradient. The lipid bilayer is fairly impermeable to water because of its polar structure. Water moves across the plasma membrane through channels called *aquaporins.* Nearly all cells have aquaporins present in their cell membranes at all times, with the exception of a few specialized cells in the kidney tubules. Net movement of water across a membrane (osmosis) occurs in response to differences in osmotic pressure on either side of the membrane and is a passive process.

Active Transport Pumps

Active transport is the process whereby protein transport pumps move solutes across the membrane against an electrochemical or concentration gradient. Primary active transport requires metabolic energy, which is supplied by ATP hydrolysis. There are three families of ATP-driven pumps: the F-type ATPases that move H^+; the P-type adenosine triphosphatase (ATPase) that pump ions across membranes; and the ATP-binding cassette (ABC) transporters that transport a wide range of solutes. The ATP synthase located on the inner mitochondrial membrane is an example of an F-type pump; however, in that location it runs backward, allowing H^+ to run down its electrochemical gradient, and uses the energy to form a bond between ADP and P_i (see Fig. 3.18). As a general principle, pumps, carriers, and channels can transport in either direction depending on the concentration of substrates on either side of the membrane.

Sodium–potassium ion pump. The sodium–potassium (Na^+–K^+) pump is a P-type ATPase present in the plasma membranes of virtually

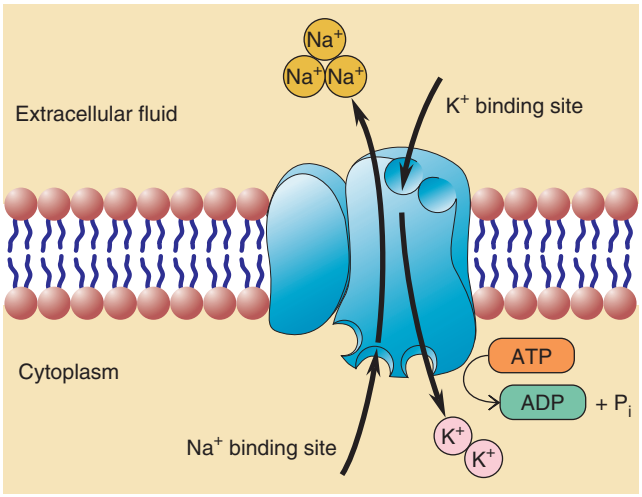


FIG 3.21 Schematic drawing of the sodium–potassium transport protein, which uses ATP to pump Na+ out of the cell and K+ into the cell against steep electrochemical gradients. This transporter is responsible for maintaining a low intracellular concentration of Na+ and a large Na+ gradient across the membrane. The energy of this Na+ gradient can be harvested by other transporters to actively transport substances.

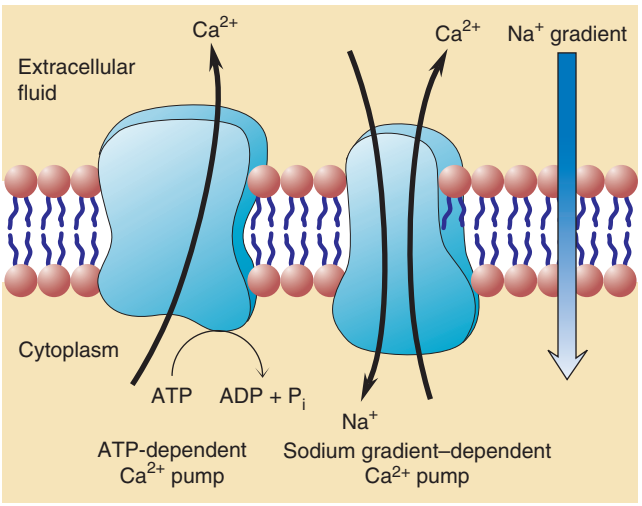


FIG 3.22 Two transporters of calcium ions are present in some cell membranes. One uses ATP as the energy source to pump calcium against a gradient (primary active transport). The other captures the potential energy of the sodium gradient to pump calcium out of the cell (secondary active transport).

all animal cells. It serves to maintain low sodium and high potassium concentrations in the cell. The Na^+–K^+ transporter must pump ions against a steep electrochemical gradient. Almost one-third of the energy of a typical cell is consumed by the Na^+–K^+ pump. ATP hydrolysis provides the energy to drive the Na^+–K^+ transporter. The Na^+–K^+ pump behaves as an enzyme in its ability to split ATP to form ADP and P_i, leading to the protein being termed *Na^+–K^+ ATPase.*

Transport of sodium and potassium ions through the Na^+–K^+ carrier protein is *coupled;* that is, the transfer of one ion must be accompanied by the simultaneous transport of the other ion. The transporter moves three sodium ions *out* of the cell for every two potassium ions moved *into* the cell (Fig. 3.21). The Na^+–K^+ pump is important in maintaining cell volume. It controls the solute concentration inside the cell, which in turn affects the osmotic forces across the membrane. If Na^+ is allowed to accumulate within the cell, the cell will swell and could burst. The role of the Na^+–K^+ pump can be demonstrated by treating cells with digitalis, a drug that inhibits Na^+–K^+ ATPase. Cells thus treated will indeed swell and often rupture. The Na^+–K^+ pump is responsible for maintaining a steep concentration gradient for Na^+ across the plasma membrane. This gradient can be harnessed to transport small molecules across the membrane in a process called *secondary active transport.* Carriers that use ATP directly are engaged in primary active transport.

Membrane calcium transporters. Numerous important cellular processes, such as cell contraction and growth initiation, are dependent on the intracellular calcium ion concentration. Intracellular Ca^{2+} is normally very low and tightly regulated. Two important calcium pumps, present in the plasma membrane and in the ER (sarcoplasmic reticulum of muscle cells), function to remove Ca^{2+} from the cell cytoplasm. Similar to the Na^+–K^+ transporter, these transporters use ATP as the energy source (Fig. 3.22, left).

If calcium ion levels in the cytoplasm become dangerously elevated, calcium pumps in the mitochondrial membrane are activated. Calcium ions are actively pumped into the mitochondria using the energy of the proton (H^+) gradient. This is the same proton gradient that the mitochondria use to synthesize ATP, and ATP production declines when the mitochondria are required to sequester Ca^{2+}. A high intracellular Ca^{2+} level is even more dangerous to the cell than a reduction in ATP production.

ABC transporters. Another important class of ATP-driven transporters is the ABC transporter family. These transporters all have a common ATP-binding domain, called the *A*TP *b*inding *c*assette (ABC), which hydrolyzes ATP to provide energy for the transport process (Fig. 3.23). This family of membrane transporters is the largest of the transporter families. A clinically important member of this family is a chloride channel in the plasma membrane of epithelial cells. A defect in this transporter is responsible for cystic fibrosis, a common genetic disorder that affects the lungs and pancreas (see Chapter 22). Bacteria use ABC transporters to pump antibiotics out of the cell, resulting in drug resistance (see Chapter 8).

Membrane Transport Carriers

Na^+-driven carriers. In animal cells, the Na^+ gradient created by the Na^+–K^+ pump is used to power a variety of transporters by secondary active transport. An important Ca^{2+} transporter located in the plasma membrane of cardiac muscle cells uses the electrochemical gradient of Na^+ to power the transport of Ca^{2+} out of the cell (see Fig. 3.22, right). The dependence of this calcium transporter on the sodium gradient helps explain the inotropic effects of the commonly prescribed drug digitalis. Digitalis is a cardiac glycoside that inhibits the Na^+–K^+ pump and allows the accumulation of intracellular Na^+. The Na^+ concentration gradient across the membrane is thus decreased, leading to less efficient calcium removal by the Na^+-dependent Ca^{2+} pump. A more forceful cardiac muscle contraction results from the increased intracellular Ca^{2+} concentration. Another example of a transporter that uses secondary active transport is the Na^+–H^+ exchange carrier, which uses the Na^+ gradient to pump out excess hydrogen ions to help maintain intracellular pH balance. The Na^+ gradient also can be used to bring substances into the cell. For example, glucose and amino acid transport into epithelial cells is coupled to Na^+ entry. As Na^+ moves through the transporter down its electrochemical gradient, the sugar or amino acid is "dragged" along. Entry of the nutrient will not occur unless Na^+ also enters the cell. The epithelial cells that line the gut and kidney tubules have large numbers of these nutrient transporters present in the luminal (apical) surfaces of their cell membranes. In this way, large amounts of glucose and amino acids can be effectively absorbed. The reuptake of numerous types of neurotransmitters from synapses also occurs via Na^+-driven

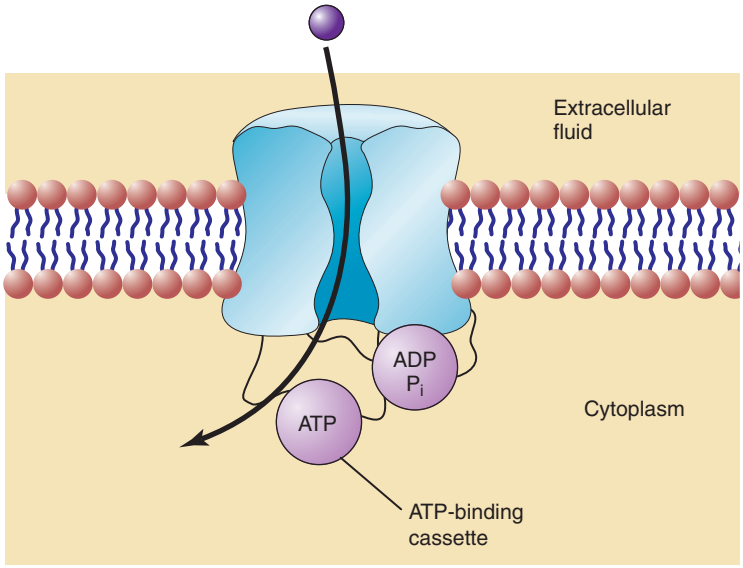


FIG 3.23 The ABC transporters are the largest known family of membrane transport proteins. They are characterized by an ATP-binding domain that causes a substrate pocket to be exposed first on one side of the membrane and then on the other as ATP is bound and hydrolyzed to ADP and Pi.

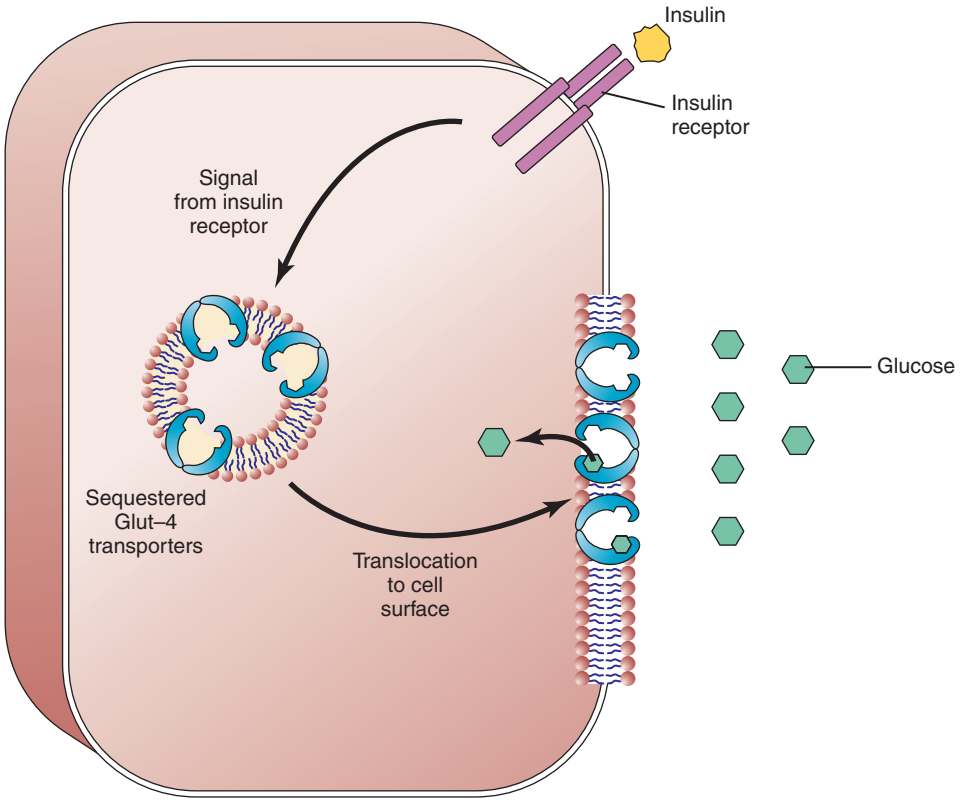


FIG 3.24 In response to insulin binding to its receptor on the cell surface, carrier proteins that transport glucose (Glut-4) are moved to the cell surface where they passively transport glucose into the cell (facilitated diffusion).

carrier proteins. The movement of Na^+ through carriers located in the presynaptic neuron drags the neurotransmitter from the synapse back into the nerve terminal, where it can be repackaged for reuse or metabolized by cellular enzymes.

Passive transport carriers. Some carriers are not linked to the Na^+ gradient and move substances across the membrane passively. The glucose transporters in many cell types belong to this class of transporters. In β cells of the pancreas, for example, the glucose transporters (Glut-1) are always present in the plasma membrane and let glucose into the cell according to its concentration in the extracellular fluid. In this way the pancreas detects blood glucose levels and releases an appropriate amount of insulin. In insulin-sensitive cells, such as muscle, liver, and adipose cells, the glucose carriers are sequestered inside the cell until insulin binds to its receptor at the cell surface. Receptor activation causes the glucose carriers (Glut-4) to move to the cell surface, where they allow passive influx of glucose (Fig. 3.24).

Membrane Channel Proteins

In contrast to carrier proteins, which bind molecules and move them across the membrane by a conformational change, channel proteins form water-filled pores in the membrane. Nearly all channel proteins are involved in transport of ions and may be referred to as *ion channels.* Ions can flow through the appropriate channel at very high rates (100 million ions/sec); this is much faster than carrier-mediated transport. However, channels are not linked to an energy source, so ions must flow passively down an electrochemical gradient. The channel proteins in the plasma membranes of animal cells are highly selective, permitting only a particular ion or class of ions to pass. Humans have about 400 genes that encode channel proteins. Ion channels are particularly important in allowing the cell to respond rapidly to a variety of external stimuli. Most channels are not continuously open, but they open and close according to membrane signals. Ion channels may be stimulated to open or close in three principal ways: (1) *voltage-gated* channels respond to a change in membrane potential; (2) *mechanically gated* channels respond to mechanical deformation; and (3) *ligand-gated* channels respond to the binding of a signaling molecule (a hormone or neurotransmitter) to a receptor on the cell surface (Fig. 3.25). In addition, some channels open without apparent stimulation and are referred to as *leak* channels. Ion channels are responsible for the development of membrane potentials and are of vital importance in nerve and muscle function, as discussed in the next section.

KEY POINTS

- Large, lipid-insoluble molecules are transported across the plasma membrane by endocytosis and exocytosis.
- Small, lipid-insoluble molecules are transported across the plasma membrane by three kinds of membrane proteins: adenosine triphosphate (ATP)-driven pumps, carriers, and channels.
- Pumps use the energy of ATP to move solutes against a gradient. Examples of ATP-driven active transport include proton pumps, Na^+–K^+ pumps, Ca^{2+} pumps, and ATP-binding cassette (ABC) transporters.
- Carriers may be passive or use the Na^+ gradient for secondary active transport. Neurotransmitter reuptake carriers and those that transport glucose and amino acids across the gut and renal tubules are examples of Na^+-driven carriers. Passive carriers include those that allow glucose entry into insulin-sensitive cells.
- Channels are always passive and allow ions to move down their electrochemical gradients when open. Channels open and close in response to specific signals, such as voltage changes, ligand binding, and mechanical pressure.

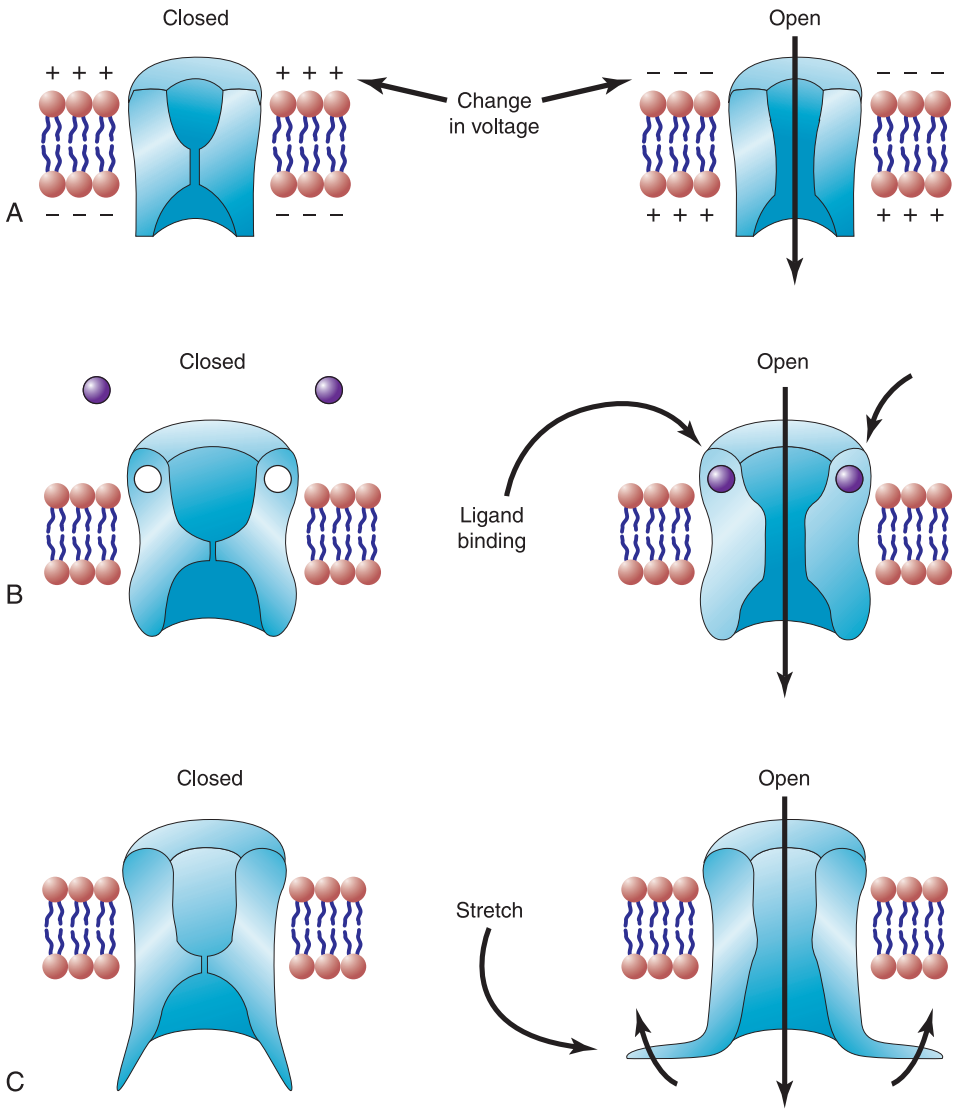


FIG 3.25 Gating of ion channels. **A,** Voltage-gated channel. **B,** Ligand-gated channel. **C,** Mechanically gated channel.

Cellular Membrane Potentials

Animal cells typically have a difference in the electrical charge across the plasma membrane. There is a slight excess of negative ions along the inner aspect of the membrane and extra positive ions along the outer membrane. This separation of charges creates a membrane potential that can be measured as a voltage. Positive and negative ions separated by the plasma membrane have a strong attraction to one another that can be used by the cell to perform work, such as the transmission of nerve impulses. A relatively large membrane potential is created by the separation of a very small number of ions along the membrane (Fig. 3.26).

Resting Membrane Potential

When there is no net ion movement across the plasma membrane, the electrical charge present inside the cell is called the *resting membrane potential* (RMP). The major determinant of the RMP is the difference in potassium ion concentration across the membrane. The concentration of potassium inside the cell is much greater (about 30 times greater) than the extracellular potassium concentration. At rest, the membrane is permeable to K^+, but not to other positively charged cations, including Na^+ and Ca^{2+}. Potassium ions remain inside the cell because of the attraction of fixed intracellular anions (negatively charged organic molecules such as proteins and phosphates that cannot diffuse out of the cell). Because the cell membrane is impermeable to Na^+ and Ca^{2+}, only K^+ is available to balance these negative intracellular ions. Thus two opposing forces are acting on the potassium ion. The negative cell interior attracts K^+ into the cell, whereas the huge K^+ concentration gradient favors movement of K^+ out of the cell. When the cell is at rest and not transmitting impulses, these forces are balanced, and although the membrane is permeable to K^+ there is no net movement. The voltage required to exactly balance a given potassium concentration gradient can be calculated from the ratio of extracellular K^+ concentration using the Nernst equation:

$$mV = 61 \log \left(K^+ \text{extracellular}/K^+ \text{intracellular}\right.$$

The measured membrane potential is quite close to that predicted mathematically and varies directly with changes in extracellular K^+ ion concentration. For example, a typical nerve cell has a normal resting potential of about −85 mV. If the extracellular K^+ level is increased, more K^+ ions will stay in the cell because of the reduced concentration gradient. These extra positive intracellular ions will neutralize more of the negative cellular anions, and the cell will *hypopolarize,* or become less negative. Conversely, if extracellular K^+ levels fall, more K^+ will exit the cell due to a greater concentration gradient. Fewer intracellular anions will be neutralized, and the cell interior will become more negative, or *hyperpolarized* (Fig. 3.27). Changes in RMP can have profound effects on the ease of action potential generation in cardiac and nerve cells.

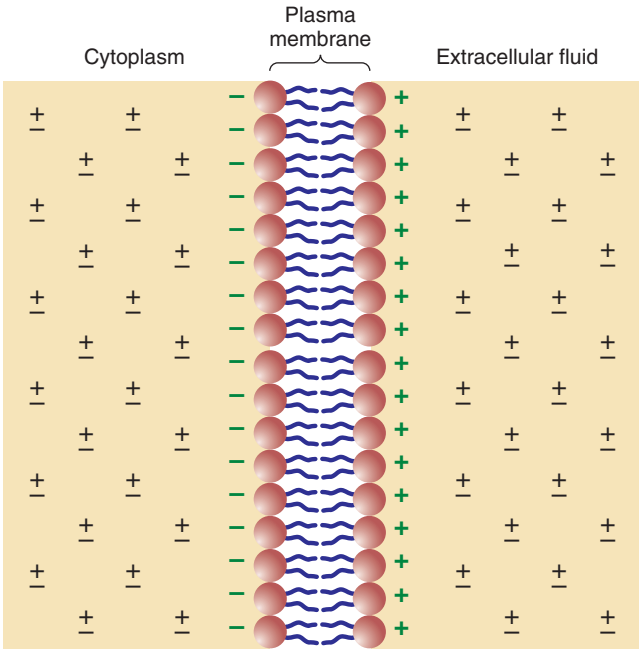


FIG 3.26 A relatively large membrane potential results from the separation of a very small number of ions across the plasma membrane.

The RMP is described by the potassium equilibrium potential because the cell is relatively impermeable to other ions at rest. Under certain conditions, the membrane may become highly permeable to an ion other than potassium. The membrane potential will reflect the equilibrium potential of the most permeant ions.

Long-term maintenance of ion gradients across the cell membrane is accomplished primarily by the Na^+–K^+ pump. The Na^+–K^+ pump also contributes to the negative RMP in that it extrudes *three* Na^+ for every *two* K^+ brought into the cell. However, this pump can be inhibited for minutes to hours in some tissues with little immediate effect on the resting membrane potential.

Action Potential

Nearly all animal cells have negative RMPs, which may vary from −20 to −200 mV, depending on the cell type and organism. The cell membranes of some specialized cell types, mainly nerve and muscle, are capable of rapid changes in their membrane potentials. These cells are electrically "excitable" and can generate and propagate action potentials. In classic experiments, action potentials were determined to be rapid, self-propagating electrical excitations of the membrane that are mediated by ion channels that open and close in response to changes in voltage across the membrane (voltage-gated ion channels). An action potential is triggered by membrane depolarization.

In nerve and muscle cells, the usual trigger for depolarization is binding of an excitatory neurotransmitter to cell surface receptors. Transmitter binding causes channels in the membrane to open, allowing ions (primarily Na^+) to enter the cell. This influx of positive ions causes a shift in the membrane potential to a less negative value, resulting in depolarization. *Threshold* is reached when a patch of the membrane becomes sufficiently depolarized (approximately −65 mV in animal neurons) to activate voltage-gated sodium channels in the membrane. At threshold, these channels open rapidly and transiently to allow the influx of Na^+ ions. A self-propagating process follows whereby Na^+ influx in one patch of membrane causes membrane depolarization of the next patch and opens more voltage-gated Na^+ channels, allowing more Na^+ to enter the cell. This process is repeated many times while the action potential proceeds along the length of the cell (Fig. 3.28). In this way, action potentials can transmit information rapidly over relatively long distances.

A typical neuronal action potential is shown in Fig. 3.29. The various changes in membrane potential during the time course of the action potential are attributable to the flow of ions through membrane ion channels. The steep upstroke of the action potential corresponds to Na^+ influx through "fast" sodium channels, as described previously. *Fast channels* are so termed because they open and close rapidly, with the entire process lasting less than 1 msec. This phase of rapid depolarization is terminated when the fast Na^+ channels suddenly close and the repolarization phase begins. Fast Na^+ channels are interesting in that they can assume at least three conformations (three-dimensional forms). In addition to the open and closed conformations, the fast Na^+ channel has a refractory form during which the channel will not reopen in response to another depolarizing stimulus (Fig. 3.30). This refractory period limits the rate at which action potentials can be generated.

Two major factors contribute to cellular repolarization: sodium conductance (inflow) is stopped by closing Na^+ channels, as described

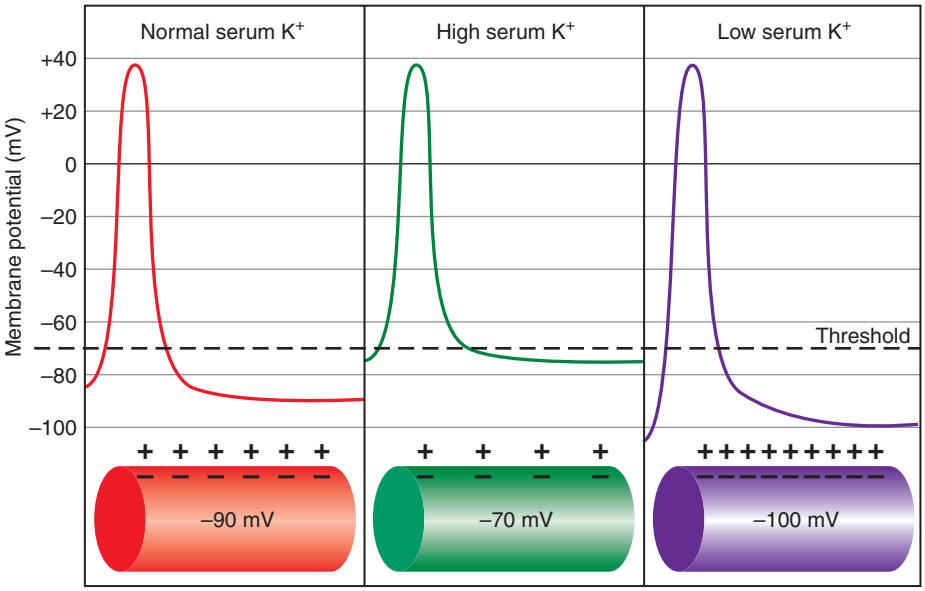


FIG 3.27 Effects of changes in extracellular K^+ level on the resting membrane potential. A high level of serum K^+ results in a hypopolarization of the membrane. A low serum K^+ level results in membrane hyperpolarization. With high serum K^+ levels, the resting membrane potential is closer to threshold, making it easier to achieve an action potential. A low serum K^+ level moves the resting membrane potential away from threshold, making it more difficult to achieve an action potential.

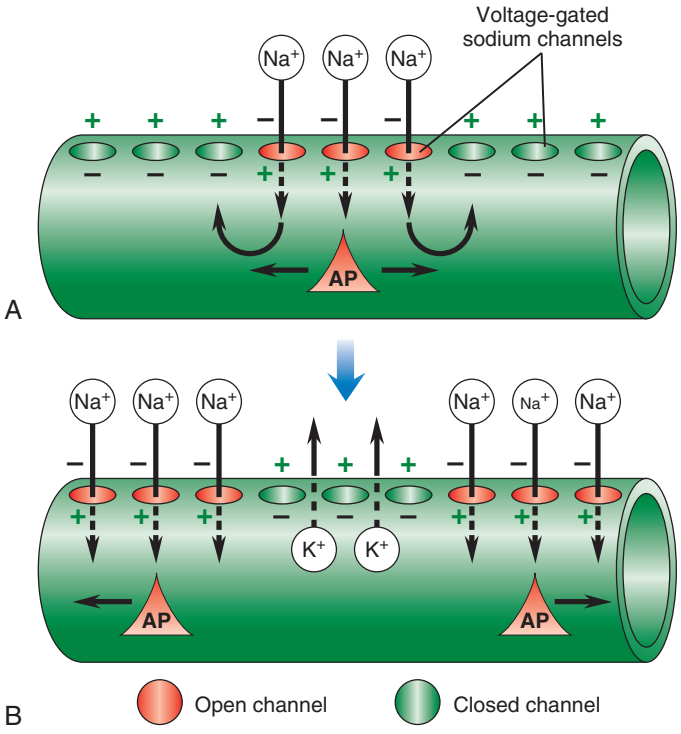


FIG 3.28 The action potential (AP) in excitable cells is propagated along the membrane by the sequential opening of voltage-gated sodium channels in adjacent sections of membrane. **A,** An action potential is initiated by the opening of sodium channels in a section of membrane. **B,** The action potential is regenerated in adjacent sections of membrane as more sodium channels open. The initial segment repolarizes as sodium channels close and potassium ions move out of the cell.

previously, and K^+ conductance (outflow) through K^+ channels increases. Although cells are always permeable to K^+, during repolarization additional voltage-gated K^+ channels open allowing a higher rate of K^+ efflux. These K^+ channels respond to depolarization of the membrane in the same manner as fast Na^+ channels, but they take much longer to open and close. When K^+ channels open, K^+ flows out of the cell because of the concentration gradient and the loss of intracellular negativity that accompanies Na^+ influx. The outward flow of positive intracellular potassium ions helps to quickly return the membrane potential to its negative RMP value.

Action potentials in cardiac muscle cells are more complex than the neuronal ones just described. Recall that contraction depends on the presence of free intracellular calcium ions. Because Ca^{2+} carries a charge, its entry into the cell cytoplasm is reflected in the membrane potential. In skeletal muscle, most of the free cytosolic calcium ions come from intracellular stores (sarcoplasmic reticulum) that are released when the cell is depolarized. In cardiac muscle cells, Ca^{2+} entry through voltage-gated channels in the plasma membrane is also important. Calcium conductance into the cell tends to prolong the action potential, resulting in a plateau phase (Fig. 3.31). This is of functional importance in cardiac tissue because it allows time for muscular contraction before another impulse is conducted and prevents the potentially disastrous condition of cardiac muscle tetany. (For a more thorough discussion of cardiac electrophysiology, see Chapter 17.)

KEY POINTS

- The negative value of the resting membrane potential (RMP) is determined by the ratio of intracellular to extracellular K^+ ion concentration. Changes in serum K^+ concentration can have profound effects on the RMP.
- Cells with voltage-gated ion channels are excitable and can produce and conduct action potentials. An action potential results from the opening of "fast" Na^+ channels, which allows Na^+ to rush into the cell.
- Repolarization is caused by closure of Na^+ channels and efflux of K^+ from the cell. In cardiac muscle, repolarization is prolonged due to Ca^{2+} influx through "slow" Ca^{2+} channels.

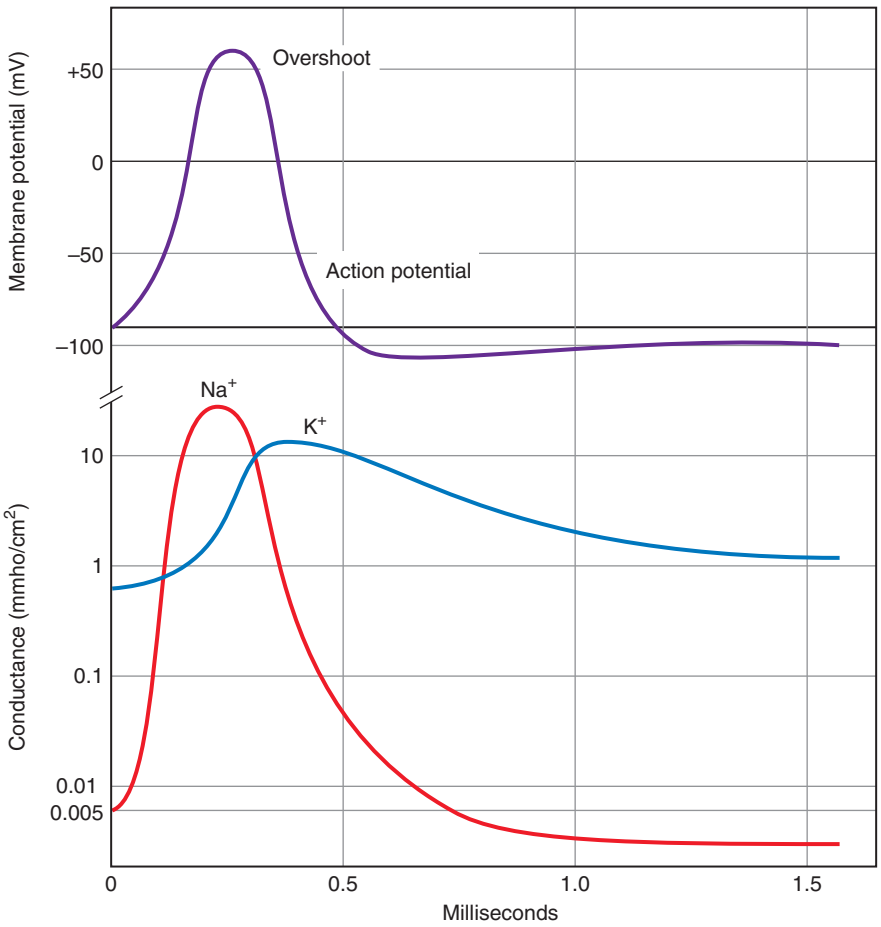


FIG 3.29 A typical neuronal action potential showing changes in membrane potential and the associated ion conductances. Note: mmho is a measure of conductance (amperes per volt), also called millisiemens (mS). The steep upstroke of the action potential is attributed to the sudden influx of Na^+ through voltage-gated "fast" sodium ion channels. Voltage-gated K^+ channels open more slowly and stay open longer to allow K^+ efflux from the cell, which aids in repolarization.

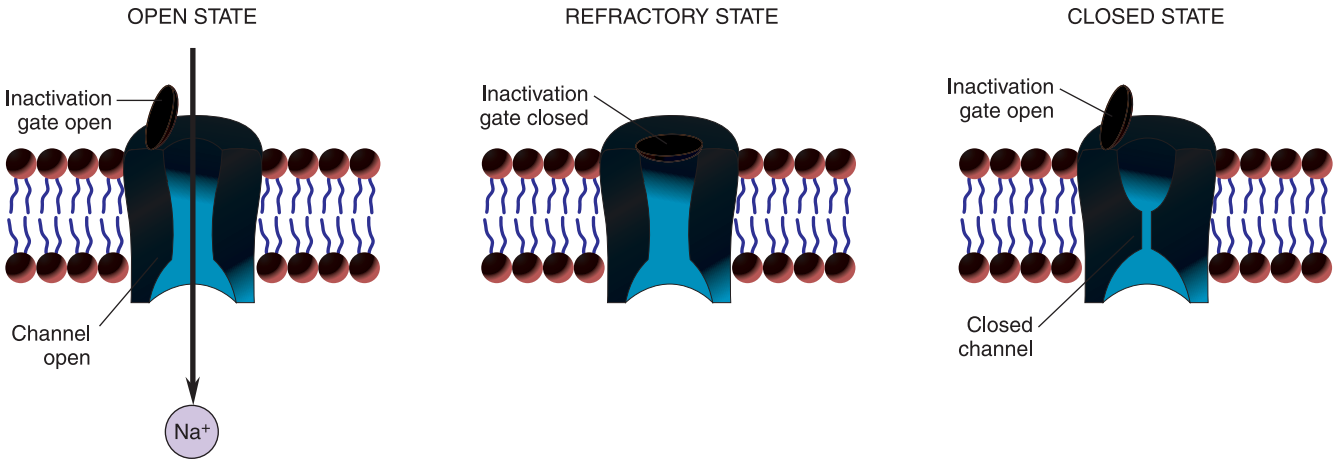


FIG 3.30 Three possible states of the voltage-gated sodium channel. In the open state, Na^+ is allowed to pass. In the refractory state, the channel is blocked by the inactivation gate and will not open in response to a depolarizing stimulus. In the closed state, the channel will open in response to a membrane depolarization.

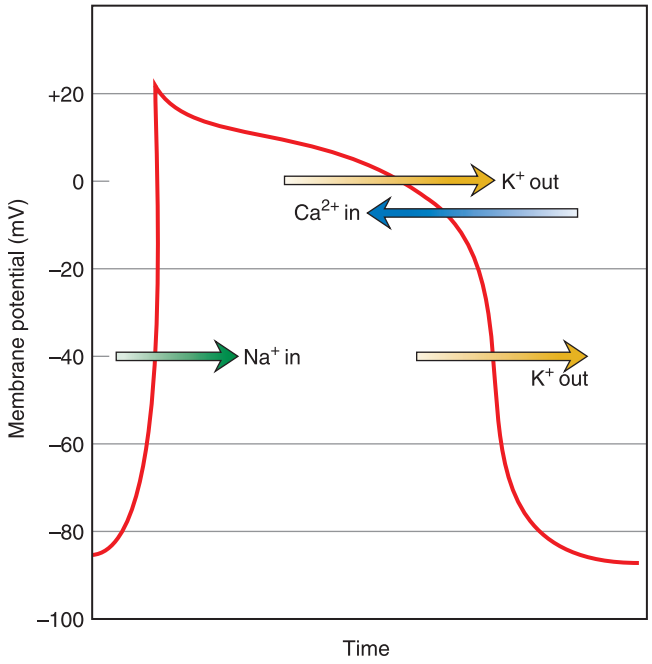


FIG 3.31 A typical cardiac muscle cell action potential showing the ion fluxes associated with each phase. Note that the repolarization phase is prolonged in comparison to the nerve action potential in Fig. 3.28. This occurs because Ca^{2+} influx offsets the repolarizing effect of K^+ efflux and a plateau in the membrane potential is seen. When the Ca^{2+} channels close, the membrane quickly repolarizes.

INTERCELLULAR COMMUNICATION AND GROWTH

Cell Signaling Strategies

Cells in multicellular organisms need to communicate with one another and respond to changes in the cellular environment. Coordination of growth, cell division, and the functions of various tissues and organ systems are accomplished by three principal means of communication: (1) through gap junctions that directly connect the cytoplasm of adjoining cells; (2) by direct cell-to-cell contact of plasma membranes or the extracellular molecules associated with the cell (extracellular matrix); and (3) by secretion of chemical mediators (ligands) that influence cells some distance away (Fig. 3.32).

Gap junctions are found in many tissues. They are connecting channels between adjacent cells that allow the passage of small molecules from one cell to the next. These junctions are formed by special transmembrane proteins called *connexins* that associate to form pores of about 1.5 nm in width. Small molecules, such as inorganic ions, glucose, amino acids, and vitamins, may pass through the pores, whereas macromolecules (e.g., proteins, polysaccharides, and nucleic acids) are too large to pass through pores. Gap junctions are particularly important in tissues in which synchronized functions are required, such as cardiac muscle contraction, vascular tone, and intestinal peristaltic movements. Gap junctions appear to be important in embryogenesis as well. Cellular differentiation may be mediated in part through chemical signaling through gap junctions. (See Chapter 5 for a discussion of the development and differentiation of tissue types.)

Direct contact of cell membrane receptors with signaling molecules present on the surface of other cells or extracellular matrix is an important means of local communication among cells in tissues. Contact-dependent signaling is particularly important for the development of the immune

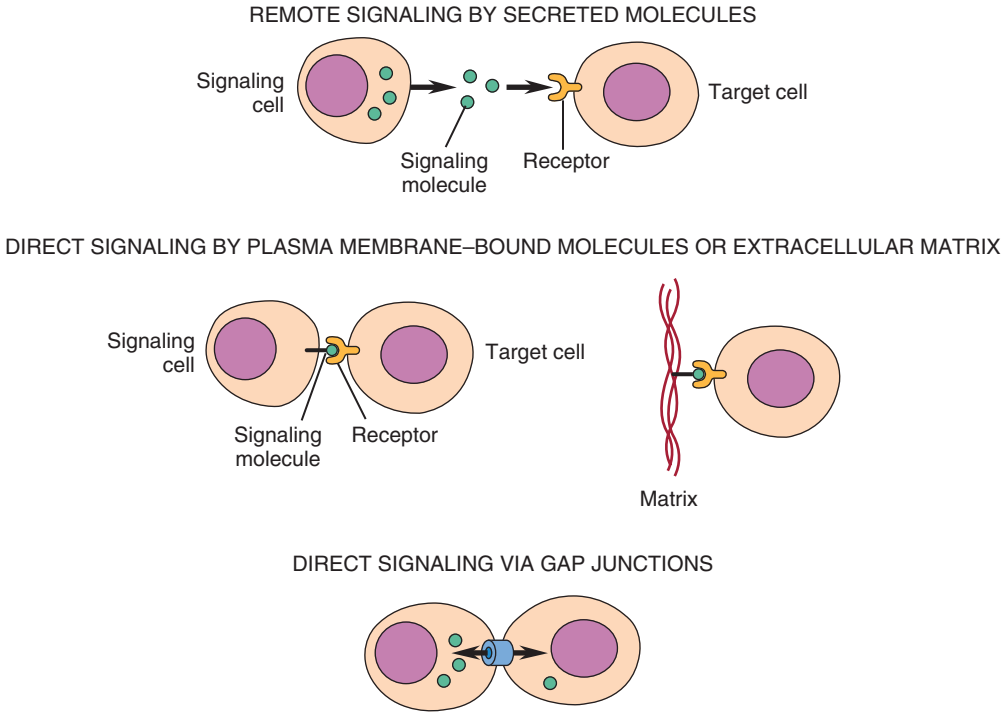


FIG 3.32 Methods used for intercellular communication.

response. Such cell-to-cell contact during fetal development is thought to allow the cells of the immune system to discriminate between foreign and self tissues and to develop self-tolerance. If cell-to-cell contact does not occur during fetal life, the immune cells may later attack the body's own cells, leading to the development of autoimmune diseases. (See Chapter 10 for a discussion of autoimmunity.) There are four major families of cell adhesion molecules (CAMs): immunoglobulin–cell adhesion molecules (Ig-CAMs), cadherins, integrins, and selectins. These cell adhesion proteins make contacts between cells and with the extracellular matrix and provide signals that maintain cell survival and differentiated cell types (Fig. 3.33).

The best understood form of cell communication is signaling through secreted molecules, or *ligands.* Three strategies of intercellular chemical signaling have been described, relating to the distances over which they operate (Fig. 3.34). *Synaptic* signaling is confined to the cells of the nervous system and occurs at specialized junctions between the nerve cell and its target cell. The neuron secretes a chemical neurotransmitter into the space between the nerve and target cell; the neurotransmitter then diffuses across this synaptic cleft and binds receptors on the postsynaptic cell. Synaptic signaling occurs over very small distances (50 nm) and involves only one or a few postsynaptic target cells. In *paracrine* signaling, chemicals are secreted into a localized area and are rapidly destroyed, so that only cells in the immediate area are affected. Growth factors (GFs), for example, act locally to promote wound healing without affecting the growth of the entire organism. *Endocrine* signaling is accomplished by specialized endocrine cells that secrete hormones that travel via the bloodstream to target cells widely distributed throughout the body. Endocrine signaling is slow in comparison to nervous signaling because it relies on diffusion and blood flow to target tissues.

A fourth type of signaling, *autocrine* signaling, occurs when cells are able to respond to signaling molecules that they secrete. Autocrine communication provides a feedback signal to the secreting cell and is commonly linked to pathways that regulate ligand secretion rates. Abnormal autocrine stimulation is thought to be a mechanism in some forms of cancer (see Chapter 7).

Target cells respond to ligand signaling through specific protein *receptors.* Cells can respond to a particular ligand only if they possess the appropriate receptor. For example, all cells of the body are exposed to thyroid-stimulating hormone (TSH) as it circulates in the blood, but only thyroid cells respond because they alone possess TSH receptors. However, cells that possess the same receptor may respond very differently to a particular ligand. For example, binding of acetylcholine to its receptor on a glandular cell may induce secretion, whereas binding to the same receptor on a cardiac muscle cell causes a decrease in contractile force. The cellular response to signaling molecules is regulated both by the array of receptors the cell carries and by the internal machinery to which the receptors are linked.

Cell Surface Receptor–Mediated Responses

Most hormones, local chemical mediators, and neurotransmitters are water-soluble molecules that are unable to pass through the lipid bilayer of the cell. These ligands exert their effects through binding with a receptor on the surface of the target cell, which then changes or transduces the external signal into an intracellular message. There are three major classes of cell surface receptor proteins: ion channel–linked, enzyme-linked, and G-protein–coupled (Fig. 3.35).

Ion channel–linked receptors bind neurotransmitters, causing specific ion channels in the membrane to open or close. This type of signaling is prevalent in the nervous system, where rapid synaptic signaling between neurons is required. *Enzyme-linked* receptors catalyze enzyme reactions when they are activated by appropriate ligands. Nearly all enzyme-linked receptors function as *protein kinases;* that is, they mediate the transfer of phosphate groups from ATP (or GTP) to proteins (phosphorylate), and thus affect the activity of those target proteins. The insulin receptor and most growth factor receptors are protein kinase receptors that phosphorylate and activate intracellular enzyme cascades. Enzyme-linked kinase receptors activate common kinase cascades, including the

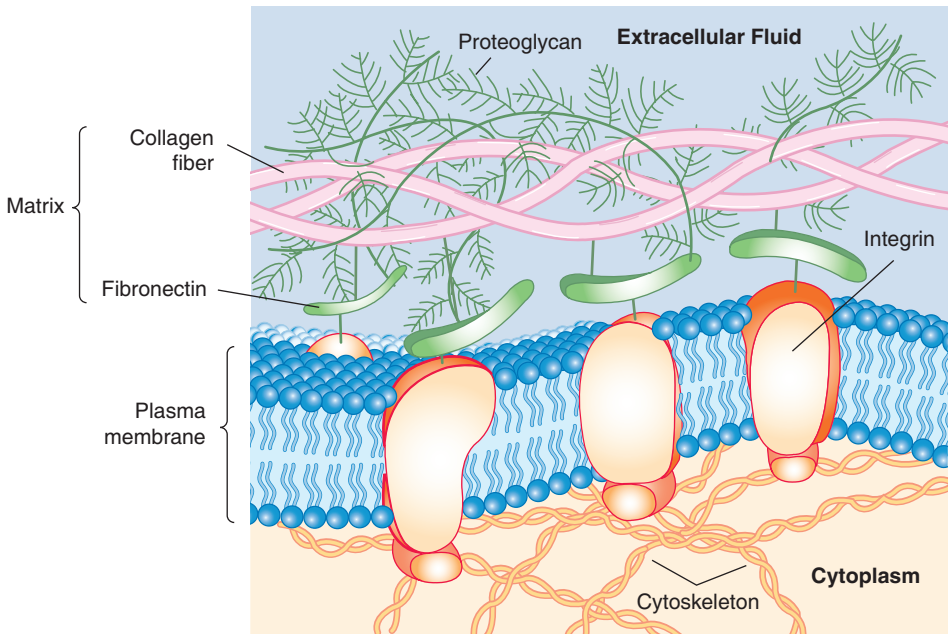


FIG 3.33 Cell adhesion proteins interact with the extracellular matrix (integrins) and with neighboring cells to maintain cell survival and differentiation. (Redrawn from Patton KT, Thibodeau GA: *Anatomy & physiology*, ed 9, St. Louis, 2016, Mosby, p. 141.)

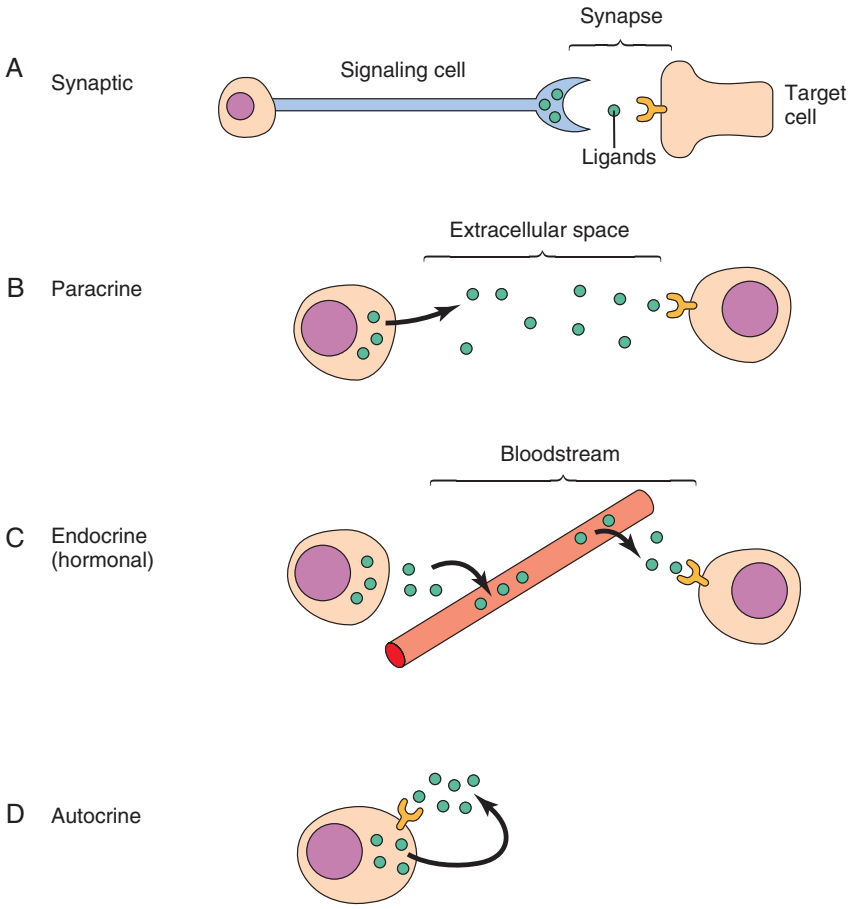


FIG 3.34 Signaling by secreted ligands can occur over variable distances. **A,** Synaptic signaling over a very small distance between neuron and target cell. **B,** Paracrine signaling through the extracellular fluid between cells in a tissue. **C,** Long-range signaling from endocrine cells through the bloodstream to distant targets. **D,** Localized autocrine signaling in which the secreting cell is also the target cell.

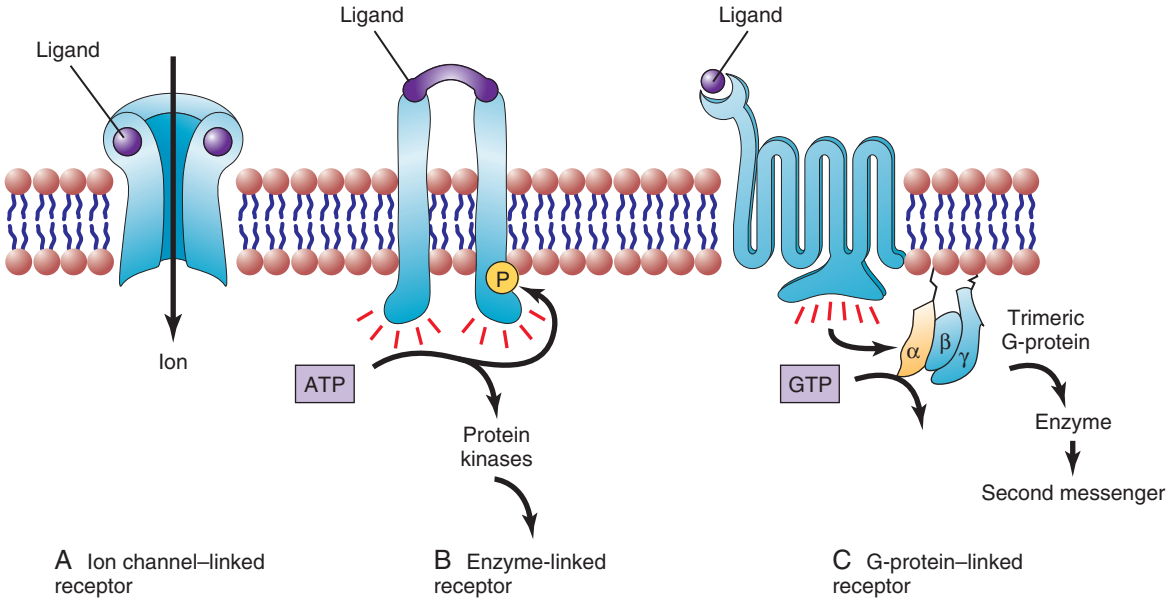


FIG 3.35 There are three major types of cell surface receptor proteins. **A,** Ion channel–linked receptors are also called *ligand-gated channels*. When the ligand binds, they open to allow specific ions through the membrane. **B,** Enzyme-linked receptors become activated kinases when a ligand binds to them. Kinases phosphorylate target proteins and change their activity. **C,** G-protein–linked (coupled) receptors have seven membrane-spanning segments with a ligand-binding pocket on the outside and a G-protein–activating portion on the inside. G-protein–linked receptors activate G-proteins, which in turn influence enzymes that produce second messengers.

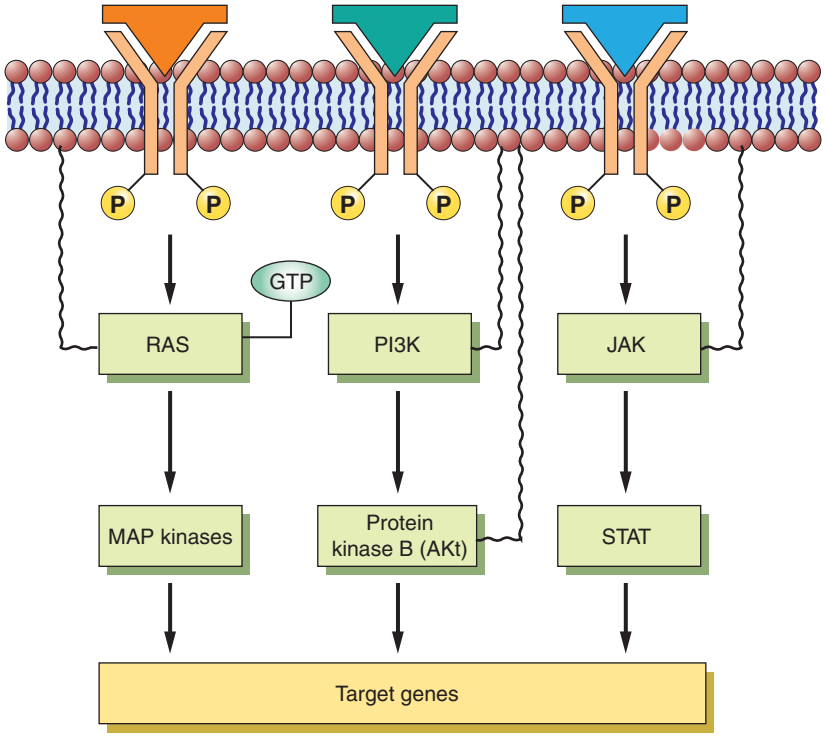


FIG 3.36 Many growth factor receptors activate protein kinase cascades within the cell. Three common pathways are shown. After binding of ligand, the receptor dimerizes and becomes phosphorylated. A cascade of kinase activations is initiated resulting in a change in target gene transcription. *GTP,* Guanosine triphosphate; *JAK,* janus kinase; *MAP,* mitogen-activated kinase; *PI3K,* phosphoinositide 3-kinase; *RAS,* rat sarcoma protein; *STAT,* signal transducer and activator of transcription.

PI3K-protein kinase B pathway, the RAS-MAP kinase pathway, and the JAK-STAT pathway (Fig. 3.36).

A large number of signaling ligands bind to *G-protein–coupled receptors* (GPCRs). Most hormones and many drugs have their effects through G-protein–linked cascades. G-protein–coupled receptors act indirectly through a membrane-bound trimeric G-protein that binds GTP when activated by the receptor. The activated α subunit of the trimeric G-protein influences the activity of specific target enzymes. The target enzymes of G-proteins produce second messengers that trigger specific intracellular cascades and alter cell function (Fig. 3.37). The α subunit of G-proteins has intrinsic enzyme activity that degrades GTP into GDP and P_i after a time. When GTP is bound, the G-protein is in the right conformation to activate its downstream targets, but when GTP is hydrolyzed to GDP and P_i, the G-protein resumes its inactive conformation and the activity of the signaling cascade is terminated.

There are three principal G-protein–coupled signaling systems that, when activated, alter the intracellular concentration of one or more second messengers (see Fig. 3.37). Numerous receptors activate trimeric G-proteins whose α subunit stimulates adenylyl cyclase to produce the second messenger cyclic adenosine monophosphate (cAMP). These G-proteins are called *Gs*. An increase in cAMP concentration is linked to different signaling cascades in different cell types. For example, cAMP causes glycogen breakdown in liver cells, increased force of contraction in cardiac cells, and increased secretion by glandular cells. Various cell types respond differently to the same second messenger because of differences in enzymes and other proteins in the cell.

Another important G-protein–coupled cascade is mediated by G-proteins called *Gq* whose α subunit stimulates the enzyme phospholipase C. Phospholipase C cleaves a membrane phospholipid (PI[4,5]P_2) to form two second messengers: inositol 1,4,5-trisphosphate (IP_3) and diacylglycerol (DAG) (see Fig. 3.37). The IP_3 travels to the ER, where it stimulates the release of Ca^{2+} into the cytoplasm. The Ca^{2+} then triggers a change in cell function. DAG remains bound to the inner surface of the plasma membrane and can trigger several different intracellular cascades. Two important targets are the protein kinase C pathway and the eicosanoid pathway. Protein kinase C is a key enzyme in the growth response. The eicosanoid pathway results in the production of several arachidonic acid derivatives, including prostaglandins. These products are often secreted by the cell as signaling molecules to other nearby cells. Prostaglandins are important mediators of inflammation and platelet function.

The third trimeric G-protein type is called *Gi* because it is inhibitory to the production of cAMP. GPCRs such as the acetylcholine receptor in the heart activate Gi, whose α subunit then inhibits adenylyl cyclase (see Fig. 3.37). In this case the γβ subunit of Gi is also activated and opens membrane potassium channels in the heart, which tend to slow the heart rate. Although Gs, Gq, and Gi are the primary trimeric G-protein signaling cascades, others have been described (Table 3.1).

In addition to the four second messengers already mentioned (cAMP, IP_3, DAG, and Ca^{2+}) there is a fifth called *cyclic guanosine monophosphate* (cGMP), which is produced by the enzyme guanylyl cyclase (Fig. 3.38). The primary activator of guanylyl cyclase is a small lipid-soluble gas molecule called *nitric oxide.* Nitric oxide is an important signaling molecule with widespread targets. It functions as a neurotransmitter in the brain and is an important smooth muscle relaxant in the vascular system. cGMP is also produced by a special class of enzyme-linked receptors (see Fig. 3.38).

To be effective at communicating signals, all the receptor systems must be quickly turned off so that they can be responsive to the next incoming signal. A variety of strategies are used to quench the signaling

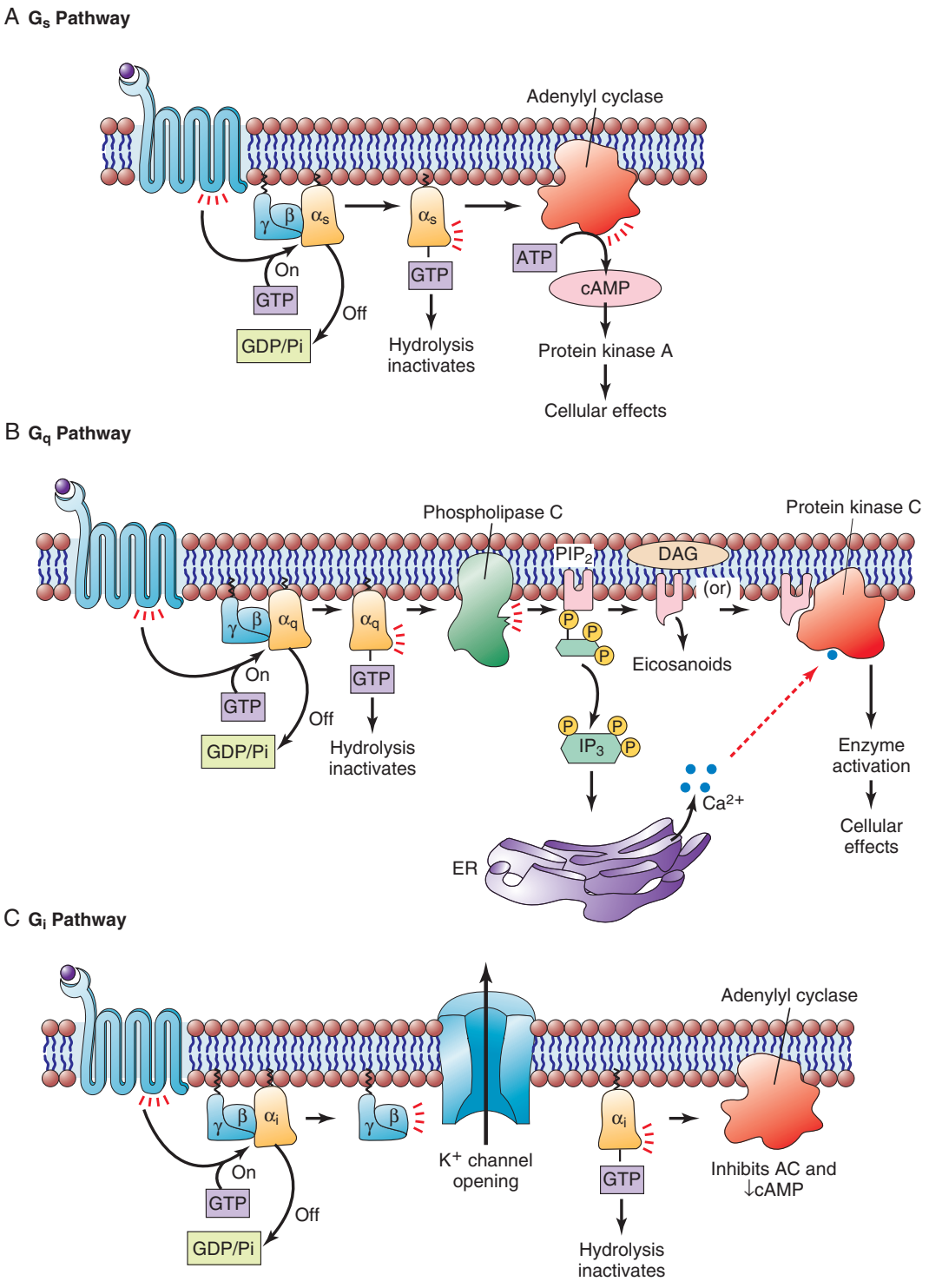


FIG 3.37 G-protein–coupled signaling. When the ligand binds to the receptor, an intracellular domain is changed into an active configuration that can interact with inactive trimeric G-proteins. The receptor induces the G-protein to release its bound GDP and Pi in exchange for a GTP molecule. When GTP binds to the α subunit of the G-protein, it is activated and diffuses away from the γβ subunits to find its target enzyme (adenylyl cyclase [AC] or phospholipase C). The α GTP stimulates its target enzyme to produce a second messenger, which in turn activates a signaling cascade within the cell. After a time, the α subunit hydrolyzes its GTP to GDP and Pi and becomes inactive. The α subunit is now in the correct conformation to reassociate with the γβ subunits and await another signal from the receptor. **A,** The Gs pathway increases the production of cyclic adenosine monophosphate (cAMP). **B,** The Gq pathway increases the production of inositol 1,4,5-trisphosphate (IP_3) and diacylglycerol (DAG). **C,** The Gi pathway is inhibitory to the production of cAMP. In some cases the γβ subunit also has functional activity and may regulate ion channels. *ER,* Endoplasmic reticulum; *PKC,* protein kinase C.

TABLE 3.1 Four Major Families of Trimeric G Proteins

Family*	Some Family Members	Subunits That Mediate Action	Some Functions
I	Gs	α	Activates adenylyl cyclase; activates Ca^{2+} channels
	Golf	α	Activates adenylyl cyclase in olfactory sensory neurons
II	Gi	α	Inhibits adenylyl cyclase
		βγ	Activates K^+ channels
	Go	βγ	Activates K^+ channels; inactivates Ca^{2+} channels
		α and βγ	Activates phospholipase C-β
	Gt (transducin)	α	Activates cyclic GMP phosphodiesterase in vertebrate rod photoreceptors
III	Gq	α	Activates phospholipase C-β
IV	G12/13	α	Activates Rho family monomeric GTPases (via Rho-GEF) to regulate the actin cytoskeleton

*Families are determined by amino acid sequence relatedness of the α subunits. Only selected examples are included. About 20 α subunits and 11 γ subunits have been described in humans.

Alberts B, et al: Cell signaling. In Alberts B, et al: *Molecular biology of the cell,* ed 6, New York, 2015, Garland Science, p 846.

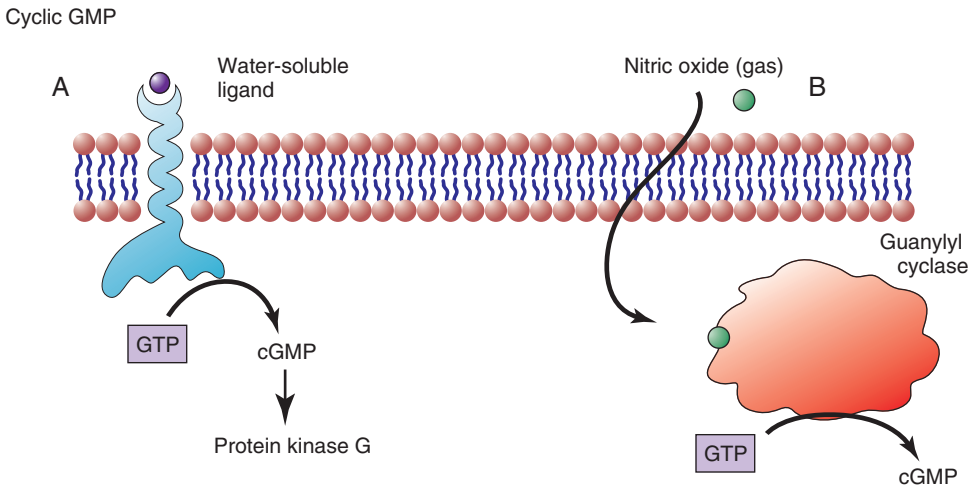


FIG 3.38 Cyclic GMP (cGMP) is an important second messenger. **A,** It can be synthesized by enzyme-linked receptors that are activated by water-soluble ligands such as atrial natriuretic peptide. **B,** Nitric oxide is an important signaling molecule that is lipid soluble and can diffuse across the cell membrane. Nitric oxide binds to and stimulates the enzyme guanylyl cyclase to produce cGMP.

cascades (Fig. 3.39). For example, phosphodiesterases are enzymes that convert the cyclic nucleotides cAMP and cGMP to their inactive forms, AMP and GMP, respectively, and help to remove these second messengers soon after they are formed. Some drugs, such as caffeine and sildenafil citrate (Viagra), are phosphodiesterase inhibitors that slow the normal breakdown of cyclic nucleotides and prolong their activity. Many of the intracellular signaling cascades rely on kinases that phosphorylate their target proteins and change their activity. The action of kinases is countered by numerous phosphatase enzymes that quickly cleave the phosphates off the target proteins and inhibit their activity.

The cell also can regulate the activity and number of receptors on the cell surface. Generally a cell decreases the number or activity of receptors when it is exposed to excessive concentrations of signaling molecules (see Fig. 3.39). Receptors can be internalized in the cell where they are inactive but are available for later use, or they can be sent to lysosomes for degradation. Destruction of receptors in lysosomes is called *down-regulation.* (The production of extra receptors is called *up-regulation.*) Receptors that remain in the membrane also can be inhibited by phosphorylation, which blocks them from interacting with their intracellular targets. Receptors that can bind ligand but do not produce a response are said to be uncoupled. The proteins that phosphorylate G-protein receptors are called G-protein–receptor kinases (GRKs). The mechanisms that "turn off" signaling cascades are vitally important to maintaining a responsive communication system.

Intracellular Receptor–Mediated Responses

A small number of hormones are lipid soluble and can pass directly through the cell membrane to interact with receptors *inside* the cell. These receptors are located in the cell cytosol (e.g., cortisol) or may be associated with the cell nucleus. Intracellular receptors are specific for a particular ligand, just as surface receptors are. Binding of the ligand causes the receptor to become activated. Because lipid-soluble ligands enter the cell directly, no second messengers are needed. An activated cytosolic steroid receptor travels to the nucleus, where it binds with specific genes and regulates their activity (Fig. 3.40). Thyroid receptors are also located within the cell. Thyroid hormone enters the cell through carriers in the membrane and travels to the nucleus. The thyroid receptor is already bound to DNA in the absence of thyroid hormone. When thyroid hormone finds its nuclear receptor, the complex dissociates and removes an inhibitory influence on gene transcription. Cellular responses to these gene regulatory receptor complexes are slow in comparison to the cell surface receptor responses and generally last longer.

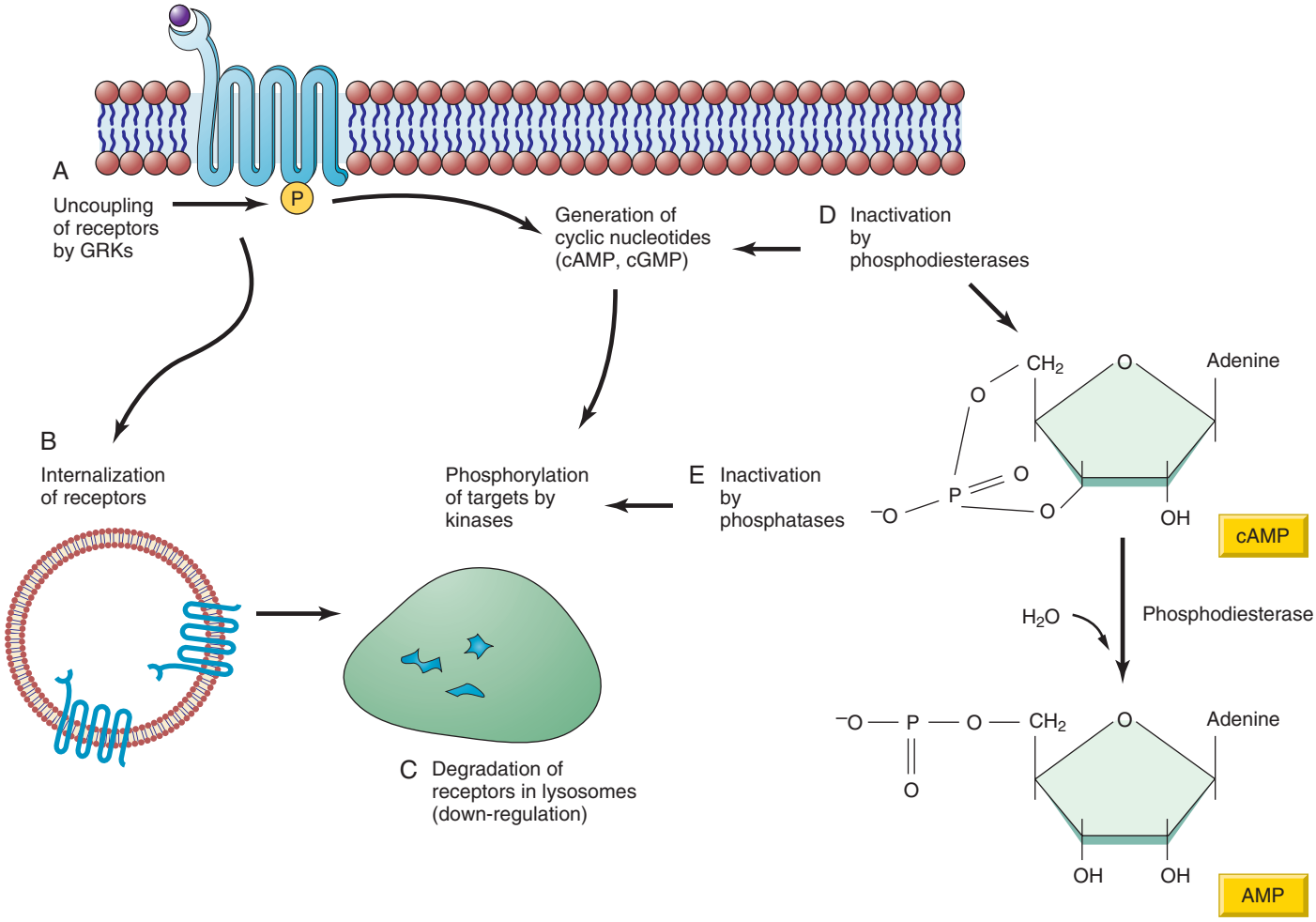


FIG 3.39 A variety of mechanisms exist to inhibit receptor-mediated signaling cascades. **A,** Phosphorylation of the receptor by receptor kinases such as G-protein receptor kinases (GRKs) uncouples the enzyme from its intracellular cascade. **B,** Receptor internalization temporarily reduces the number of receptors displayed at the cell surface. **C,** Receptor degradation results in a long-term reduction in receptors (down-regulation). **D,** The cyclic nucleotide second messengers can be degraded by phosphodiesterase enzymes to stop the intracellular cascade. **E,** Phosphatase enzymes counteract the phosphorylating activities of kinases and inhibit the intracellular cascade.

Regulation of Cellular Growth and Proliferation

In multicellular organisms such as humans, the growth and proliferation of cells and tissues must be strictly controlled to maintain a balance between cell birth rate and cell death rate. The system must be capable of rapidly increasing proliferation of a particular tissue to replace cells lost to injury and normal wear and tear while simultaneously inhibiting unwanted growth or proliferation of other cells. Special intercellular communication systems function to regulate the replication of individual cells in the body. Two important strategies of cell cycle control have been described. First, a variety of protein mitogens and growth factors are required in specific combinations for growth and proliferation of particular cell types. Second, cells respond to spatial signals from the extracellular matrix (from integrin receptors) and neighboring cells (from cell adhesion proteins) that indicate how much room is available. When conditions favor cell proliferation, the cell proceeds through the stages of the cell cycle (Fig. 3.41). Dormant cells can remain in G_1 phase indefinitely (called G_0). Cycling cells proceed through G_1, S phase (synthesis), G_2, M phase (mitosis), and cell division. S phase is characterized by duplication of DNA and synthesis of intracellular components in preparation for cell division. M phase, or mitosis, proceeds through six stages, beginning with prophase, in which the chromosomes condense and become visible, and ending with cytokinesis, when cell division is accomplished. The chromosomes of body cells are duplicated and distributed equally to the cell's progeny when it divides by mitosis, such that each daughter cell receives an identical full set of 46 chromosomes. The stages of mitotic cell division are explained in Fig. 3.42. Mitosis is responsible for the proliferation of body cells in which little genetic variation is needed or desired. A more elaborate cell division process, meiosis, occurs in the germ cells (egg and sperm), where significant chromosomal rearrangements occur (see Chapter 6).

The cell cycle has been the subject of intense study in recent years because of its importance in cancer biology. Cancer cells continue to grow and divide unchecked, despite the lack of appropriate signals to stimulate them. Of particular interest are the events that prod the cell from its dormant state and cause it to begin the cycle. A simplified picture of a major component of this complex process is shown in Fig. 3.43. The Rb protein (or pRb) is of central importance in preventing a cell from proceeding through the cell cycle. The Rb protein functions to bind gene transcription factors called *E2F* so that they are unable to

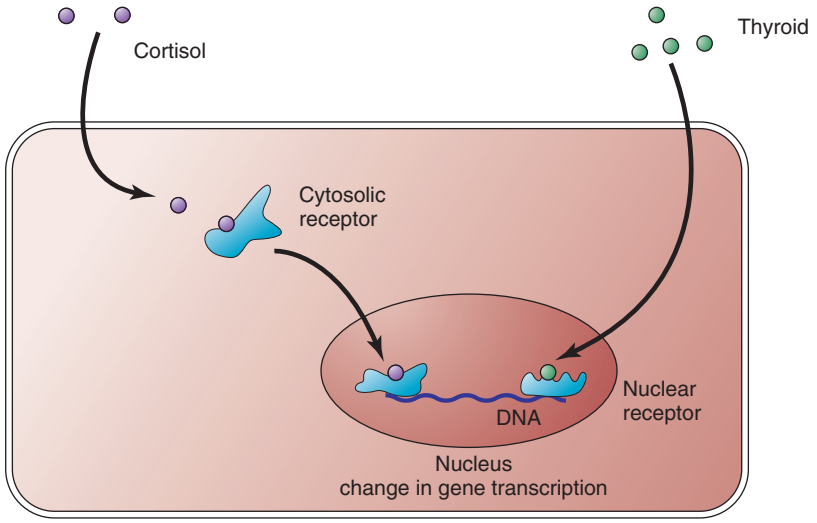


FIG 3.40 Lipid-soluble ligands, such as steroid hormones and gases, can diffuse across the cell membrane and interact with receptors located within the cell cytoplasm or nucleus. Thyroid hormone enters the cell through a carrier to interact with its intracellular receptor. When the ligand binds to its intracellular receptor, it forms a functional gene regulatory protein that affects the rate of transcription of its target genes. The response of the cell to intracellular ligands is generally slow and long lasting.

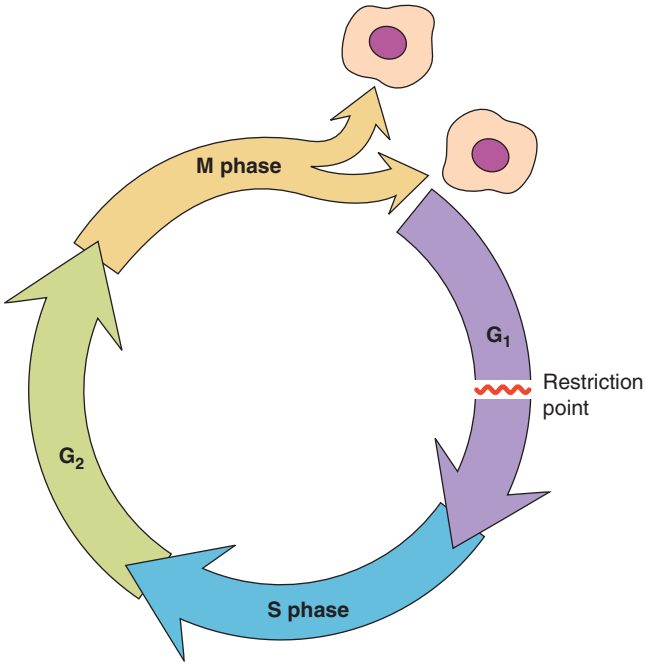


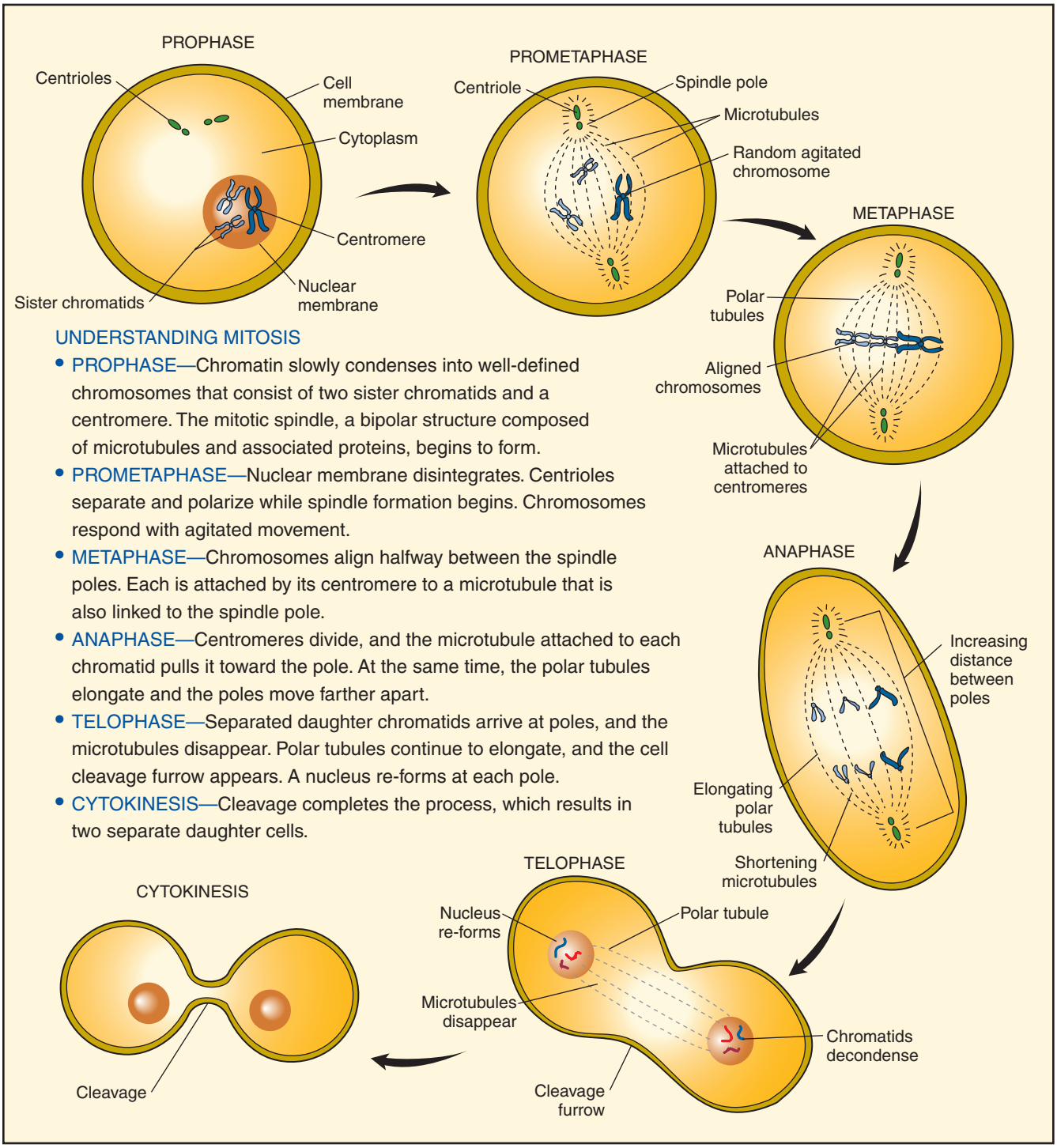
FIG 3.41 Events of the cell cycle. The cycle begins late in G_1 when the cell passes a restriction point. The cell then proceeds systematically through the S phase (synthesis), G_2, and M phase (mitosis).

bind to DNA promoter regions and begin the processes of cell replication. The Rb protein can be induced to release the E2F transcription factors when appropriate mitogen signals arrive at the cell surface. These proliferation-promoting signals at the cell surface are transmitted to pRb by way of cyclin-dependent signaling pathways within the cell. Proteins called *cyclins* accumulate in the cell and then bind to and activate cyclin-dependent kinases (Cdk). The Cdk then phosphorylates pRb, changing its affinity for E2F so that it is released. The E2F then translocates to specific regions of DNA where it regulates more than 500 genes and promotes cell replication.

To respond to a mitogen growth factor, a cell must have the corresponding receptor on its cell surface. Many cells in the body synthesize and secrete mitogens, which then influence the proliferation of other cell types in a paracrine or endocrine fashion. Platelet-derived growth factor (PDGF) was one of the first mitogens to be discovered. It is secreted by platelets when they form blood clots in response to an injury. PDGF stimulates fibroblasts and smooth muscle cells in the damaged area to divide and replace cells lost to the injury. Numerous mitogens have been identified, and most cells require an appropriate combination of mitogen signals before they can enter the cell cycle. There are many signaling steps in the pathway from mitogen receptor to DNA activation. Somatic cells respond to GFs by increasing cell size, whereas stem cell populations undergo cell division. Thus the same signaling ligands may have different effects depending on cell type and conditions. Similar signaling pathways may also trigger cell death (apoptosis) when cells have to be reduced or removed during tissue development and remodeling. The processes of abnormal cellular proliferation and cancer are further detailed in Chapter 7. The process of apoptosis is described in Chapter 4.

KEY POINTS

- Intercellular communication is accomplished by three principal means: (1) gap junctions, which directly connect the cytoplasm of adjoining cells; (2) direct cell-to-cell surface contact; and (3) secretion of chemical mediators (ligands). Most ligands are water-soluble molecules that interact with receptors on the cell surface. These receptors are of three general types: ion channel linked, enzyme linked, and G-protein coupled.
- Binding of a ligand to a G-protein receptor controls the production of second messengers (cAMP, IP_3, DAG, Ca^{2+}) within the target cell that initiate changes in cell function.
- Somatic cells divide by a process called *mitosis* in which daughter cells each receive an identical and complete set of 46 chromosomes.
- Cell replication normally requires specific extracellular mitogens that activate signaling systems within the cell. Cyclin proteins and Cdk alter the function of Rb protein, causing it to release transcription factors that begin the process of cell replication.



UNDERSTANDING MITOSIS

- PROPHASE—Chromatin slowly condenses into well-defined chromosomes that consist of two sister chromatids and a centromere. The mitotic spindle, a bipolar structure composed of microtubules and associated proteins, begins to form.
- PROMETAPHASE—Nuclear membrane disintegrates. Centrioles separate and polarize while spindle formation begins. Chromosomes respond with agitated movement.
- METAPHASE—Chromosomes align halfway between the spindle poles. Each is attached by its centromere to a microtubule that is also linked to the spindle pole.
- ANAPHASE—Centromeres divide, and the microtubule attached to each chromatid pulls it toward the pole. At the same time, the polar tubules elongate and the poles move farther apart.
- TELOPHASE—Separated daughter chromatids arrive at poles, and the microtubules disappear. Polar tubules continue to elongate, and the cell cleavage furrow appears. A nucleus re-forms at each pole.
- CYTOKINESIS—Cleavage completes the process, which results in two separate daughter cells.

FIG 3.42 Six stages of mitotic cell division. (Redrawn from Nichols FH, Zwelling E, editors: *Maternal-newborn nursing: theory and practice,* Philadelphia, 1997, Saunders, p 307.)

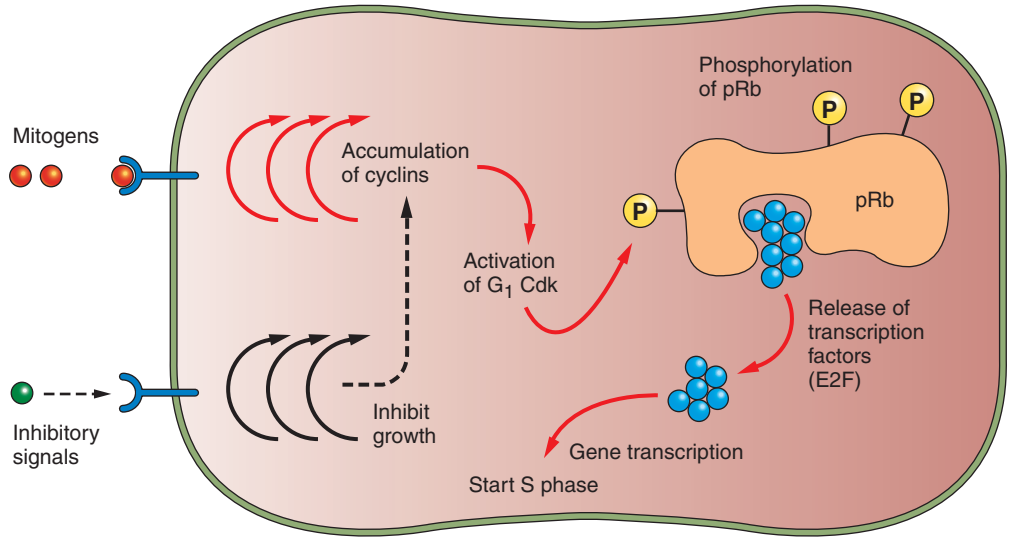


FIG 3.43 The mechanism of initiation of cellular replication requires appropriate stimulation by extracellular growth factors that bind their complementary receptors on the cell surface. Activation of the receptor stimulates signaling pathways within the cell that increase cyclin proteins. The cyclins bind to cyclin-dependent kinases (Cdks) to form active enzyme complexes. The active cyclin-Cdk enzymes phosphorylate Rb protein (pRb), inducing it to release E2F transcription factors that initiate replication. In the absence of appropriate growth factor signals, the Rb protein functions to inhibit unwanted cell proliferation.

SUMMARY

Detailed knowledge of cell physiology is essential to understanding disease processes. Cells are complex, membrane-bound units that perform a variety of functions necessary to the maintenance of life. The major cell components and their functions are summarized in Table 3.2. The cell membrane is an important cellular structure that protects the cell interior and mediates information transfer to and from the extracellular environment. Proteins embedded in the membrane lipid bilayer perform most of the membrane functions, including transduction of extracellular messages, membrane transport, electrical excitation, and cell-to-cell communication.

Human cells have several important intracellular organelles. These include the cytoskeleton, which organizes the intracellular compartment; the nucleus, which holds the cell's genetic material and directs the daily activities of the cell; the ER and the Golgi apparatus, which produce, package, and transport proteins and lipids to the plasma membrane and lysosomes; the lysosomes and peroxisomes, which perform the task of intracellular digestion of organic waste; and the mitochondria, which produce cellular energy in the form of ATP. The energy released by ATP hydrolysis is used by the cell to drive the many energetically unfavorable reactions needed to maintain cellular functions. Multicellular organisms have developed complex communication systems to control cell behavior, such as growth and differentiation into specialized cell types. Disruption of these cellular processes is at the root of pathophysiologic processes and disease.

TABLE 3.2 Structure and Function of Major Cellular Components

Cellular Structure	Functions	Cellular Structure	Functions
Plasma membrane	Protective barrier separates life from nonlife	Nucleus	Protection of genetic material
	Extracellular message transduction		Regulation of cell type and function through control of protein synthesis
	Transport of materials into and out of cell	Endoplasmic reticulum	Protein and lipid synthesis
	Maintenance and transmission of membrane potentials		Lipid metabolism and detoxification
	Cell-to-cell recognition, interaction	Golgi apparatus	Protein and lipid modification and sorting
Cytoskeleton	Maintenance of cell shape		Transport of proteins and lipids to appropriate destinations
	Cell movement	Lysosomes	Hydrolytic breakdown of organic waste
	Trafficking within cell	Peroxisomes	Oxidative breakdown of organic waste
		Mitochondria	Cellular energy production (ATP)

ATP, Adenosine triphosphate.

RESOURCES

General Cell Structure and Function

Alberts B, et al: Cells and genomes. In Alberts B, et al, editors: *Molecular biology of the cell*, ed 6, New York, 2015, Garland Science, pp 1–42.

Alberts B, et al: Membrane structure. In Alberts B, et al, editors: *Molecular biology of the cell*, ed 6, New York, 2015, Garland Science, pp 565–596. See comment in PubMed Commons below.

Alberts B, et al: DNA, chromosomes and genomes. In Alberts B, et al, editors: *Molecular biology of the cell*, ed 6, New York, 2015, Garland Science, pp 175–236.

Alberts B, et al: Intracellular membrane traffic. In Alberts B, et al, editors: *Molecular biology of the cell*, ed 6, New York, 2015, Garland Science, pp 695–725.

Alberts B, et al: The cytoskeleton. In Alberts B, et al, editors: *Molecular biology of the cell*, ed 6, New York, 2015, Garland Science, pp 889–962.

Cymer F, von Heijne G, White SH: Mechanisms of integral membrane protein insertion and folding. *J Mol Biol* 427(5):999–1022, 2015.

Furukawa K, Ohkawa Y, Yamauchi Y, et al: Fine tuning of cell signals by glycosylation. *J Biochem* 151(6):573–578, 2012.

Johnson AE, et al: Structure, function, and regulation of free and membrane-bound ribosomes: the view from their substrates and products. *Cold Spring Harb Symp Quant Biol* 66:531–541, 2001.

Meisenberg G, Simmons WH: The cytoskeleton. In Meisenberg G, Simmons WH, editors: *Principles of medical biochemistry*, ed 3, Philadelphia, 2012, Saunders, pp 198–211.

Monnard P, Walde P: Current ideas about prebiological compartmentalization. *Life* 5(2):1239–1263, 2015.

Parenti G, Pignata C, Vajro P, Salerno M: New strategies for the treatment of lysosomal storage diseases (review). *Int J Mol Med* 31(1):11–20, 2013.

Singer SJ, Nicolson GL: The fluid mosaic model of the structure of cell membranes. *Science* 175:720–731, 1972.

Wandke C, Kulay U: Enclosing chromatin: reassembly of the nucleus after open mitosis. *Cell* 152(6):1222–1225, 2013.

Cellular Energy Metabolism

Alberts B, et al: Energy conversion: mitochondria and chloroplasts. In Alberts B, et al, editors: *Molecular biology of the cell*, ed 6, New York, 2015, Garland Science, pp 753–812.

Meisenberg G, Simmons WH: Glycolysis, tricarboxylic acid cycle, and oxidative phosphorylation. In Meisenberg G, Simmons WH, editors: *Principles of medical biochemistry*, ed 3, Philadelphia, 2012, Saunders, pp 347–373.

Pollard T, Earnshaw W: *Cell biology*, ed 2, Philadelphia, 2008, Saunders, p 332.

Cellular Electrical Potentials

Alberts B, et al: Membrane transport of small molecules and the electrical properties of membranes. In Alberts B, et al, editors: *Molecular biology of the cell*, ed 6, New York, 2015, Garland Science, pp 597–640.

Baker PF, Hodgkin AL, Shaw T: The effects of changes in internal ionic concentrations of the electrical properties of perfused giant axons. *J Physiol* 164:355–374, 1962.

Catterall WA: Structure and function of voltage-gated sodium channels at atomic resolution. *Exp Physiol* 99(1):35–51, 2014.

Hodgkin AL: *The conduction of the nervous impulse*, Liverpool, England, 1971, Liverpool University Press.

Hodgkin AL, Huxley AF: Currents carried by sodium and potassium ions through the membrane of the giant axon of Loligo. *J Physiol* 116:449–472, 1952.

Hodgkin AL, Huxley AF, Katz B: Measurement of current-voltage relations in the membrane of the giant axon of Loligo. *J Physiol* 116:424–448, 1952.

Hodgkin AL, Katz B: The effect of sodium ions on the electrical activity of the giant axon of the squid. *J Physiol* 108:37–77, 1949.

Lamas JA, Reboreda A, Codesido V: Ionic basis of the resting membrane potential in cultured rat sympathetic neurons. *Neuroreport* 13(5):585–591, 2002.

Pollard T, Earnshaw W: *Cell biology*, ed 2, Philadelphia, 2008, Saunders, p 149.

Cell Growth Regulation

Alberts B, et al: Cell signaling. In Alberts B, et al, editors: *Molecular biology of the cell*, ed 6, New York, 2015, Garland Science, pp 813–888.

Dick FA, Rubin SM: Molecular mechanisms underlying RB protein function. *Nat Rev Mol Cell Biol* 14(5):297–306, 2013.

Meisenberg G, Simmons WH: Cellular growth control and cancer. In Meisenberg G, Simmons WH, editors: *Principles of medical biochemistry*, ed 3, Philadelphia, 2012, Saunders, pp 307–332.

4

Cell Injury, Aging, and Death

Jacquelyn L. Banasik

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- What are the usual cellular responses to reversible injury?
- How are reversible and irreversible cellular injuries differentiated?
- How do necrosis and apoptosis differ?
- To what kind of injuries are cells susceptible?
- What are the usual physiologic changes of aging, and how are these differentiated from disease?

CHAPTER OUTLINE

Reversible Cell Injury, 59
Hydropic Swelling, 60
Intracellular Accumulations, 60
Cellular Adaptation, 63
Atrophy, 63
Hypertrophy, 63
Hyperplasia, 63
Metaplasia, 63
Dysplasia, 64
Irreversible Cell Injury, 64
Necrosis, 64
Apoptosis, 66
Etiology of Cellular Injury, 68
Ischemia and Hypoxic Injury, 68
Nutritional Injury, 70
Infectious and Immunologic Injury, 70
Chemical Injury, 72
Physical and Mechanical Injury, 72
Cellular Aging, 74
Cellular Basis of Aging, 74
Physiologic Changes of Aging, 75
Somatic Death, 75

Disease and injury are increasingly being understood as cellular and genomic phenomena. Although pathophysiologic processes are often presented in terms of systemic effects and manifestations, ultimately it is the *cells* that make up the systems that are affected. Even complex multisystem disorders such as cancer ultimately are the result of alterations in cell function. As the mysterious mechanisms of diseases are understood on the cellular and molecular levels, more specific methods of diagnosis, treatment, and prevention can be developed. This chapter presents the general characteristics of cellular injury, adaptation, aging, and death that underlie the discussions of systemic pathophysiologic processes presented in later chapters of this text.

Cells are confronted by many challenges to their integrity and survival and have efficient mechanisms for coping with an altered cellular environment. Cells respond to environmental changes or injury in three general ways: (1) when the change is mild or short lived, the cell may withstand the assault and completely return to normal. This is called a *reversible cell injury.* (2) The cell may adapt to a persistent but sublethal injury by changing its structure or function. Generally, adaptation also is reversible. (3) Cell death may occur if the injury is too severe or prolonged. Cell death is irreversible and may occur by two different processes termed *necrosis* and *apoptosis.* Necrosis is cell death caused by external injury, whereas apoptosis is triggered by intracellular signaling cascades that result in regulated cell death. Necrosis is considered to be a pathologic process associated with significant tissue damage, whereas apoptosis may be a normal physiologic process in some instances and pathologic in others.

REVERSIBLE CELL INJURY

Regardless of the cause, reversible injuries and the early stages of irreversible injuries often result in cellular swelling and the accumulation of excess substances within the cell. These changes reflect the cell's inability to perform normal metabolic functions because of insufficient cellular energy in the form of adenosine triphosphate (ATP) or dysfunction of

associated metabolic enzymes. Once the acute stress or injury has been removed, by definition of a reversible injury, the cell returns to its preinjury state.

Hydropic Swelling

Cellular swelling attributable to accumulation of water—called *hydropic swelling* or *oncosis*—is the first manifestation of most forms of reversible cell injury. Hydropic swelling results from malfunction of the sodium–potassium (Na^+–K^+) pumps that normally maintain ionic equilibrium of the cell. Failure of the Na^+–K^+ pump results in accumulation of sodium ions within the cell, creating an osmotic gradient for water entry. Because Na^+–K^+ pump function is dependent on the presence of cellular ATP, any injury that causes insufficient energy production will result in hydropic swelling (Fig. 4.1). Hydropic swelling is characterized by a large, pale cytoplasm; dilated endoplasmic reticulum; and swollen mitochondria. With severe hydropic swelling, the endoplasmic reticulum may rupture and form large water-filled vacuoles. Generalized swelling in the cells of a particular organ may cause the organ to increase in size and weight. Organ enlargement is indicated by the suffix *-megaly* (e.g., *splenomegaly* denotes an enlarged spleen, *hepatomegaly* denotes an enlarged liver).

Intracellular Accumulations

Excess accumulations of substances in cells may result in cellular injury because the substances are toxic or provoke an immune response, or merely because they occupy space needed for cellular functions. In some cases, accumulations do not in themselves appear to be injurious, but rather are indicators of cell injury. Intracellular accumulations may be categorized as (1) excessive amounts of normal intracellular substances such as fat, (2) accumulation of abnormal substances produced by the cell because of faulty metabolism or synthesis, and (3) accumulation of pigments and particles that the cell is unable to degrade (Fig. 4.2).

Normal intracellular substances that tend to accumulate in injured cells include lipids, carbohydrates, glycogen, and proteins. Faulty metabolism of these substances within the cell results in excessive intracellular storage. In some cases, the enzymes required for breaking down a particular substance are absent or abnormal as a result of a genetic defect. In other cases, altered metabolism may be due to excessive intake, toxins, or other disease processes.

A common site of intracellular lipid accumulation is the liver, where many fats are normally stored, metabolized, and synthesized. Fatty liver is often associated with excessive intake of alcohol. Mechanisms whereby alcohol causes fatty liver remain unclear, but it is thought to result from direct toxic effects, as well as the preferential metabolism of alcohol instead of lipid (see Chapter 38 for a discussion of fatty liver). Lipids may also contribute to atherosclerotic diseases and accumulate in blood vessels, kidney, heart, and other organs. Fat-filled cells tend to compress cellular components to one side and cause the tissue to appear yellowish and greasy (Fig. 4.3). In several genetic disorders, the enzymes needed to metabolize lipids are impaired; these include Tay-Sachs disease and Gaucher disease, in which lipids accumulate in neurologic tissue.

Glycosaminoglycans (mucopolysaccharides) are large carbohydrate complexes that normally compose the extracellular matrix of connective tissues. Connective tissue cells secrete most of the glycosaminoglycan into the extracellular space, but a small portion remains inside the cell and is normally degraded by lysosomal enzymes. The *mucopolysaccharidoses* are a group of genetic diseases in which the enzymatic degradation of these molecules is impaired and they collect within the cell. Mental disabilities and connective tissue disorders are common findings.

Like other disorders of accumulation, *excessive glycogen storage* can be the result of inborn errors of metabolism, but a common cause is diabetes mellitus. Diabetes mellitus is associated with impaired cellular uptake of glucose, which results in high serum and urine glucose levels. Cells of the renal tubules reabsorb the excess filtered glucose and store it intracellularly as glycogen. The renal tubule cells also are a common site for abnormal accumulations of proteins. Normally, very little protein escapes the bloodstream into the urine. However, with certain disorders, renal glomerular capillaries become leaky and allow proteins to pass through them. Renal tubule cells recapture some of the escaped proteins through endocytosis, resulting in abnormal accumulation.

Cellular stress may lead to accumulation and aggregation of denatured proteins. The abnormally folded intracellular proteins may cause serious cell dysfunction and death if they are allowed to persist in the cell. A family of stress proteins (also called *chaperone or heat-shock proteins*) is responsible for binding and refolding aberrant proteins back into their correct three-dimensional forms (Fig. 4.4). If the chaperones are unsuccessful in correcting the defect, the abnormal proteins form complexes with another protein called *ubiquitin*. Ubiquitin targets the abnormal proteins to enter a proteosome complex, where they are digested into fragments that are less injurious to cells (see Fig. 4.4).

Finally, a variety of pigments and inorganic particles may be present in cells. Some pigment accumulations are normal, such as the accumulation

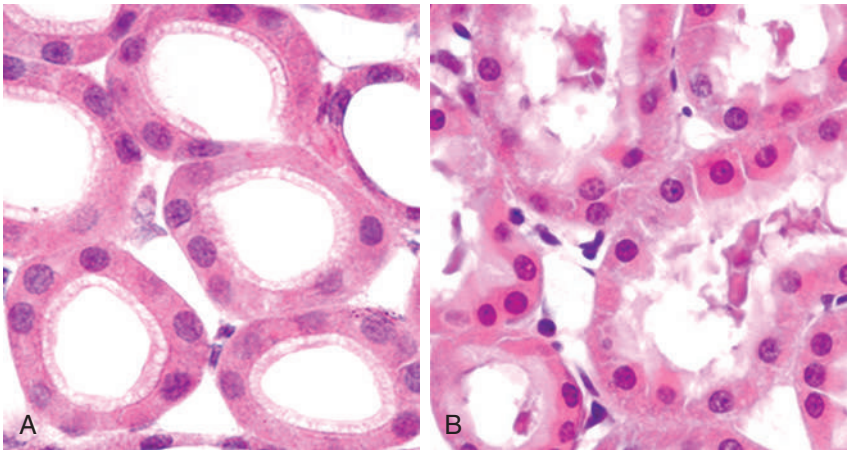


FIG 4.1 Cellular swelling in kidney tubule epithelial cells. **A,** Normal kidney tubule with cuboidal cells. **B,** Early ischemic changes showing surface blebs and swelling of cells. (From Kumar V et al: *Robbins and Cotran pathologic basis of disease*, ed 9, Philadelphia, 2015, Saunders, p 42. Photograph courtesy Drs. Neal Pinckard and M. A. Venkatachalam, University of Texas Health Sciences Center, San Antonio, TX.)

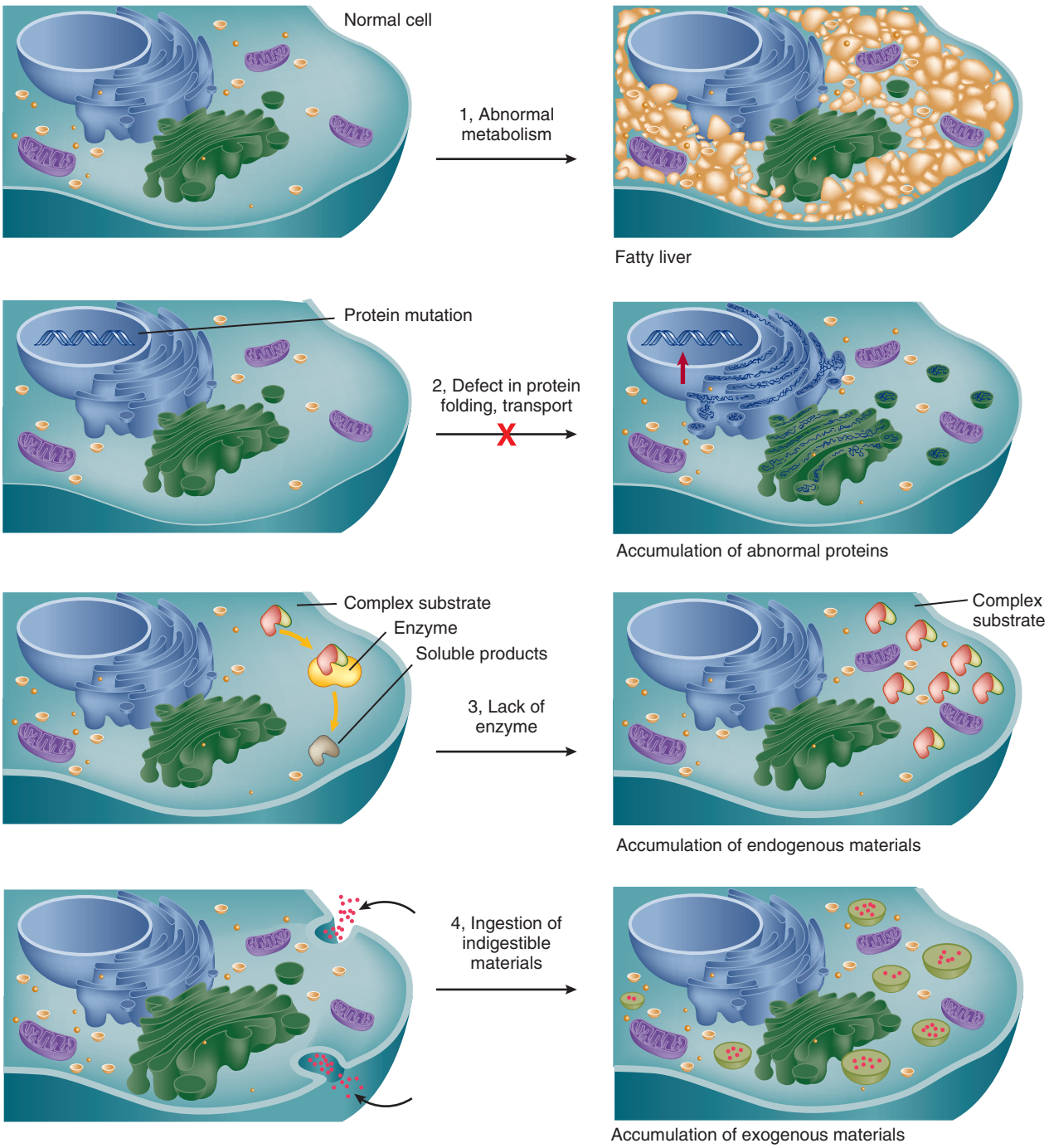


FIG 4.2 General mechanisms of intracellular accumulation: (1) abnormal metabolism as in fatty change in the liver, (2) mutations causing alterations in protein folding and transport so that defective proteins accumulate, (3) deficiency of critical enzyme responsible for lysosomal degradation, and (4) an inability to degrade phagocytosed particles such as coal dust.

of melanin in tanned skin, whereas others signify pathophysiologic processes. Pigments may be produced by the body (endogenous) or may be introduced from outside sources (exogenous). In addition to melanin, the iron-containing substances hemosiderin and bilirubin are endogenous pigments that, when present in excessive amounts, indicate disease processes. Hemosiderin and bilirubin are derived from hemoglobin. Excessive amounts may indicate abnormal breakdown of red blood cells (RBCs), prolonged administration of iron, or the presence of hepatobiliary disorders. Inorganic particles that may accumulate include calcium, tar, and mineral dusts such as coal, silica, iron, lead, and silver. Mineral dusts generally are inhaled and accumulate in lung tissue (Fig. 4.5). Inhaled dusts cause chronic inflammatory reactions in the lung, which generally result in destruction of pulmonary alveoli and capillaries and the formation of scar tissue. Over many years, the lung may become stiff and difficult to expand because of extensive scarring (see Chapter 23).

Deposits of calcium salts occur in conditions of altered calcium intake, excretion, or metabolism. Impaired renal excretion of phosphate

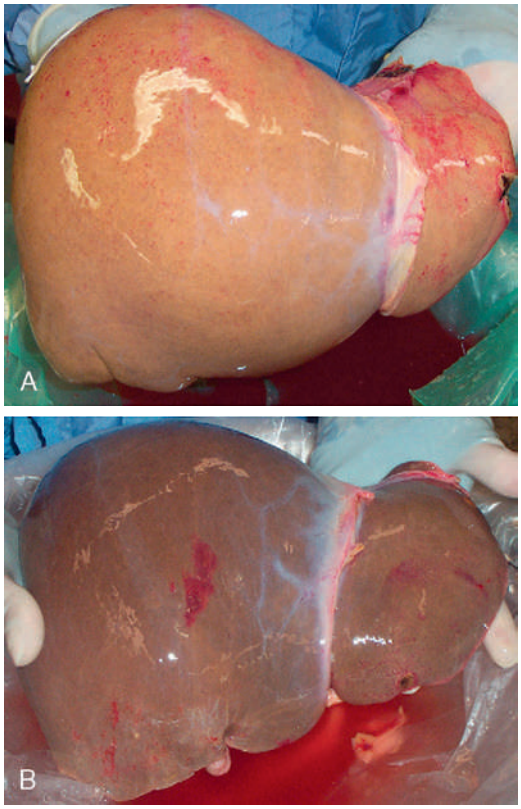


FIG 4.3 Gross appearance of a fatty donor liver **(A)** compared with a normal nonfatty donor liver **(B).** (From Odze RD, Goldblum JR. *Surgical pathology of the GI tract, liver, biliary tract, and pancreas*, ed 2, St Louis, 2009, Elsevier.)

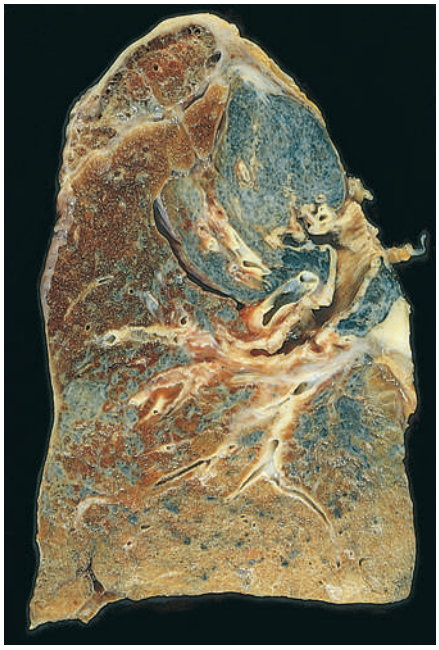


FIG 4.5 Accumulations of silicon dust in tissues of the lung. (From Kumar V et al: *Robbins and Cotran pathologic basis of disease,* ed 9, Philadelphia, 2015, Saunders, p 690. Photograph courtesy Dr. John Goldeski, Brigham and Women's Hospital, Boston, MA.)

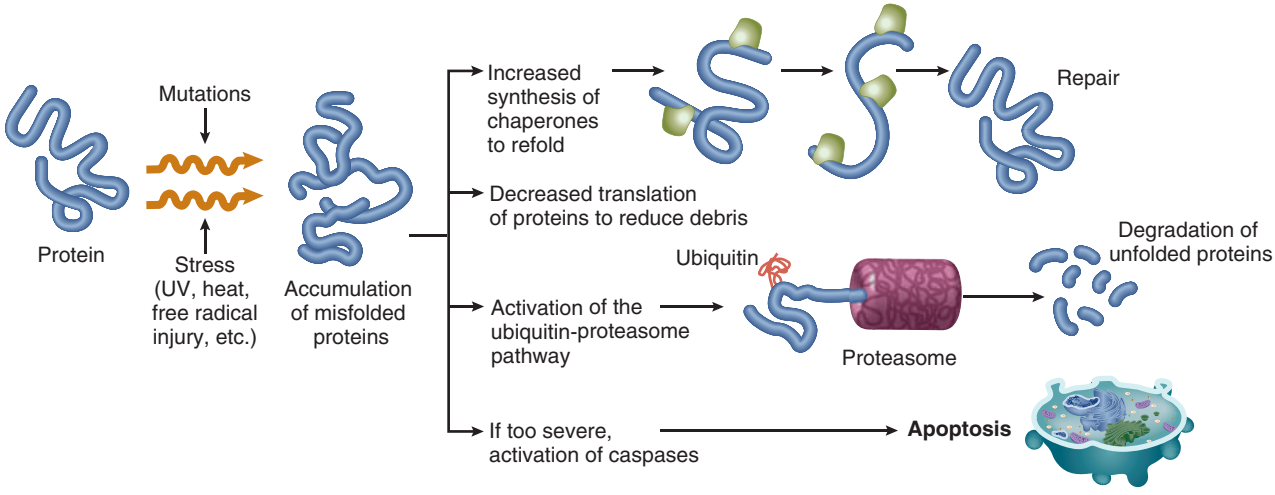


FIG 4.4 Roles of chaperone proteins in protein refolding and ubiquitin in protein degradation after stress-induced protein damage.

may result in the formation of calcium phosphate salts that are deposited in the tissues of the eye, heart, and blood vessels. Calcification of the heart valves may cause obstruction to blood flow through the heart or interfere with valve closing. Calcification of blood vessels may result in narrowing of vessels and insufficient blood flow to distal tissues. Dead and dying tissues often become calcified (filled with calcium salts) and appear as dense areas on x-ray films. For example, lung damage resulting from tuberculosis often is apparent as calcified areas, called *tubercles.*

With the exception of inorganic particles, the intracellular accumulations generally are reversible if the causative factors are removed.

KEY POINTS

- Hydropic swelling is an early indicator of cell injury. It results from Na^+–K^+ pump dysfunction at the cell membrane.
- Intracellular accumulations of abnormal endogenous or exogenous particles indicate a disorder of cellular metabolism.
- Damage from accumulation of abnormal intracellular protein is limited by chaperone proteins that attempt to refold the protein into its correct shape and by the ubiquitin–proteosome system that digests targeted proteins into fragments.

CELLULAR ADAPTATION

The cellular response to persistent, sublethal stress reflects the cell's efforts to adapt. Cellular stress may be due to an increased functional demand or a reversible cellular injury. Although the term *adaptation* implies a change for the better, in some instances an adaptive change may not be beneficial. The common adaptive responses are atrophy (decreased cell size), hypertrophy (increased cell size), hyperplasia (increased cell number), metaplasia (conversion of one cell type to another), and dysplasia (disorderly growth) (Fig. 4.6). Each of these changes is potentially reversible when the cellular stress is relieved.

Atrophy

Atrophy occurs when cells shrink and reduce their differentiated functions in response to a variety of normal and injurious factors. The general causes of atrophy may be summarized as (1) disuse, (2) denervation, (3) ischemia, (4) nutrient starvation, (5) interruption of endocrine signals, (6) and persistent cell injury. Apparently, atrophy represents an effort by the cell to minimize its energy and nutrient consumption by decreasing the number of intracellular organelles and other structures.

A common form of atrophy is the result of a reduction in functional demand, sometimes called disuse atrophy**.** For example, immobilization by bed rest or casting of an extremity results in shrinkage of skeletal muscle cells. On resumption of activity, the tissue resumes its normal size. Denervation of skeletal muscle results in a similar decrease in muscle size caused by loss of nervous stimulation. Inadequate blood supply to a tissue is known as ischemia**.** If the blood supply is totally interrupted, the cells will die, but chronic sublethal ischemia usually results in cell atrophy. The heart, brain, kidneys, and lower leg are common sites of ischemia. Atrophic changes in the lower leg attributable to ischemia include thin skin, muscle wasting, and hair loss. Atrophy can be a consequence of chronic nutrient starvation, whether the result of poor intake, absorption, or distribution to the tissues. Many glandular tissues throughout the body depend on growth-stimulating (trophic) signals to maintain size and function. For example, the adrenal cortex, thyroid, and gonads are maintained by trophic hormones from the pituitary gland and will atrophy in their absence. Atrophy that results from persistent cell injury is most commonly related to chronic inflammation and infection.

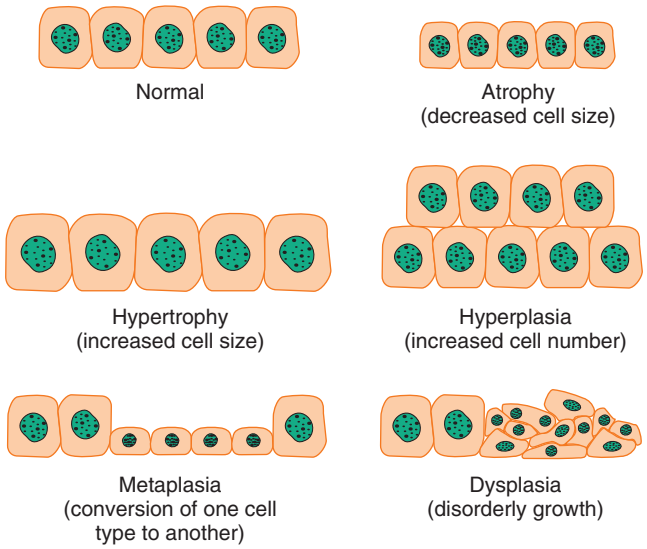


FIG 4.6 The adaptive cellular responses of atrophy, hypertrophy, hyperplasia, metaplasia, and dysplasia.

The biochemical pathways that result in cellular atrophy are imperfectly known; however, two pathways for protein degradation have been implicated. The first is the previously mentioned ubiquitin–proteosome system, which degrades targeted proteins into small fragments (see Fig. 4.4). The second involves the lysosomes that may fuse with intracellular structures, leading to hydrolytic degradation of the components (autophagy). Certain substances apparently are resistant to degradation and remain in the lysosomal vesicles of atrophied cells. For example, lipofuscin is an age-related pigment that accumulates in residual vesicles in atrophied cells, giving them a yellow-brown appearance.

Hypertrophy

Hypertrophy is an increase in cell mass accompanied by an augmented functional capacity. Cells hypertrophy in response to increased physiologic or pathophysiologic demands. Cellular enlargement results primarily from a net increase in cellular protein content. Like the other adaptive responses, hypertrophy subsides when the increased demand is removed; however, the cell may not entirely return to normal because of persistent changes in connective tissue structures. Organ enlargement may be a result of both an increase in cell size (hypertrophy) and an increase in cell number (hyperplasia). For example, an increase in skeletal muscle mass and strength in response to repeated exercise is primarily the result of hypertrophy of individual muscle cells, although some increase in cell number is also possible because muscle stem cells (satellite cells) are able to divide. Physiologic hypertrophy occurs in response to a variety of trophic hormones in sex organs—the breast and uterus, for example. Certain pathophysiologic conditions may place undue stress on some tissues, causing them to hypertrophy. Liver enlargement in response to bodily toxins and cardiac muscle enlargement in response to high blood pressure (Fig. 4.7) represent hyperplastic and hypertrophic adaptations to pathologic conditions. Hypertrophic adaptation is particularly important for cells, such as differentiated muscle cells, that are unable to undergo mitotic division.

Hyperplasia

Cells that are capable of mitotic division generally increase their functional capacity by increasing the number of cells (hyperplasia) as well as by hypertrophy. Hyperplasia usually results from increased physiologic demands or hormonal stimulation. Persistent cell injury also may lead to hyperplasia. Examples of demand-induced hyperplasia include an increase in RBC number in response to high altitude and liver enlargement in response to drug detoxification. Trophic hormones induce hyperplasia in their target tissues. Estrogen, for example, leads to an increase in the number of endometrial and uterine stromal cells. Dysregulation of hormones or growth factors can result in pathologic hyperplasia, such as that which occurs in thyroid or prostate enlargement.

Chronic irritation of epithelial cells often results in hyperplasia. Calluses and corns, for example, result from chronic frictional injury to the skin. The epithelium of the bladder commonly becomes hyperplastic in response to the chronic inflammation of cystitis.

Metaplasia

Metaplasia is the replacement of one differentiated cell type with another. This most often occurs as an adaptation to persistent injury, with the replacement cell type better able to tolerate the injurious stimulation. Metaplasia is fully reversible when the injurious stimulus is removed. Metaplasia often involves the replacement of glandular epithelium with squamous epithelium. Chronic irritation of the bronchial mucosa by cigarette smoke, for example, leads to the conversion of ciliated columnar epithelium to stratified squamous epithelium. Metaplastic cells generally remain well differentiated and of the same tissue type, although cancerous

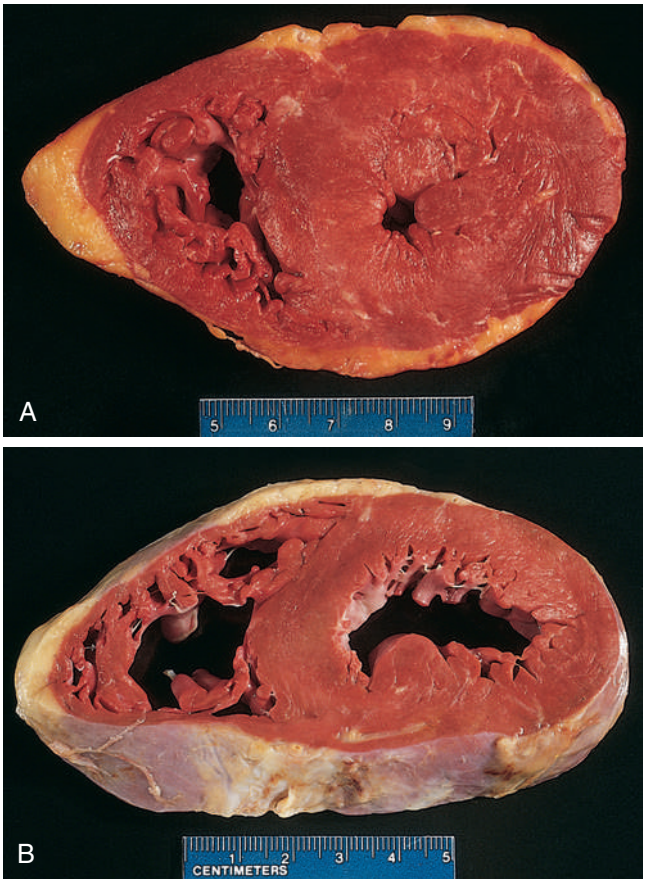


FIG 4.7 **A,** Hypertrophy of cardiac muscle in the left ventricular chamber. **B,** Compare with the thickness of the normal left ventricle. This is an example of cellular adaptation to an increased cardiac workload. (From Kumar V et al: *Robbins and Cotran pathologic basis of disease*, ed 9, Philadelphia, 2015, Saunders, p 33.)

transformations can occur. Some cancers of the lung, cervix, stomach, and bladder appear to derive from areas of metaplastic epithelium.

Dysplasia

Dysplasia refers to the disorganized appearance of cells because of abnormal variations in size, shape, and arrangement. Dysplasia occurs most frequently in hyperplastic squamous epithelium, but it may also be seen in the mucosa of the intestine. Dysplasia probably represents an adaptive effort gone astray. Dysplastic cells have significant potential to transform into cancerous cells and are usually regarded as *pre*neoplastic lesions. (See Chapter 7 for a discussion of cancer.) Dysplasia that is severe and involves the entire thickness of the epithelium is called *carcinoma in situ.* Mild forms of dysplasia may be reversible if the inciting cause is removed.

KEY POINTS

- Adaptive cellular responses indicate cellular stress caused by altered functional demand or chronic sublethal injury.
- Hypertrophy and hyperplasia generally result from increased functional demand. Atrophy results from decreased functional demand or chronic ischemia. Metaplasia and dysplasia result from persistent injury.

IRREVERSIBLE CELL INJURY

Pathologic cellular death occurs when an injury is too severe or prolonged to allow cellular adaptation or repair. Two different processes may contribute to cell death in response to injury: necrosis and apoptosis. Necrosis usually occurs as a consequence of ischemia or toxic injury and is characterized by cell rupture, spilling of contents into the extracellular fluid, and inflammation. Apoptosis (from a Greek word meaning *falling off,* as in leaves from a tree) occurs in response to injury that does not directly kill the cell but triggers intracellular cascades that activate a cellular suicide response. Apoptotic cells generally do not rupture and are ingested by neighboring cells with minimal disruption of the tissue and without inflammation. Apoptosis is not always a pathologic process and occurs as a necessity of development and tissue remodeling.

Necrosis

Necrotic cells demonstrate typical morphologic changes, including a shrunken (pyknotic) nucleus that is subsequently degraded (karyolysis), a swollen cell volume, dispersed ribosomes, and disrupted plasma and organelle membranes (Fig. 4.8). The disruption of the permeability barrier of the plasma membrane appears to be a critical event in the death of the cell.

Localized injury or death of tissue is generally reflected in the entire system as the body attempts to remove dead cells and works to compensate for loss of tissue function. Several manifestations indicate that the system is responding to cellular injury and death. A general inflammatory response is often present, with general malaise, fever, increased heart rate, increased white blood cell (WBC) count, and loss of appetite. With the death of necrotic cells, intracellular contents are released and often find their way into the bloodstream. The presence of specific cellular proteins in the blood is used as an indicator of the location and extent of cellular death. For example, an elevated serum amylase level indicates pancreatic damage, and an elevated creatine kinase (MB isoenzyme) or cardiac troponin level indicates myocardial damage. The location of pain caused by tissue destruction may also aid in the diagnosis of cellular death.

Four different types of tissue necrosis have been described: coagulative, liquefactive, fat, and caseous (Fig. 4.9). They differ primarily in the type of tissue affected. *Coagulative* necrosis is the most common. Manifestations of coagulative necrosis are the same, regardless of the cause of cell death. In general, the steps leading to coagulative necrosis may be summarized as follows: (1) ischemic cellular injury, leading to (2) loss of the plasma membrane's ability to maintain electrochemical gradients, which results in (3) an influx of calcium ions and mitochondrial dysfunction, and (4) degradation of plasma membranes and nuclear structures (Fig. 4.10). The area of coagulative necrosis is composed of denatured proteins and is relatively solid. The coagulated area is then slowly dissolved by proteolytic enzymes, and the general tissue architecture is preserved for a relatively long time (weeks). This is in contrast to liquefactive necrosis.

When the dissolution of dead cells occurs very quickly, a liquefied area of lysosomal enzymes and dissolved tissue may result and form an abscess or cyst. This type of necrosis, called *liquefactive necrosis,* may be seen in the brain, which is rich in degradative enzymes and contains little supportive connective tissue. Liquefaction may also result from a bacterial infection that triggers a localized collection of WBCs. The phagocytic WBCs contain potent degradative enzymes that may completely digest dead cells, resulting in liquid debris.

Fat necrosis refers to death of adipose tissue and usually results from trauma or pancreatitis. The process begins with the release of activated digestive enzymes from the pancreas or injured tissue. The

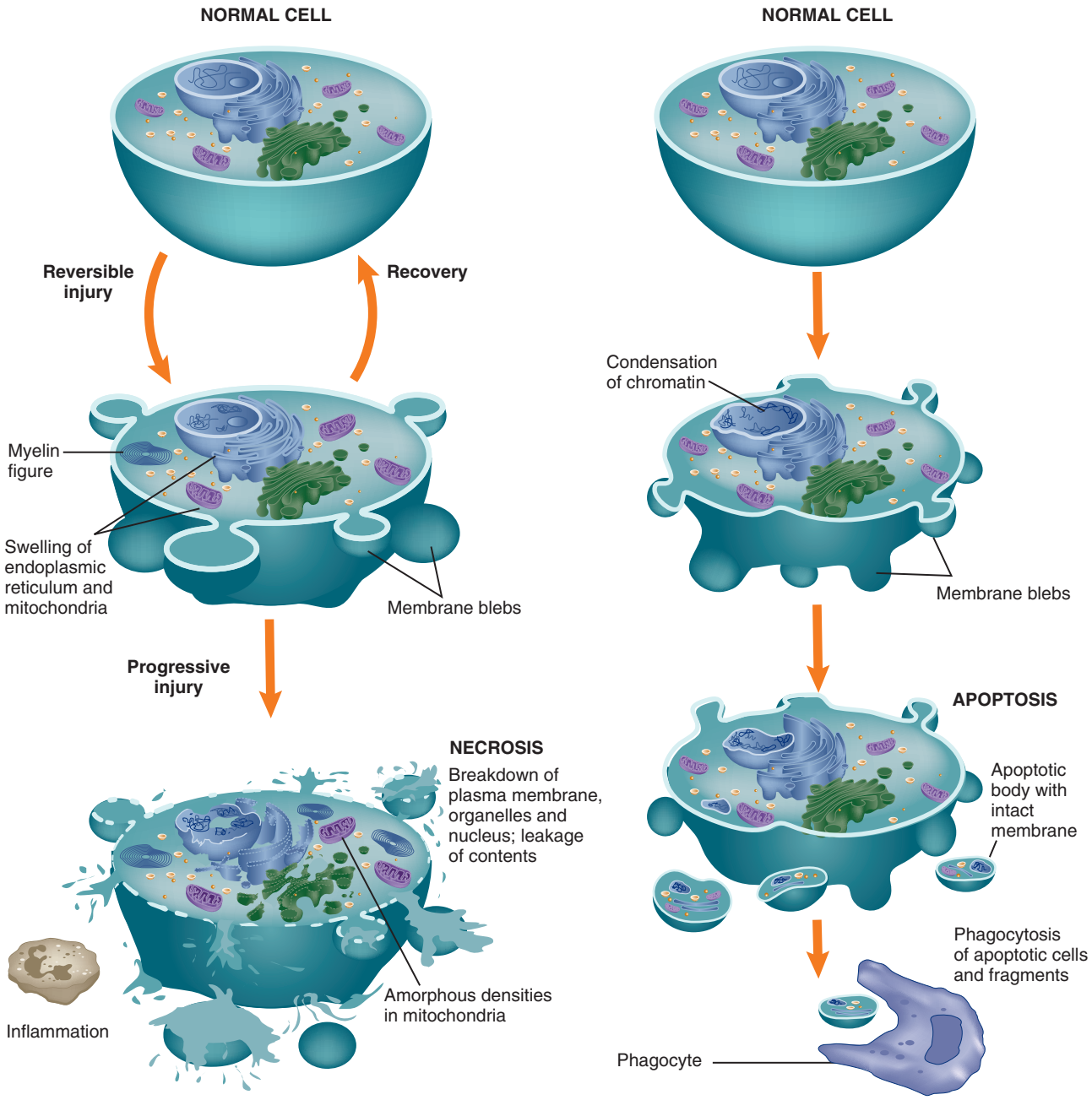


FIG 4.8 Comparison of cellular changes in necrosis and apoptosis.

enzymes attack the cell membranes of fat cells, causing release of their stores of triglycerides. Pancreatic lipase can then hydrolyze the triglycerides to free fatty acids and glycerol, which precipitate as calcium soaps (saponification). Fat necrosis appears as a chalky white area of tissue.

Caseous necrosis is characteristic of lung tissue damaged by tuberculosis. The areas of dead lung tissue are white, soft, and fragile, resembling clumpy cheese. Dead cells are walled off from the rest of the lung tissue by inflammatory WBCs. In the center, the dead cells lose their cellular structure but are not totally degraded. Necrotic debris may persist indefinitely.

Gangrene is a term used to describe cellular death involving a large area of tissue. Gangrene usually results from interruption of the major blood supply to a particular body part, such as the toes, leg, or bowel. Depending on the appearance and subsequent infection of the necrotic tissue, it is described as dry gangrene, wet gangrene, or gas gangrene. *Dry* gangrene is a form of coagulative necrosis characterized by blackened, dry, wrinkled tissue that is separated from adjacent healthy tissue by an obvious line of demarcation (see Fig. 4.9A). It generally occurs only on the extremities. Liquefactive necrosis may result in *wet* gangrene, which is typically found in internal organs, appears cold and black, and may be foul smelling because of the invasion of bacteria. Rapid spread of tissue damage and the release of toxins into the bloodstream make wet gangrene a life-threatening problem. *Gas* gangrene is characterized by the formation of bubbles of gas in damaged tissue. Gas gangrene is the result of infection of necrotic tissue by anaerobic bacteria of the genus *Clostridium.* These bacteria produce toxins and degradative enzymes that allow the infection to spread rapidly through the necrotic tissue. Gas gangrene may be fatal if not managed rapidly and aggressively.

FIG 4.9 The four primary types of tissue necrosis. **A,** Coagulative. **B,** Liquefactive. **C,** Fat. **D,** Caseous. (A, From Crowley L: *Introduction to human disease*, ed 4, Sudbury, MA, 1996, Jones and Bartlett, www.jbpub.com. Reprinted with permission. B–D, From Kumar V et al: *Robbins and Cotran pathologic basis of disease*, ed 9, Philadelphia, 2015, Saunders, pp 43–44.)

Apoptosis

The number of cells in tissues is tightly regulated by controlling the rate of cell division and the rate of cell death. If cells are no longer needed, they activate a cellular death pathway resulting in regulated cell death. In contrast to necrosis, which is messy and results in inflammation and collateral tissue damage, apoptosis is tidy and does not elicit inflammation. Apoptosis is not a rare event; large numbers of cells are continually undergoing programmed cell death as tissues remodel. During fetal development, for example, more than half of the nerve cells that form undergo apoptosis. It is estimated that more than 95% of the T lymphocytes that are generated in the bone marrow are induced to undergo apoptosis after reaching the thymus. These are normal physiologic processes that regulate normal system function.

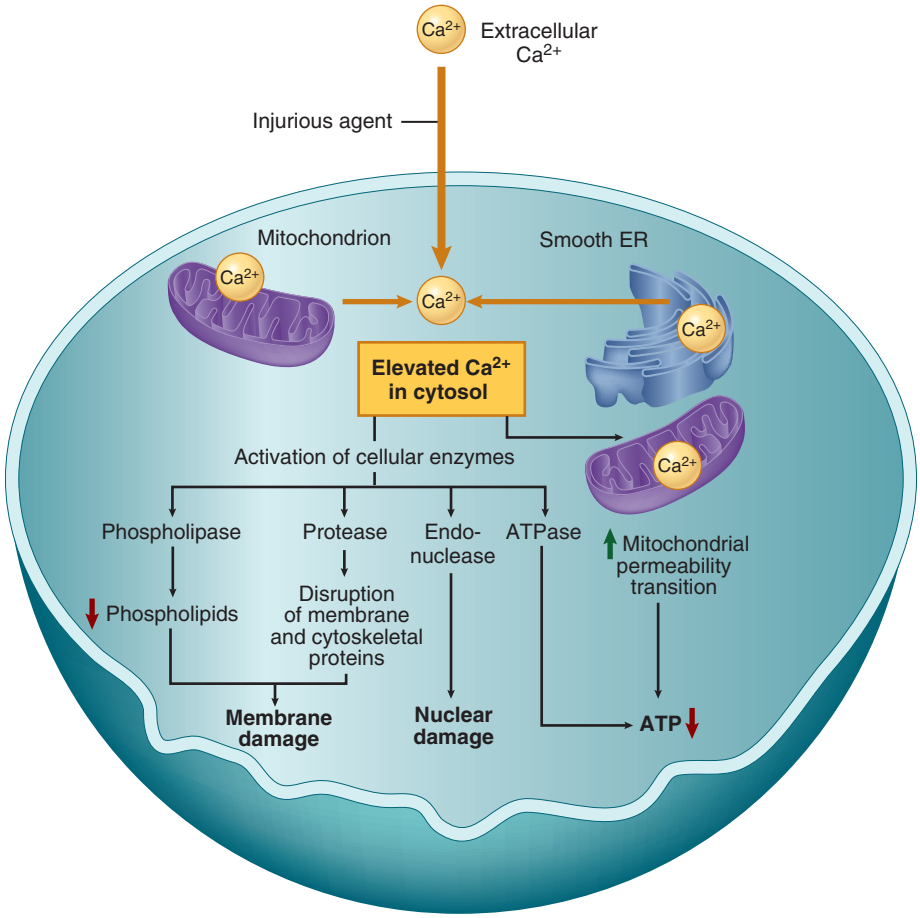


FIG 4.10 Cellular injury as a consequence of intracellular calcium overload.

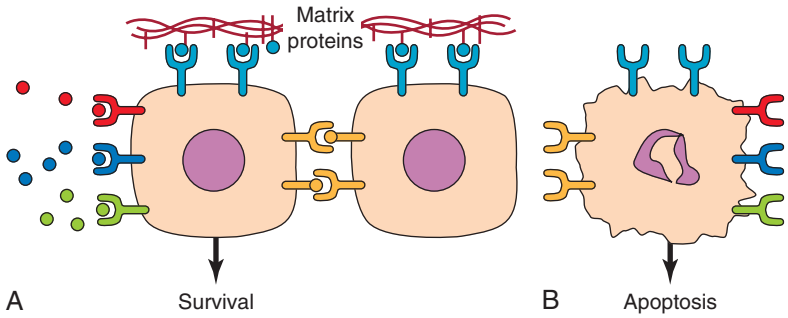


FIG 4.11 Each cell displays a set of receptors that enable it to respond to extracellular signals that control growth, differentiation, and survival. **A,** Extracellular signals are provided by the neighboring cells, secreted signaling molecules, and the extracellular matrix. **B,** Withdrawal of these survival signals induces the cell to initiate apoptosis.

Apoptosis also has been implicated in pathologic cell death and disease. For example, it has been estimated that the area of tissue death after a myocardial infarction (heart attack) is about 20% necrotic and 80% apoptotic. It is difficult to measure the degree of apoptotic cell death because neighboring cells rapidly ingest their apoptotic neighbors, and few are ever present in the tissue but the rates are high. Death of cancer cells in response to radiation or chemotherapy is believed to be primarily caused by apoptotic mechanisms. When the rate of apoptosis is greater than the rate of cell replacement, tissue or organ function may be impaired. Apoptosis is now recognized as a primary factor in diseases such as heart failure (Chapter 19) and dementia (Chapter 45). The mechanisms regulating apoptosis are complex, and only major concepts are included here.

Environmental or extrinsic signals may induce apoptosis. Apoptosis may be triggered by withdrawal of "survival" signals that normally suppress the apoptotic pathways. Normal cells require a variety of signals from neighboring cells and from the extracellular matrix in order to stay alive (Fig. 4.11). If these contacts or signals are removed, the cell

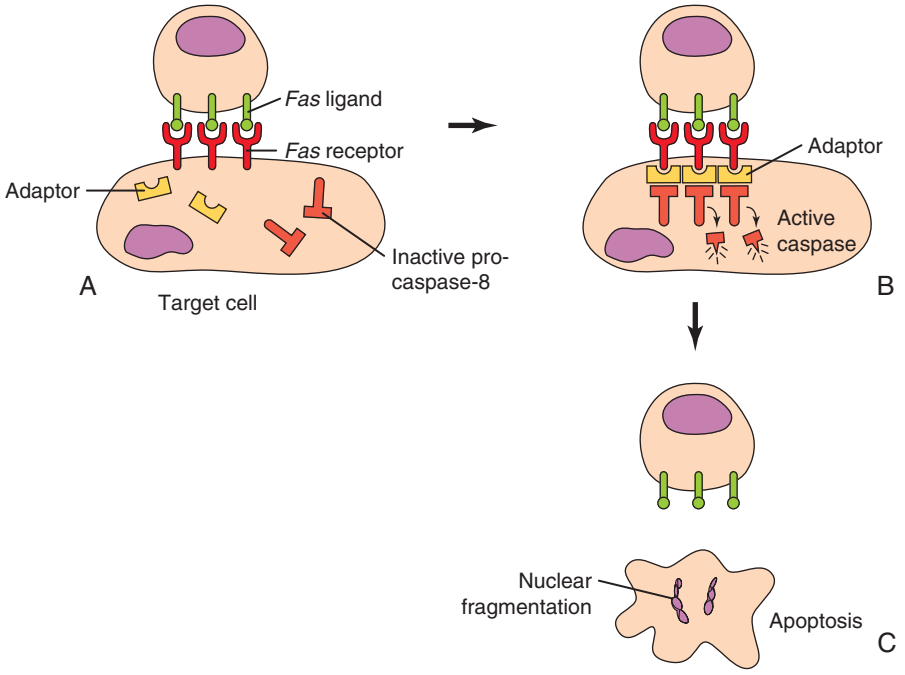


FIG 4.12 Induction of apoptosis by Fas ligand. **A,** Target cell binds to Fas ligand on a signaling cell. **B,** Active Fas receptors organize and activate caspases. **C,** The caspases degrade the nucleus and trigger cell death.

death cascade is activated. A second mechanism of triggering apoptosis involves extracellular signals, such as the *Fas* ligand, that bind to the cell and trigger the death cascade through activation of "death receptors" (Fig. 4.12).

Apoptosis can also be triggered by internal pathways. Cells have ways to monitor their condition and usefulness. When excessive, irreparable damage occurs to the cell's DNA or other vital structures, growth and division stall for a while to permit repair. If the damage is too great, the cell can trigger its own death. Mitochondrial damage with leakage of cytochrome *c* into the cytoplasm is a critical activator of the intrinsic apoptotic pathway. This pathway is governed in part by a protein called *p53.* The amount of p53 in a cell is normally quite low but increases in response to cellular DNA damage. If high levels of p53 are sustained, apoptosis will occur. Thus p53 is important in preventing the proliferation of cells with damaged DNA. A large number of cancers (50%) are associated with a mutation in the *P53* gene, which allows cancer cells to escape this monitoring system.

Regardless of the initiating event, apoptosis involves numerous intracellular signals and enzymes (Fig. 4.13). A family of enzymes called *caspases* is the main component of the proteolytic cascade that degrades key intracellular structures leading to cell death. The caspases are proenzymes that are activated in a cascade. Activation of a few *initiator* caspases at the beginning of the cascade results in a rapid domino effect of caspase activation. Some caspases cleave key proteins, such as the nuclear lamina, to destroy the nuclear envelope, whereas others activate still more enzymes that chop up the DNA. All of this destruction requires cellular energy and is contained within an intact plasma membrane, and the cell remnants are then assimilated by their neighbors. Neighboring cells are prompted to ingest apoptotic cells, because a phospholipid that is normally located only on the cytoplasmic side of a healthy cell (phosphatidylserine) flips to the outside of the lipid bilayer. This membrane lipid signals neighbors and tissue macrophages to bind and assimilate the cell components and suppresses the inflammatory response that normally accompanies phagocytosis.

KEY POINTS

- Necrosis occurs when the injury is too severe or prolonged to allow adaptation and is usually a consequence of disrupted blood supply.
- Local and systemic indicators of necrotic cell death include pain, elevated serum enzyme levels, inflammation (fever, elevated white blood cell [WBC] count, malaise), and loss of function.
- Different tissues exhibit necrosis of different types: heart (coagulative), brain (liquefactive), lung (caseous), and pancreas (fat).
- Gangrene refers to a large area of necrosis that may be described as dry, wet, or gas gangrene. Gas gangrene and wet gangrene may be rapidly fatal.
- Apoptosis is cell death resulting from activation of intracellular signaling cascades. Apoptosis is tidy and not usually associated with systemic manifestations of inflammation. Apoptosis requires adequate amounts of adenosine triphosphate (ATP) be present in the cell.

ETIOLOGY OF CELLULAR INJURY

Cellular injury and death result from a variety of cellular assaults, including lack of oxygen and nutrients, infection and immune responses, chemicals, and physical and mechanical factors. The extent of cell injury and death depends in part on the duration and severity of the assault and in part on the prior condition of the cells. Well-nourished and somewhat adapted cells may withstand the injury better than cells that are poorly nourished or not adapted.

Ischemia and Hypoxic Injury

Living cells must receive a continuous supply of oxygen to produce sufficient ATP to power energy-requiring functions. The lack of oxygen (hypoxia) results in power failure within the cell. Tissue hypoxia is most often caused by *ischemia,* or the interruption of blood flow to an area, but it may also result from heart failure, lung disease, and RBC disorders. Ischemia is the most common cause of cell injury in clinical medicine and injures cells faster than hypoxia alone. Faster injury occurs because

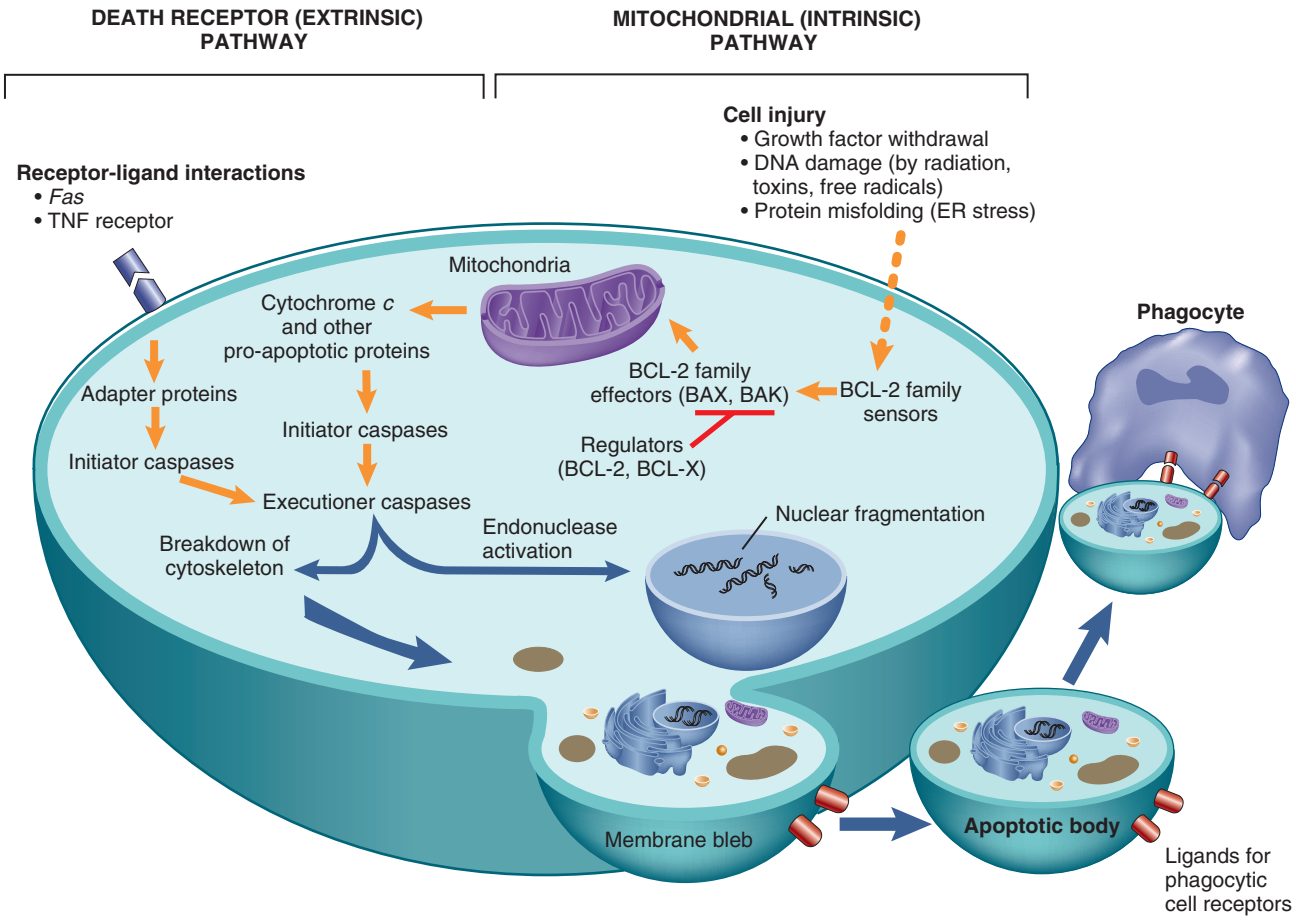


FIG 4.13 Schematic of the events of apoptosis. Numerous triggers can initiate apoptosis through intrinsic cell injury pathways (mitochondrial), such as withdrawal of survival factors, various cell injuries, and protein overload or misfolding; or through extrinsic cell injury pathways (death receptors), such as binding to Fas or tumor necrosis factor receptors. A number of intracellular regulatory proteins may inhibit or promote the activation of caspases, which, when activated, begin the process of cellular degradation and apoptotic cell fragmentation. Fragments are internalized by phagocytic cells.

ischemia not only disrupts the oxygen supply but also allows metabolic wastes to accumulate and deprives the cell of nutrients for glycolysis. The cellular events that follow oxygen deprivation are shown in Fig. 4.14. Decreased oxygen delivery to the mitochondria causes ATP production in the cell to stall and ATP-dependent pumps, including the Na^+–K^+ and Ca^{2+} pumps, to fail. Sodium accumulation within the cell creates an osmotic gradient favoring water entry, resulting in hydropic swelling. Excess intracellular calcium collects in the mitochondria, further interfering with mitochondrial function. A small amount of ATP is produced by anaerobic glycolytic pathways, which metabolize cellular stores of glycogen. The pyruvate end products of glycolysis accumulate and are converted to lactate, causing cellular acidification. Lactate can escape into the bloodstream, resulting in lactic acidosis, which can be detected by laboratory tests. Cellular proteins and enzymes become progressively more dysfunctional as the pH falls. Up to a point, ischemic injury is reversible, but when the plasma, mitochondrial, and lysosomal membranes are critically damaged, cell death ensues.

Cell death resulting from ischemia may be slow to develop, generally taking many minutes to hours. In fact, most cellular damage occurs after the blood supply to the tissues has been restored—a so-called *reperfusion injury.* Ischemia–reperfusion is a complex phenomenon, but three critical components have been identified: (1) calcium overload, (2) formation of reactive oxygen molecules (free radicals), and (3) subsequent inflammation.

Restoration of blood flow to ischemic cells bathes them in a fluid high in calcium ions at a time when their ATP stores are depleted and they are unable to control ion flux across the cell membrane. Accumulation of calcium ions in the cytoplasm can trigger apoptosis or activate enzymes that degrade lipids in the membrane (lipid peroxidation).

The ischemic episode also primes cells for abnormal generation of reactive oxygen molecules, such as superoxide (O_2–), peroxide (H_2O_2), and hydroxyl radicals (OH–). These reactive oxygen molecules are free radicals that have an unpaired electron in an outer orbital. They steal hydrogen atoms and form abnormal molecular bonds. Molecules that react with free radicals may be converted to free radicals themselves, continuing the destructive cascade. Reactive oxygen species damage cell membranes, denature proteins, and disrupt cell chromosomes. Oxygen free radicals also have been linked to initiation of the inflammatory cascade.

Ischemia primes cells for the generation of oxygen radicals by allowing the buildup of ATP precursors, such as adenosine diphosphate and pyruvate, during the period of hypoxia. When oxygen supply is reestablished, a disorganized burst of high-energy electrons partially reduces oxygen and forms oxygen radicals. The ischemia–reperfusion

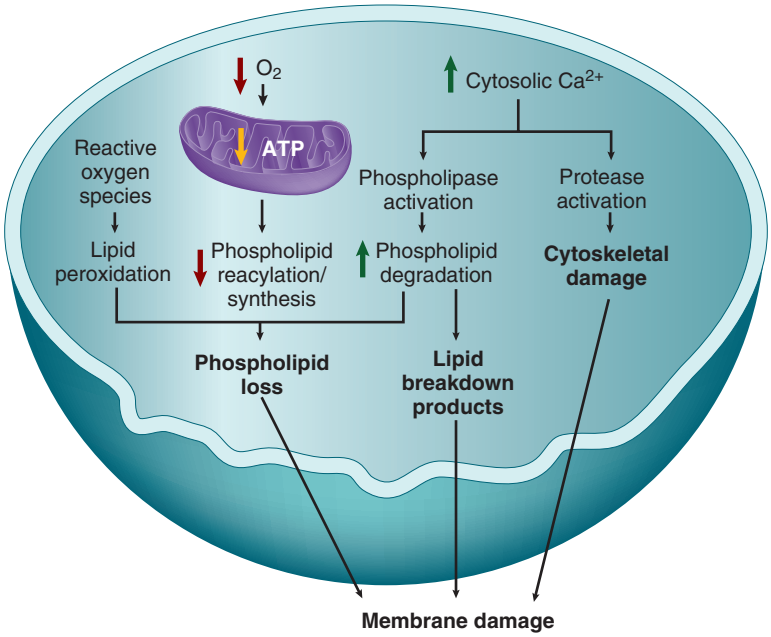


FIG 4.14 Mechanisms of ischemia-induced cell injury and membrane damage. Loss of cell membrane integrity is a critical event in necrosis.

event frequently is followed by a generalized inflammatory state that may lead to ongoing cellular and organ damage for days and weeks after the initial event. WBCs recruited to the area release enzymes and other chemicals that further damage the cells in the area. (Mechanisms and causes of ischemic tissue injury are described further in Chapter 20.)

Nutritional Injury

Adequate amounts of fats, carbohydrates, proteins, vitamins, and minerals are essential for normal cellular function. Most of these essential nutrients must be obtained from external sources because the cell is unable to manufacture them. The cell is unable to synthesize many of the 20 amino acids needed to form the proteins of the body. Likewise, most vitamins and minerals must be obtained from exogenous sources. Cell injury results from deficiencies as well as excesses of essential nutrients.

Certain cell types are more susceptible to injury from particular nutritional imbalances. Iron deficiency, for example, primarily affects RBCs, whereas vitamin D deficiency affects bones. All cell types must receive glucose for energy, as well as fatty acid and amino acid building blocks to synthesize and repair cellular components. Nutritional deficiencies result from poor intake, altered absorption, impaired distribution by the circulatory system, or inefficient cellular uptake. Common causes of malnutrition include (1) poverty, (2) chronic alcoholism, (3) acute and chronic illness, (4) self-imposed dietary restrictions, and (5) malabsorption syndromes. Vitamin deficiencies are common even in industrialized countries because of pervasive use of processed foods. Some examples of vitamin deficiency disorders are shown in Table 4.1. Deficiencies of minerals, especially iron, also are common (Table 4.2).

Nutritional excesses primarily result from excessive intake, although deficient cellular uptake by one cell type may contribute to excess nutrient delivery to other cell types. For example, in the condition of diabetes mellitus, some cell types have deficient receptors for insulin-dependent glucose uptake, which causes excessive amounts of glucose to remain in the bloodstream. As a result, cells that do not require insulin to take in glucose, such as neurons, may have abnormally high intracellular glucose levels. An excess of caloric intake above metabolic use produces overweight and obesity syndromes. Excess body fat can be estimated by measuring the ratio of body weight (in kilograms) to height (in meters squared) to derive the body mass index (BMI). A BMI greater than 25 kg/m^2 imparts a health risk, and a BMI greater than 30 kg/m^2 is considered indicative of obesity. Numerous health problems are associated with excess body fat, including heart and blood vessel disease, musculoskeletal strain, diabetes, hypertension, and gallbladder disease. Metabolism and obesity are explored in Chapter 42.

Infectious and Immunologic Injury

Bacteria and viruses are common infectious agents that may injure cells in a variety of ways. The virulence of a particular biological agent depends on its ability to gain access to the cell and its success in altering cellular functions. (See Chapter 8 for a detailed discussion of infectious processes.) Some of the injurious effects are directly from the biological agent, but added injury may be done indirectly by triggering the body's immune response.

Most bacteria do not gain entry into the cell and so accomplish their injurious effects from the outside. (Notable exceptions include *Mycobacterium tuberculosis, Shigella, Legionella, Salmonella,* and *Chlamydia.*) Some bacteria produce and secrete powerful destructive enzymes that digest cellular membranes and connective tissues. For example, collagenase and lecithinase are produced by *Clostridium perfringens.* Other bacteria produce exotoxins, which interfere with specific cellular functions when released from the bacterium. *Clostridium botulinum* and *Clostridium tetani,* for example, produce life-threatening toxins that disrupt normal neuromuscular transmission. Cholera and diphtheria are well-known examples of exotoxin-related diseases. Most exotoxins are proteins and are generally susceptible to destruction by extremes of heat. Certain gram-negative bacteria (e.g., *Escherichia coli, Klebsiella pneumoniae*) contain another type of toxin, endotoxin, in their cell wall. When bacteria are killed, the endotoxin is released, causing fever, malaise, and even circulatory shock.

The indirect cellular injury attributable to the bacteria-evoked immune response may be more damaging than the direct effects of the

TABLE 4.1 Vitamins: Major Functions and Deficiency Syndromes

Vitamin	Functions	Deficiency Syndromes
Fat Soluble		
Vitamin A	A component of visual pigment Maintenance of specialized epithelia Maintenance of resistance to infection	Night blindness, xerophthalmia, blindness Squamous metaplasia Vulnerability to infection, particularly measles
Vitamin D	Facilitates intestinal absorption of calcium and phosphorus and mineralization of bone	Rickets in children Osteomalacia in adults
Vitamin E	Major antioxidant; scavenges free radicals	Spinocerebellar degeneration
Vitamin K	Cofactor in hepatic carboxylation of procoagulants—factors II (prothrombin), VII, IX, and X; and protein C and protein S	Bleeding diathesis
Water-Soluble		
Vitamin B_1 (thiamine)	As pyrophosphate, is coenzyme in decarboxylation reactions	Dry and wet beriberi, Wernicke syndrome, Korsakoff syndrome
Vitamin B_2 (riboflavin)	Converted to coenzymes flavin mononucleotide and flavin adenine dinucleotide, cofactors for many enzymes in intermediary metabolism	Ariboflavinosis, cheilosis, stomatitis, glossitis, dermatitis, corneal vascularization
Niacin	Incorporated into NAD and NAD phosphate; involved in a variety of redox reactions	Pellagra—"three D's": dementia, dermatitis, diarrhea
Vitamin B_6 (pyridoxine)	Derivatives serve as coenzymes in many intermediary reactions	Cheilosis, glossitis, dermatitis, peripheral neuropathy
Vitamin B_{12}	Required for normal folate metabolism and DNA synthesis Maintenance of myelinization of spinal cord tracts	Megaloblastic pernicious anemia and degeneration of posterolateral spinal cord tracts
Vitamin C	Serves in many oxidation-reduction (redox) reactions and hydroxylation of collagen	Scurvy
Folate	Essential for transfer and use of 1-carbon units in DNA synthesis	Megaloblastic anemia, neural tube defects
Pantothenic acid	Incorporated in coenzyme A	No nonexperimental syndrome recognized
Biotin	Cofactor in carboxylation reactions	No clearly defined clinical syndrome

From Kumar V et al: *Robbins and Cotran pathologic basis of disease,* ed 9, Philadelphia, 2015, Saunders, p 442.
NAD, Nicotinamide adenine dinucleotide.

TABLE 4.2 Selected Trace Elements and Deficiency Syndromes

Element	Function	Basis of Deficiency	Clinical Features
Zinc	Component of enzymes, principally oxidases	Inadequate supplementation in artificial diets Interference with absorption by other dietary constituents Inborn error of metabolism	Rash around eyes, mouth, nose, and anus called acrodermatitis enteropathica Anorexia and diarrhea Growth retardation in children Depressed mental function Depressed wound healing and immune response Impaired night vision Infertility
Iron	Essential component of hemoglobin as well as a number of iron-containing metalloenzymes	Inadequate diet Chronic blood loss	Hypochromic microcytic anemia
Iodine	Component of thyroid hormone	Inadequate supply in food and water	Goiter and hypothyroidism
Copper	Component of cytochrome *c* oxidase, dopamine β-hydroxylase, tyrosinase, lysyl oxidase, and unknown enzymes involved in cross-linking collagen	Inadequate supplementation in artificial diet Interference with absorption	Muscle weakness Neurologic defects Abnormal collagen cross-linking
Fluoride	Mechanism unknown	Inadequate supply in soil and water Inadequate supplementation	Dental caries
Selenium	Component of glutathione peroxidase Antioxidant with vitamin E	Inadequate amounts in soil and water	Myopathy Cardiomyopathy (Keshan disease)

From Kumar V et al: *Robbins and Cotran pathologic basis of disease,* ed 9, Philadelphia, 2015, Saunders, p 443.

infectious agent. WBCs secrete many enzymes and chemicals meant to destroy the invading organism, including histamines, kinins, complement, proteases, lymphokines, and prostaglandins. Normal body cells may be exposed to these injurious chemicals because they are too close to the site of immunologic battle. Immune cells are particularly adept at producing free radicals, which can attack host cell membranes and induce significant cell injury.

Viruses are small pieces of genetic material that are able to gain entry into the cell. They may be regarded as intracellular parasites that use the host cell's metabolic and synthetic machinery to survive and

replicate. In some cases the virus remains in the cell for a considerable time without inflicting lethal injury. In other cases the virus causes rapid lysis and destruction of the host cell.

Virally infected cells may trigger their own destruction when they express viral proteins on the cell surface that are foreign to the host's immune system. The hepatitis B virus is an example of such an indirectly cytopathic virus that causes immune-mediated cell death. The hepatitis B virus consists of double-stranded DNA that becomes incorporated into the host cell's nucleus, where it can be transcribed by the normal DNA polymerases. The mRNA transcripts of the viral genes are transported to the cytoplasm and translated into structural proteins and enzymes, which are used to make more copies of the virus. Some of these viral components are displayed on the cell surface where cytotoxic cells can recognize them, triggering death of the infected cells (see Chapter 9). Virally infected cells may remain functional virus factories until they are destroyed by the host's immune system.

Chemical Injury

Toxic chemicals or poisons are plentiful in the environment (Tables 4.3 and 4.4). Some toxic chemicals cause cellular injury directly, whereas others become injurious only when metabolized into reactive chemicals by the body. Carbon tetrachloride (CCl_4) is an example of the latter. Carbon tetrachloride, a formerly used dry-cleaning agent, is converted to a highly toxic free radical, CCl_3–, by liver cells. The free radical is highly reactive, forming abnormal chemical bonds in the cell and ultimately destroying the cellular membranes of liver cells, causing liver failure. In high doses, acetaminophen, a commonly used analgesic, may have similar toxic effects on the liver.

Many toxins are inherently reactive and do not require metabolic activation to exert their effects. Common examples are heavy metals (e.g., lead and mercury), toxic gases, corrosives, and antimetabolites. Some toxins have an affinity for a particular cell type or tissue, whereas others exert widespread systemic effects. For example, carbon monoxide binds tightly and selectively to hemoglobin, preventing the RBC from carrying sufficient oxygen. Lead poisoning, however, has widespread effects, including effects on nervous tissue, blood cells, and the kidney. Extremely acidic or basic chemicals are directly corrosive to cellular structures. Certain chemicals interfere with normal metabolic processes of the cell. Some of these antimetabolites have been utilized in the form of cytotoxic agents for the management of cancer (Chapter 7).

Physical and Mechanical Injury

Injurious physical and mechanical factors include extremes of temperature, abrupt changes of atmospheric pressure, mechanical deformation, electricity, and ionizing radiation.

Extremes of cold result in the hypothermic injury known as *frostbite*. Before actual cellular freezing, severe vasoconstriction and increased blood viscosity may result in ischemic injury. With continued exposure to cold, a rebound vasodilatory response may occur, leading to intense swelling and peripheral nerve damage. The cytoplasmic solution may freeze, resulting in the formation of intracellular ice crystals and rupture of cellular components. Frostbite generally affects the extremities, ears, and nose and is often complicated by gangrenous necrosis.

Extremes of heat result in hyperthermic injury or burns. High temperatures cause microvascular coagulation and may accelerate metabolic processes in the cell. Burns result from direct tissue destruction by high temperatures and are classified according to the degree of tissue destruction. Burns are discussed in Chapter 54.

Abrupt changes in atmospheric pressure may result from high-altitude flying, deep-sea diving, and explosions. Pressure changes may interfere with gas exchange in the lungs, cause the formation of gas emboli in the bloodstream, collapse the thorax, and rupture internal organs. A well-known example of pressure injury is the condition of "the bends," which afflicts deep-sea divers who surface too quickly. The rapid decrease in water pressure results in the formation of bubbles of nitrogen gas in the blood, which may obstruct the circulation and cause ischemic injury.

Destruction of cells and tissues resulting from mechanical deformation ranges from mild abrasion to severe lacerating trauma. Cell death may result from direct trauma to cell membranes and resulting blood loss or from obstruction of blood flow and hypoxia. Nonpenetrating trauma generally results from physical impact with a blunt object such as a fist, a car steering wheel, or the pavement. Surgery is a common cause of tissue trauma. Other causes of penetrating trauma are bite, knife, and

TABLE 4.3 Health Effects of Outdoor Air Pollutants

Pollutant	Populations at Risk	Effects*
Ozone	Healthy adults and children Athletes, outdoor workers Asthmatics	Decreased lung function Increased airway reactivity Lung inflammation Decreased exercise capacity Increased hospitalizations
Nitrogen dioxide	Healthy adults Asthmatics Children	Increased airway reactivity Decreased lung function Increased respiratory infections
Sulfur dioxide	Healthy adults Patients with chronic lung disease Asthmatics	Increased respiratory symptoms Increased mortality Increased hospitalization Decreased lung function
Acid aerosols	Healthy adults Children Asthmatics	Altered mucociliary clearance Increased respiratory infections Decreased lung function Increased hospitalizations
Particulates	Children Individuals with chronic lung or heart disease Asthmatics	Increased respiratory infections Decreased lung function Excess mortality Increased attacks

From Kumar V et al: *Robbins and Cotran pathologic basis of disease*, ed 9, Philadelphia, 2015, Saunders, p 408.
*See Chapters 22 and 23 for a discussion of respiratory disorders.

TABLE 4.4 Selected Indoor Air Pollutants With Significant Health Risks

Pollutant	Source
Carbon monoxide	Fuel combustion, fire, furnace
Wood smoke	Fireplaces, woodstoves
Formaldehyde	Manufacture of construction materials
Radon	Natural ground radiation
Asbestos fibers	Old insulation, shingles
Manufactured mineral fibers	Insulation, building materials
Aerosols	Spray bottle propellants

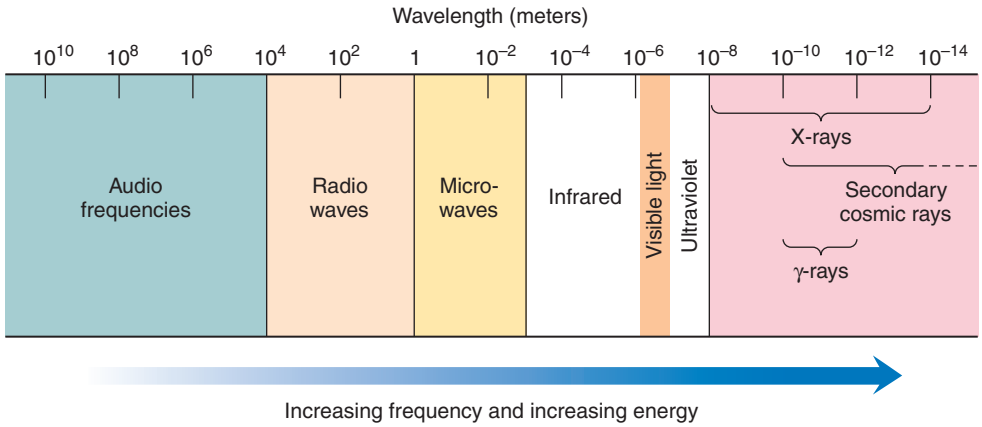


FIG 4.15 Types of electromagnetic radiation.

gun wounds. Trauma-induced inflammatory swelling may further compromise injured tissues.

Electrical injury may occur when the cells of the body act as conductors of electricity. The electrical current damages tissues in two ways: (1) by disruption of neural and cardiac impulses and (2) by hyperthermic destruction of tissues. Resistance to the flow of electrons results in heat production, which damages the tissues. The current tends to follow the path of least resistance—through neurons and body fluids—causing violent muscle contractions, thermal injury, and coagulation in blood vessels. In general, greater injury is suffered with high-voltage alternating current applied to a low-resistance area (e.g., wet skin).

There are many forms of electromagnetic radiation, ranging from low-energy radio waves to high-energy γ-rays or photons (Fig. 4.15). Radiation is capable of injuring cells directly by breaking chemical bonds and indirectly by generating free radicals. Cellular DNA is particularly susceptible to damage from radiation exposure. A direct hit of the radiant energy on the DNA molecule may result in breakage of the chemical bonds holding the linear DNA together. This type of direct bond breakage generally results from the high-energy forms of radiation, such as x-rays and γ-rays. The molecular bonds of DNA also may be indirectly disrupted by ionizing radiation. *Ionization* refers to the ability of the radiant energy to split water molecules by knocking off orbital electrons (radiolysis). Radiolysis creates activated free radicals that steal electrons from other molecules and disrupt chemical bonds. Many forms of radiation are capable of ionization, but the medium-energy α and β particles that result from decay of atomic nuclei are especially destructive. Low-energy electromagnetic radiation, such as that created by microwaves, ultrasound, computers, and infrared light, cannot break chemical bonds, but it can cause rotation and vibration of atoms and molecules. The rotational and vibrational energy is then converted to heat. It is possible that the resulting localized hyperthermia may result in cellular injury. Early studies reported a higher incidence of certain cancers in persons occupationally exposed to radiofrequency microwave electromagnetic radiation, but further analysis failed to confirm these findings.

At the cellular level, radiation has two primary effects: (1) genetic damage and (2) acute cell destruction (Fig. 4.16). The vulnerability of a tissue to radiation-induced genetic damage depends on its rate of proliferation. Genetic damage to the DNA of a long-lived, nonproliferating cell may be of little consequence, whereas tissues with rapid cellular division have less opportunity to repair damaged DNA before passing it on to the next generation of cells. (Genetic mutation is discussed in Chapter 6.) Hematopoietic, mucosal, gonadal, and fetal cells are particularly susceptible to genetic radiation damage.

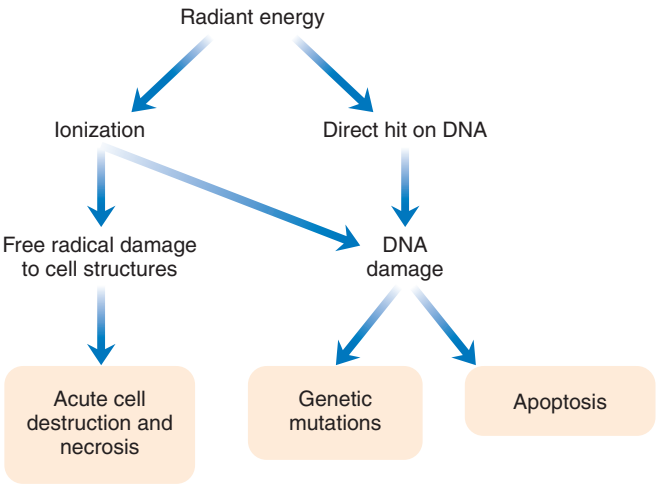


FIG 4.16 The mechanism of radiation-induced genetic and cell injury.

Radiation-induced cell death is attributed primarily to the radiolysis of water, with resulting free radical damage to the plasma membrane. Whole-body exposure to sufficiently high levels of radiation (300 rad) results in acute radiation sickness with hematopoietic failure, destruction of the epithelial layer of the gastrointestinal tract, and neurologic dysfunction. The high levels of irradiation that cause acute radiation sickness are associated with events such as nuclear accidents and bombings. Radiation exposure from diagnostic x-rays, cosmic rays, and natural radiant chemicals in the earth is far below the level that would result in acute radiation sickness. The signs and symptoms of acute radiation sickness are shown in Fig. 4.17. The fact that radiation induces cell death in proliferating cells is used to advantage in the management of some forms of cancer. Radiation therapy may be used when a cancerous growth is confined to a particular area. Injury associated with radiation therapy is generally localized to the irradiated area. Small arteries and arterioles in the area may be damaged, leading to blood clotting and fibrous deposits that compromise tissue perfusion. Most irradiated cells are thought to die through the process of apoptosis rather than from direct killing effects of radiation. Radiation induces cell damage that triggers the apoptotic pathway in cells that cannot efficiently repair the damage. Cells most susceptible to apoptotic death are those that tend to have high rates of mitosis or meiosis.

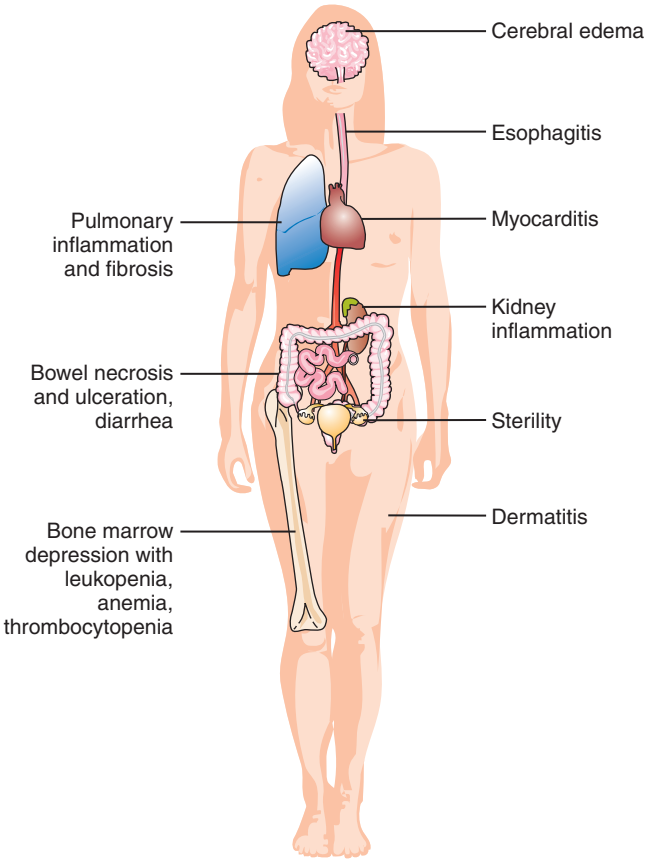


FIG 4.17 Signs and symptoms of acute radiation sickness.

KEY POINTS

- Hypoxia is an important cause of cell injury that usually results from poor oxygenation of the blood (hypoxemia) or inadequate delivery of blood to the cells (ischemia).
- Reperfusion injury to cells may occur when circulation is restored as a result of the production of partially reduced oxygen molecules that damage cell membranes and trigger immune-mediated injury.
- Nutritional injury is a common cause of dysfunction and disease. Malnutrition is rampant in many poor countries, whereas industrialized nations are facing an epidemic of obesity-related disorders, including heart disease and diabetes.
- Cellular damage attributable to infection and immunologic responses is common. Some bacteria and viruses damage cells directly, whereas others stimulate the host's immune system to destroy the host's cells.
- Chemical, physical, and mechanical factors cause cell injury in various ways. Chemicals may interfere with normal metabolic processes in the cell. Injury resulting from physical factors, such as burns and frostbite, causes direct destruction of tissues. Radiation-induced cell death is primarily a result of radiolysis of water, with resulting free radical damage to the cell membrane.

CELLULAR AGING

The inevitable process of aging and death has been the subject of interest and investigation for centuries. Despite scientific study and the search for the "fountain of youth," a satisfactory explanation for the process of cellular aging and methods for halting the aging process have not been revealed. The maximal human life span currently is about 120 years, and the oldest-known human to have lived was 122 years old. It seems apparent that aging is distinct from disease and that the life span is limited by the aging process itself rather than by the ravages of disease. Although the elderly are certainly more vulnerable to diseases, the aging process and disease processes are generally viewed as different phenomena. In practice, the distinction between aging and disease may be difficult to make. For example, the aging skeleton normally loses some bone mass, but too much bone loss results in osteoporosis—a disease process. Likewise, a loss of blood vessel elasticity is generally viewed as a normal aging change, but at what point does too much arterial stiffness become abnormal? This confusion results from the continued inability to identify the irreversible and universal processes of cellular aging as separate from the potentially reversible effects of disease.

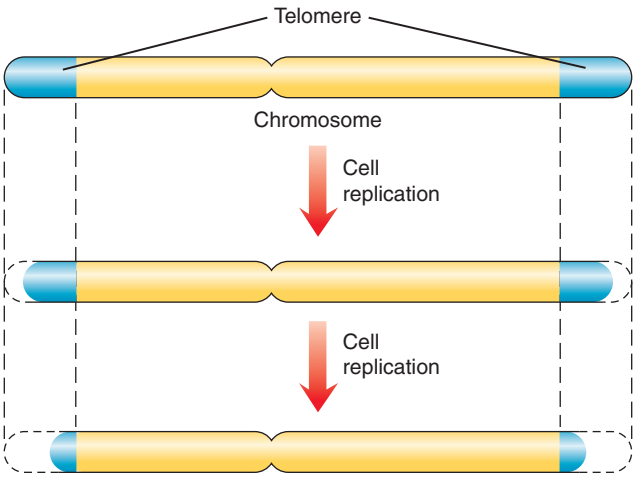


FIG 4.18 The end caps of the chromosomes are called telomeres. In most body cells, the telomeres progressively shorten with each cell replication until a critical point is reached, at which time the cell becomes dormant or dies.

Cellular Basis of Aging

Cellular aging is the cumulative result of a progressive decline in the proliferative and reparative capacity of cells coupled with exposure to environmental factors that cause accumulation of cellular and molecular damage. Several mechanisms are believed to be responsible for cellular aging. These include DNA damage, reduced proliferative capacity of stem cells, and accumulation of metabolic damage.

Damage to cellular DNA is a common phenomenon resulting from various factors, including ultraviolet radiation, oxidative stress from normal metabolism, and errors in DNA replication. A host of DNA repair mechanisms is present in normal cells to prevent accumulation of DNA damage. With aging these repair systems appear to become less able to keep pace with DNA damage, and cell replication may be inhibited or apoptosis initiated. Support for this idea comes from the premature aging syndromes that are associated with defective DNA repair mechanisms.

The *programmed senescence theory* states that aging is the result of an intrinsic genetic program. Support for the theory of a genetically programmed life span comes primarily from studies of cells in culture. In classic experiments by Hayflick, fibroblastic cells in culture were shown to undergo a finite number of cell divisions. Fibroblasts taken from older individuals underwent fewer cell divisions than those from younger individuals. Given an adequate environment, the information encoded in the cellular genome is thought to dictate the number of possible cell replications, after which damaged or lost cells are no longer replaced. It has been postulated that cells undergo a finite number of replications because the chromosomes shorten slightly with each cell division until some critical point is reached (Fig. 4.18), at which time the cell becomes dormant or dies. The end caps of the chromosomes, called *telomeres*, are the sections that shorten with each cell division.

Certain cells (germ cells, such as egg and sperm) are able to replenish their telomeres, which gives them potential immortality. The enzyme that rebuilds the telomeres has been named *telomerase.* Stem cells, which are capable of mitosis, also express telomerase, but at lower levels. Progressive loss of telomerase gene expression with aging may contribute to reduced proliferative capacity. Interestingly, a number of cancer cell types have been found to produce telomerase, whereas most differentiated somatic cells do not (Chapter 7).

Aging may also be a result of accumulated metabolic cell damage over time. The *free radical theory* was prompted in part by the observation that larger animals, which have slower metabolic rates, generally have longer life spans. Metabolic rate, in turn, determines the production of activated oxygen free radicals. Aging is thought to result from the cumulative and progressive damage to cell structures, particularly the cell membrane, by these oxygen radicals. Protection from metabolic damage is provided by a number of antioxidant mechanisms. Over time these protective mechanisms may become less efficient, allowing metabolic damage to accumulate in cells. Accumulated damage may eventually trigger apoptotic mechanisms leading to tissue degeneration. Altered metabolism from chronic low-calorie diets has been shown to prolong life span in studies.

Physiologic Changes of Aging

All the body systems show age-related changes that can be generally described as a decrease in functional reserve or impaired ability to adapt to environmental demands. An overview of the tissue and systemic changes of aging is presented in Table 4.5. The details of age-related changes in the various body systems are described in later chapters of this book.

KEY POINTS

- Aging is theoretically distinct from disease. The maximal life span is limited by the aging process itself rather than by the ravages of disease.
- Aging is thought to be the result of accumulated DNA damage, decreased proliferative capacity of stem cells, and accumulated metabolic damage. Cells may age more quickly when DNA repair mechanisms are faulty and when metabolic damage is excessive because of reduced antioxidant activity.
- Age-related changes in body systems can generally be described as a decrease in functional reserve and a reduced ability to adapt to environmental demands.

TABLE 4.5 Overview of the Physiologic Changes of Aging

System	Physiologic Changes
Cardiovascular	↓ Vessel elasticity caused by calcification of connective tissue (↑ pulmonary vascular resistance)
	↓ Number of heart muscle fibers with ↑ size of individual fibers (hypertrophy)
	↓ Filling capacity
	↓ Stroke volume
	↓ Sensitivity of baroreceptors
	Degeneration of vein valves
Respiratory	↓ Chest wall compliance resulting from calcification of costal cartilage
	↓ Alveolar ventilation
	↓ Respiratory muscle strength
	Air trapping and ↓ ventilation due to degeneration of lung tissue (↓ elasticity)
Renal/urinary	↓ Glomerular filtration rate due to nephron degeneration (↓ one third to one half by age 70)
	↓ Ability to concentrate urine
	↓ Ability to regulate H+ concentration
Gastrointestinal	↓ Muscular contraction
	↓ Esophageal emptying
	↓ Bowel motility
	↓ Production of HCl, enzymes, and intrinsic factor
	↓ Hepatic enzyme production and metabolic capacity
	Thinning of stomach mucosa
Neurologic/sensory	Nerve cells degenerate and atrophy
	↓ Of 25%–45% of neurons
	↓ Number of neurotransmitters
	↓ Rate of conduction of nerve impulses
	Loss of taste buds
	Loss of auditory hair cells and sclerosis of eardrum
Musculoskeletal	↓ Muscle mass
	↑ Bone demineralization
	↑ Joint degeneration, erosion, and calcification
Immune	↓ Inflammatory response
	↓ In T cell function due to involution of thymus gland
Integumentary	↓ Subcutaneous fat
	↓ Elastin
	Atrophy of sweat glands
	Atrophy of epidermal arterioles causing altered temperature regulation

SOMATIC DEATH

Death of the entire organism is called *somatic death.* In contrast to localized cell death, no immunologic or inflammatory response occurs in somatic death. The general features of somatic death include the absence of respiration and heartbeat. However, this definition of death is insufficient because, in some cases, breathing and cardiac activity may be restored by resuscitative efforts. Within several minutes of cardiopulmonary arrest, the characteristics of irreversible somatic death become apparent. Body temperature falls, the skin becomes pale, and blood and body fluids collect in dependent areas. Within 6 hours, the accumulation of calcium and the depletion of ATP result in perpetual actin–myosin cross-bridge formation in muscle cells. The presence of stiffened muscles throughout the body after death is called *rigor mortis.* Rigor mortis progresses to limpness or flaccidity as the tissues of the body begin to deteriorate. Tissue deterioration or putrefaction becomes apparent 24 to 48 hours after death. Putrefaction is associated with the widespread release of lytic enzymes in tissues throughout the body, a process called *postmortem autolysis.*

The determination of "brain death" has become necessary because of the technological ability to keep the heart and lungs working through artificial means, even though the brain is no longer functional. Criteria for determining brain death as proof of somatic death may vary by geographic area but generally include unresponsiveness, flaccidity, absence of brainstem reflexes (e.g., swallowing, gagging, pupil and eye movements), absence of respiratory effort when the subject is removed from the mechanical ventilator, absence of electrical brain waves, and lack of cerebral blood flow.

KEY POINTS

- Somatic death is characterized by the absence of respirations and heartbeat. Definitions of brain death have been established to describe death in instances in which heartbeat and respiration are maintained mechanically.
- After death, body temperature falls, blood and body fluids collect in dependent areas, and rigor mortis ensues. Within 24 to 48 hours the tissues begin to deteriorate and rigor mortis gives way to flaccidity.

SUMMARY

Cells and tissues face many challenges to survival, including injury from lack of oxygen and nutrients, infection and immune responses, chemicals, and physical and mechanical factors. Cells respond to environmental changes or injury in three general ways: (1) If the change is mild or short lived, the cell may withstand the assault and return to its preinjury status. (2) The cell may adapt to a persistent but sublethal injury by changing its structure or function. (3) Cell death by apoptosis or necrosis may occur if the injury is too severe or prolonged. Characteristics of reversible cell injury include hydropic swelling and the accumulation of abnormal substances. Cell necrosis is characterized by irreversible loss of function, release of internal cellular components into the bloodstream, and an inflammatory response. The disruption of the permeability barrier of the plasma membrane is a critical event in necrotic cellular death. Apoptosis is characterized by a tidy, noninflammatory autodigestion of the cell.

Aging is a normal physiologic process characterized by a progressive decline in functional capacity and adaptive ability. The biological basis of aging remains largely a mystery, but several theories have been proposed to explain certain aspects of the process. At present, most sources differentiate between the biological alterations of aging and the alterations consequent to disease processes. In practice, however, the distinction may be difficult to make.

RESOURCES

Cellular Responses to Injury

Alberts B, et al: Cancer. In Alberts B, et al, editors: *Molecular biology of the cell*, ed 6, New York, 2015, Garland Science, p 1115.

Alberts B, et al: Cell death. In Alberts B, et al, editors: *Molecular biology of the cell*, ed 6, New York, 2015, Garland Science, pp 1021–1034.

Kumar V, Abbas A, Aster J: Cellular responses to stress and toxic insults: adaptation, injury, and death. In Kumar V, Abbas A, Aster J, editors: *Robbins and Cotran pathologic basis of disease*, ed 9, Philadelphia, 2015, Saunders, pp 31–68.

Maitra A: The endocrine system. In Kumar V, Abbas A, Aster J, editors: *Robbins and Cotran pathologic basis of disease*, ed 9, Philadelphia, 2015, Saunders, pp 1073–1140.

Nadal-Ginard B, et al: Myocyte death, growth, and regeneration in cardiac hypertrophy and failure. *Circ Res* 92:139–150, 2003.

Nikoletopoulou V, Markaki M, Palikaras K: Crosstalk between apoptosis, necrosis, and autophagy. *Biochem Biophys Acta* 2013.

Theise N, Liu C: Liver and gallbladder. In Kumar V, Abbas A, Aster J, editors: *Robbins and Cotran pathologic basis of disease*, ed 9, Philadelphia, 2015, Saunders, pp 821–882.

Van Berlo JH, Maillet M, Molkentin JD: Signaling effectors underlying pathologic growth and remodeling of the heart. *J Clin Invest* 123:37, 2013.

Types and Causes of Cell Injury

D'Andrea JA, Ziriax JM, Adair ER: Radio frequency electromagnetic fields: mild hyperthermia and safety standards. *Prog Brain Res* 162:107–135, 2007.

Guo MF, Yu JZ, Ma CG: Mechanisms related to neuron injury and death in cerebral hypoxic ischaemia. *Folia Neuropathol* 49(2):78–87, 2011.

Jaeschke H: Reactive oxygen and mechanisms of inflammatory liver injury: present concepts. *J Gastroenterol Hepatol* 26(Suppl 1):173–179, 2011.

Kumar V, Abbas A, Aster J: Environmental and nutritional diseases. In Kumar V, Abbas A, Aster J, editors: *Robbins and Cotran pathologic basis of disease*, ed 9, Philadelphia, 2015, Saunders, pp 403–450.

McAdam A, Milner D, Sharpe A: Infectious diseases. In Kumar V, Abbas A, Aster J, editors: *Robbins and Cotran pathologic basis of disease*, ed 9, Philadelphia, 2015, Saunders, pp 341–402.

National Institutes of Health (NIH): *Managing overweight and obesity in adults: systematic evidence review from the obesity expert panel*, 2013. http://www.nhlbi.nih.gov/guidelines.

Speirs CK, Hwang M, Kim S, et al: Harnessing the cell death pathway for targeted cancer treatment. *Am J Cancer Res* 1(1):43–61, 2011.

Todar K: *Todar's online textbook of bacteriology*, 2015. www.textbookofbacteriology.net.

Weber LW, Boll M, Stampfl A: Hepatotoxicity and mechanism of action of haloalkanes: carbon tetrachloride as a toxicological model. *Crit Rev Toxicol* 33(2):105–136, 2003.

Williams D: Radiation carcinogenesis: lessons from Chernobyl. *Oncogene* 27(Suppl 2):S9–S18, 2008.

Cellular Aging and Death

Hayflick L: The biology of human aging. *Adv Pathobiol* 7(2):80–99, 1980.

Rodriguez T: Chromosome caps offer clues to aging. *Scientific American Mind* 24(1):2013.

Shennan T: *Postmortems and morbid anatomy*, ed 3, Baltimore, 1935, William Wood.

Vaupel J: Biodemography of human aging. *Nature* 464(7288):536–542, 2010.

5

Genome Structure, Regulation, and Tissue Differentiation

Jacquelyn L. Banasik

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- How is genetic information stored in the cell and transmitted to progeny of daughter cells during replication?
- How does the simple four-base structure of DNA serve as a template for synthesis of proteins that may contain 20 varieties of amino acids?
- What roles do genes play in determining cell structure and function?
- How is gene expression regulated?
- By what mechanisms can the cells of an organism, which all contain identical genes, become differentiated into divergent cell types?
- What are the general structures and functions of the four main tissue types: epithelial, connective, muscle, and nerve?

CHAPTER OUTLINE

Structure and Function of DNA, 78
Structure of DNA, 78
DNA Replication, 79
Genetic Code, 80
Transcription, 81
Translation, 82
Regulation of the Genome, 84
Transcriptional Controls, 84
Differentiation of Tissues, 86
Cell Diversification and Cell Memory, 86
Mechanisms of Development, 86
Differentiated Tissues, 87
Epithelial Tissue, 87
Connective Tissue, 89
Muscle Tissue, 90
Nervous Tissue, 91

The ability of scientists to study and manipulate genes has evolved at an incredible pace, including the complete sequencing of all 3.2 billion nucleotide base pairs in an entire human genome. A better understanding of the role that genetics plays in cellular function and disease has spurred efforts to develop therapies to correct genetic abnormalities. The science of genetics developed from the premise that invisible, information-containing elements called genes exist in cells and are passed on to daughter cells when a cell divides. The nature of these elements was at first difficult to imagine: What kind of molecule could direct the daily activities of the organism and be capable of nearly limitless replication? The answer to this question was discovered in the late 1940s and was almost unbelievable in its simplicity. It is now common knowledge that genetic information is stored in long chains of stable molecules called deoxyribonucleic acid (DNA). The human genome contains nearly 20,000 protein-encoding genes containing only four different building blocks. These molecules are the deoxyribonucleotides containing the bases *adenine* (A), *cytosine* (C), *guanine* (G), and *thymine* (T). Genes are composed of varying sequences of these four bases, which are linked together by chains of sugar–phosphate bonds. By serving as the templates for the production of body proteins, genes ultimately affect all aspects of an organism's structure and function. The term *genome* was coined to be inclusive of the entire set of genetic and regulatory material in the nucleus and not just the genes. When the sequencing of an entire human genome was completed in 2004, it became clear that the genome is much more complex than the sum of its genes. Only 1.3% of chromosomal DNA codes for proteins, and many DNA sequences code for ribonucleic acid (RNA) molecules that function in the nucleus to regulate gene function. Methods to rapidly survey the DNA sequences of a particular person are available, and genetics is an increasingly important consideration in the etiology, pathogenesis, and pharmacologic treatment of a variety of diseases. However, genetic inheritance involves more than the transfer of genes from parent to offspring. For example, the nutritional exposures of grandparents may influence the metabolic physiology of grandchildren through a process known as epigenetics. Epigenetics is further explored in Chapter 6. Knowledge of the basic principles of genetics and gene regulation is a prerequisite to understanding not only conventional genetic diseases but also nearly every pathophysiologic process. This chapter examines the structure and

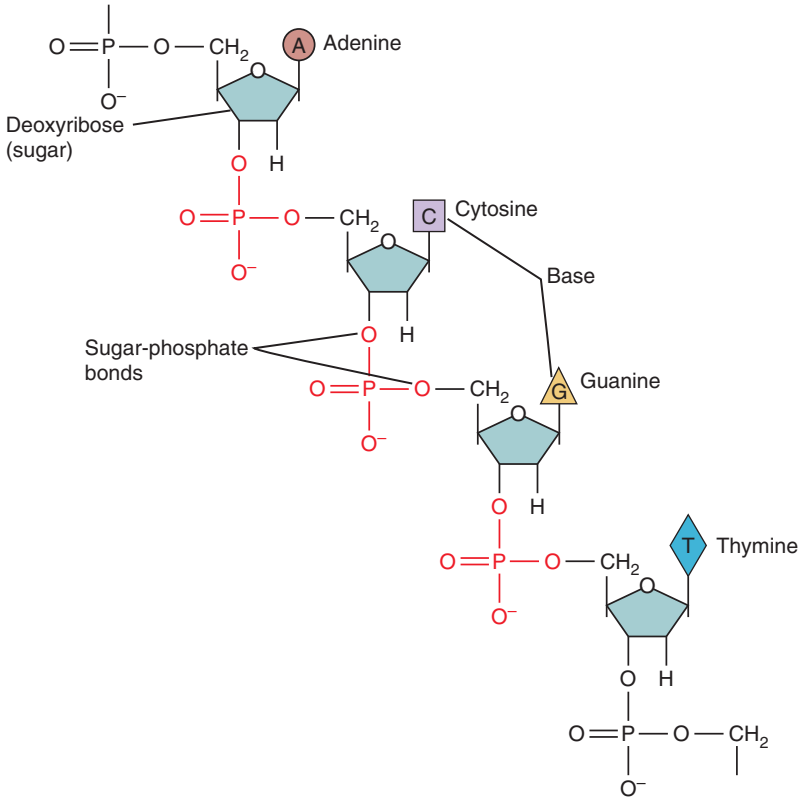


FIG 5.1 A nucleotide consists of a sugar (deoxyribose), a phosphate group, and one of the four nucleotide bases. Nucleotides are joined by repeating sugar–phosphate bonds to form long chains, called polymers. A, Adenine; C, cytosine; G, guanine; T, thymine.

function of DNA, the regulation of gene expression, and the processes of tissue differentiation. Principles of genetic inheritance precede the discussion of genetic diseases in Chapter 6.

STRUCTURE AND FUNCTION OF DNA

Structure of DNA

In humans, DNA encodes genetic information in 46 long, double-stranded chains of nucleotides called *chromosomes.* The nucleotides consist of a five-carbon sugar (deoxyribose), a phosphate group, and one of the four nucleotide bases (Fig. 5.1). The nucleotide bases are divided into two types based on their chemical structure. The pyrimidines, cytosine and thymine, have single-ring structures. The purines, guanine and adenine, have double-ring structures (Fig. 5.2). DNA polymers are formed by the chemical linkage of these nucleotides. The sugar–phosphate linkages, also called *phosphodiester bonds,* join the phosphate group on one sugar (attached to the five-carbon) to the three-carbon of the next sugar (see Fig. 5.1). The four kinds of bases (A, C, G, T) are attached to the repeating sugar–phosphate chain. The bases of one strand of DNA form weak bonds with the bases of another strand of DNA. These noncovalent hydrogen bonds are specific and complementary (Fig. 5.3). The bases G and C always bond together, and the bases A and T always bond together. Nucleotides that are able to bond together in this complementary way are called *base pairs.*

In the early 1950s, Watson and Crick proposed that the structure of DNA was a double helix. In this model, DNA can be envisioned as a twisted ladder, with the sugar–phosphate bonds as the sides of the ladder and the bases forming the rungs (see Fig. 5.3). There is one complete turn of the helix every 10 base pairs. The two strands of DNA must be complementary to form the double helix; that is, the bases of

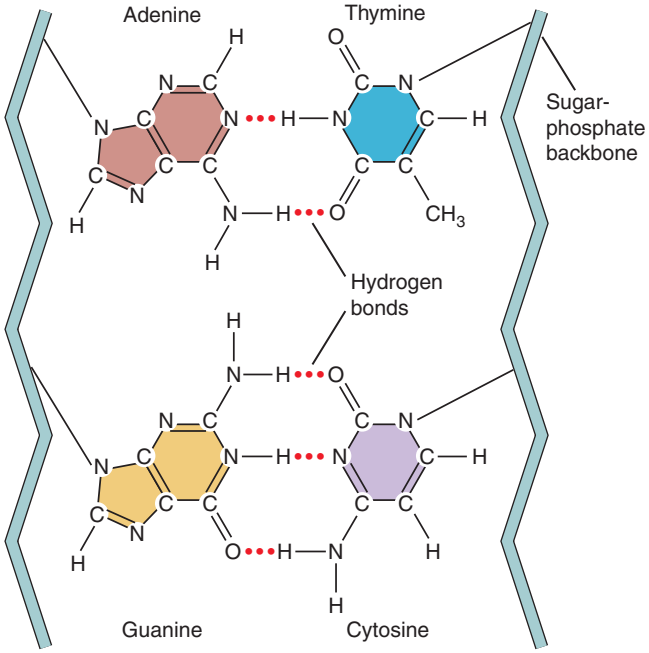


FIG 5.2 The two types of DNA bases are the single-ring pyrimidines and the double-ring purines. Thymine (T) and cytosine (C) are pyrimidines, and adenine (A) and guanine (G) are purines. Base pairing occurs between A and T and between C and G because of hydrogen bonds (dots).

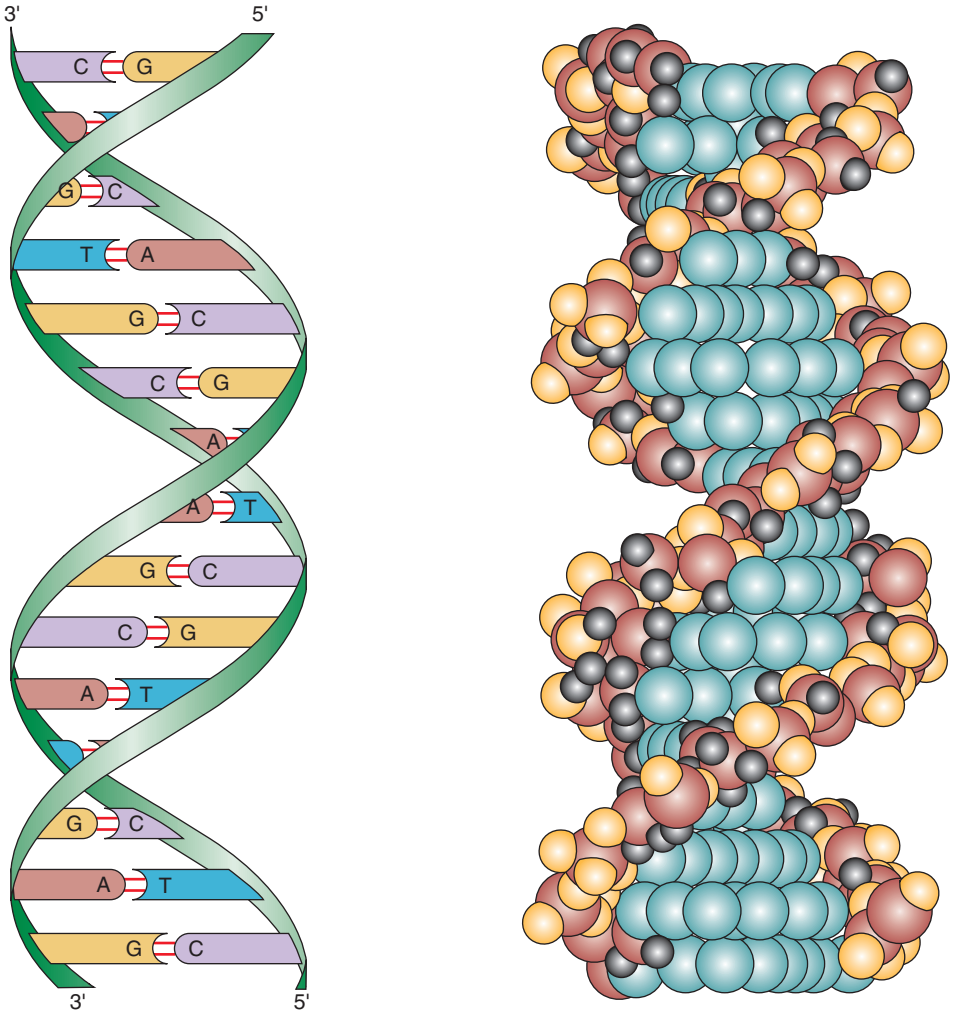


FIG 5.3 A schematic and space-filling model of the DNA double helix as proposed by Watson and Crick. The pairing of bases is specific and complementary: cytosine (C) always pairs with guanine (G), and adenine (A) always pairs with thymine (T).

one strand must pair exactly with their complementary bases on the other strand. The helix is wound around proteins called *histones* to form *nucleosomes* (Fig. 5.4). DNA coupled to histones and other nuclear proteins is termed *chromatin*. When a cell is not dividing, the chromatin is loosely packed within the nucleus and not visible under the light microscope. During cell division, the chromatin becomes tightly condensed into the 46 chromosomes that become visible during mitosis.

The discovery of the double-helix model was profound because it immediately suggested how information transfer could be accomplished by such simple molecules. Because each DNA strand carries a nucleotide sequence that is exactly complementary to the sequence of its partner, both strands can be used as templates to create an exact copy of the original DNA double helix. When a cell divides to form two daughter cells, each daughter cell must receive a complete copy of the parent cell's DNA. The process of DNA replication requires separation of the DNA double helix by breaking the hydrogen bonds between the base pairs. Specific replication enzymes then direct the attachment of the correct (complementary) nucleotides to each of the single-stranded DNA templates. In this way, two identical copies of the original DNA double helix are formed and passed on to the two daughter cells during cell division.

DNA Replication

Although the underlying principle of DNA replication is simple, the cellular machinery required to carry out the replication process is complex, involving a host of enzymes and proteins. These "replication machines" can duplicate DNA at a rate of 1000 nucleotides per second and complete the duplication of the entire genome in about 8 hours. The DNA double helix must first separate so that new nucleotides can be paired with the old DNA template strands. The DNA double helix is normally very stable: the base pairs are locked in place so tightly that they can withstand temperatures approaching the boiling point. In addition, DNA is wrapped around histones and bound by a host of DNA-binding proteins through which the replication machinery must navigate. DNA replication is started by special initiator proteins that pry the DNA strands apart at specific places along the chromatin, called *replication origins*. Then DNA helicases are needed to rapidly unwind and separate the DNA strands, whereas helix-destabilizing proteins (also called *single-stranded DNA-binding proteins*) bind to the exposed DNA strands to keep them apart until replication can be accomplished (Fig. 5.5). Human cells have 30,000 to 50,000 replication origins activated along the chromosomes during DNA replication to speed the process.

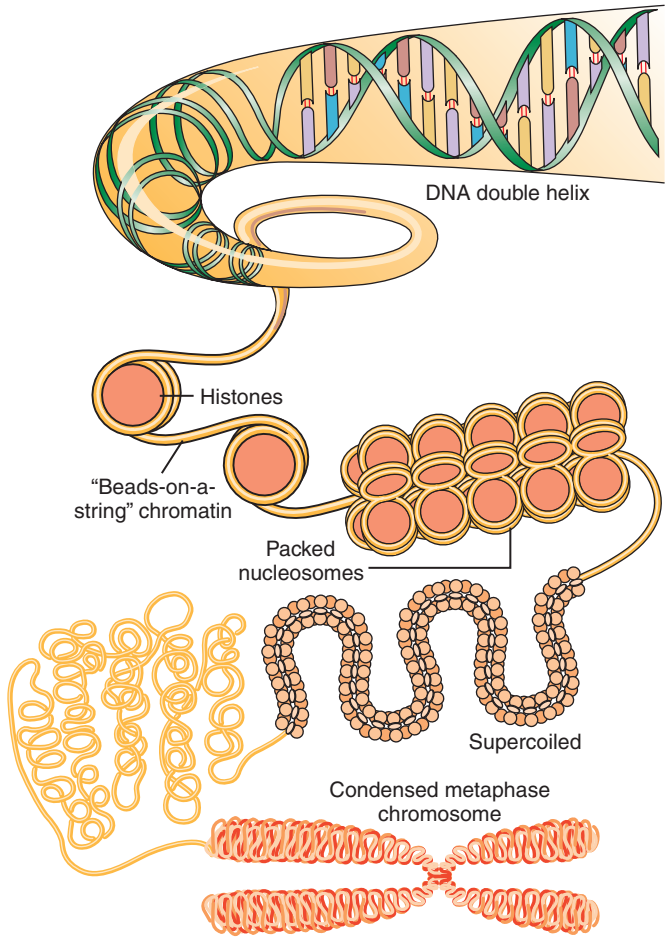


FIG 5.4 DNA is packaged by wrapping around protein complexes called histones to form beadlike structures called nucleosomes. During cell division, the coiled DNA becomes highly condensed into chromosomes that are visible under the light microscope. During interphase and when genes are being transcribed, the DNA is more loosely packaged and not visible.

As the DNA is unwound in the replication fork, it becomes overly twisted downstream, so another set of enzymes, topoisomerases, cuts nicks in the DNA and allows it to unwind to prevent tangling. Ligases repair the nicks.

Once a portion of the DNA double helix has been separated, an enzyme complex, DNA polymerase, binds the single strands of DNA and begins the process of forming a new complementary strand of DNA. DNA polymerase requires a "primer" to begin the process. Primers are formed from RNA by the enzyme DNA primase. The primase must be removed and replaced by DNA once the DNA chain is initiated. The DNA polymerases match the appropriate base to the template base and catalyze the formation of the sugar–phosphate bonds that form the backbone of the DNA strand. Replication proceeds along the DNA strand in one direction only: from the 3′ end toward the 5′ end. The ends of the DNA strands are labeled 3′ and 5′ according to the exposed carbon atom at that end. Because two complementary DNA strands are antiparallel, DNA replication is asymmetric; one strand, the leading strand, is replicated as a continuous polymer, but the lagging strand must be synthesized in short sections in a "backstitching" process (see Fig. 5.5). The backstitched fragments of DNA, called *Okazaki fragments,* are then sealed together by DNA ligase to form the unbroken DNA strand. DNA polymerase is unable to replicate DNA located at the very ends of the chromosomes (the telomeres), so another special enzyme complex, telomerase, is needed for this. The telomeres are fairly short, being composed of approximately 1000 repeats of a GGGTTA sequence. When the telomeres are replicated, one side of the double helix (3′ end) is always longer and loops around and tucks back into the strand. This looping structure is stabilized by a protein called *shelterin* that covers and protects the ends of the chromosomes. This prevents nuclear enzymes from mistaking the ends of the chromosomes for broken DNA ends and trying to attach them to each other. In many somatic cell types, telomerase activity is low and the cell's chromosomes become slightly shorter with each cell division. Chromosomal shortening has been proposed as a mechanism of "counting" the number of replications and may be important in cellular aging and mechanisms of cancer (see Chapter 7). DNA replication is said to be semiconservative because each of the two resulting DNA double helices contains one newly synthesized strand and one original, conserved, strand (Fig. 5.6).

The DNA polymerase also has the ability to proofread the newly synthesized strands for errors in base pairing. If an error is detected, the enzyme will reverse, remove the incorrect nucleotide, and replace it with the correct one. The fidelity of copying during DNA replication is such that only about one error is made for every 10 nucleotide replications. The self-correcting function of the DNA polymerases is extremely important because errors in replication will be transmitted to the next generation of cells.

Genetic Code

How do an organism's genes influence its structural and functional characteristics? A central theory in biology maintains that a gene directs the synthesis of a protein. It is the presence (or absence) and relative activity of various structural proteins and enzymes that produce the characteristics of the cell. This definition of genes as protein-coding elements is not entirely correct because many "genes" code for RNA molecules as their final functional products, and many genes code for more than one protein product through alternate splicing of the RNA messages. Protein synthesis still holds a predominant place in understanding how genes direct cell structure and function.

Proteins are composed of one or more chains of amino acids (polypeptides) that fold into complex three-dimensional structures. Cells contain 20 different types of amino acids that connect in a specific sequence to form a particular protein (Table 5.1). Each type of protein has a unique sequence of amino acids that dictates its structure and activity.

If genes are to direct the synthesis of proteins, the information contained in just four kinds of DNA nucleotide bases must code for 20 different amino acids. This so-called genetic code was deciphered in the early 1960s. It was determined that a series of three nucleotides (triplet) was needed to code for each of the 20 amino acids. Because there are four different bases, there are 4^3, or 64, different possible triplet combinations. This is far more than needed to code for the 20 known amino acids. Three of the nucleotide triplets, or codons**,** do not code for amino acids and are called *stop codons* because they signal the end of a protein code. The remaining 61 codons code for 1 of the 20 amino acids (see Table 5.1). Obviously, some of the amino acids are specified by more than one codon. For example, the amino acid arginine is determined by six different codons. The code has been highly conserved during evolution and is essentially the same in organisms as diverse as humans and bacteria.

Several intermediate molecules are involved in the process of DNA-directed protein synthesis, including the complex protein-synthesizing machinery of the ribosomes and several types of RNA. RNA is structurally similar to DNA, except that the sugar molecule is ribose rather than deoxyribose, and one of the four bases is different in that uracil replaces

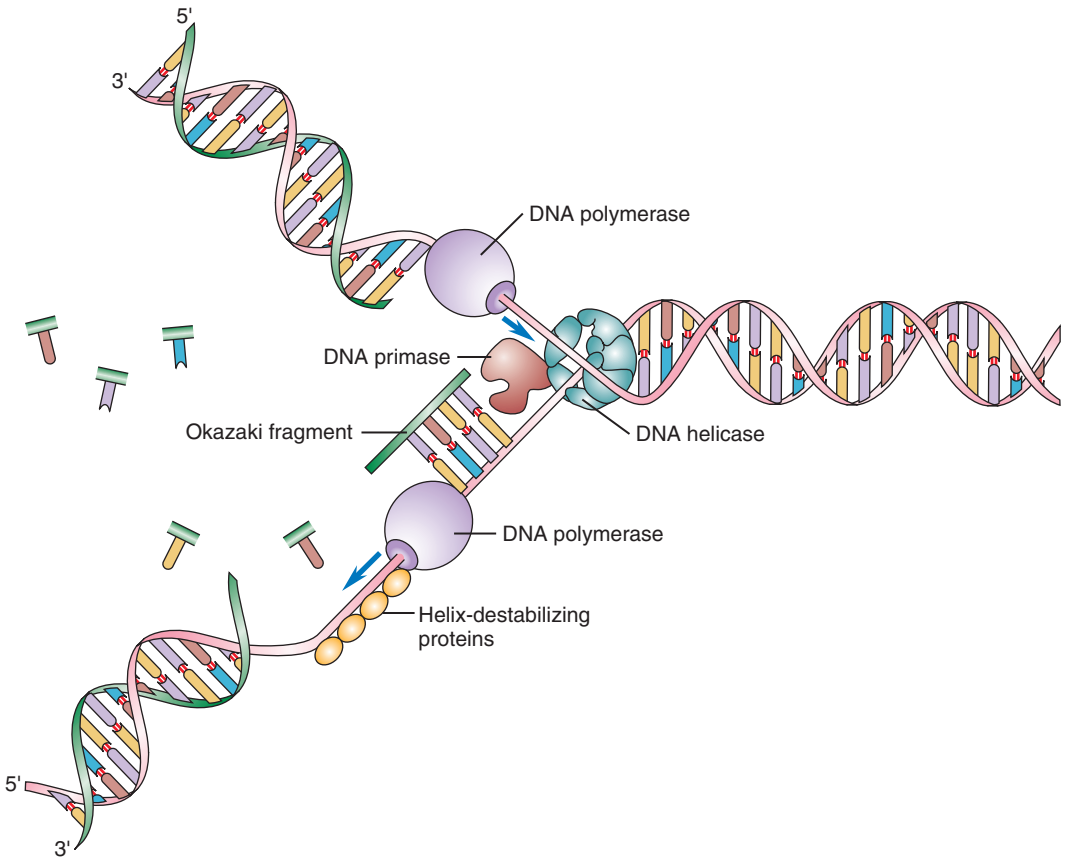


FIG 5.5 Summary of the major proteins of the DNA replication fork. Helicase unwinds the DNA double helix, whereas helix-destabilizing proteins keep the strands from reuniting. The leading strand *(top)* can be replicated in a continuous manner, whereas the lagging strand *(bottom)* must be synthesized in pieces. Okazaki fragments are formed in a "backstitching" direction and then sealed together with DNA ligase. DNA primase synthesizes a short primer of RNA to initiate DNA polymerase activity.

thymine. Because of the biochemical similarity of uracil and thymine, both can form base pairs with adenine. In addition, RNA forms stable single-stranded molecules, whereas DNA strands form a double-stranded molecule.

Several functionally different types of RNA are involved in protein synthesis and cell function. The number and variety of RNA molecules existing within the nucleus is large (Box 5.1), and the exact function of most has yet to be determined. Some perform messenger RNA (mRNA) splicing, ribosome assembly, and quality control of RNA messages before they are transferred to the cytoplasm. The roles of three types of RNA that participate in protein production are well understood. Ribosomal RNA (rRNA) is found associated with the ribosome (see Chapter 3) in the cell cytoplasm. Messenger RNA is synthesized from the DNA template in a process termed transcription and carries the protein code to the cytoplasm, where the proteins are manufactured. The amino acids that will be united to form proteins are carried in the cytoplasm by a third type of RNA, transfer RNA (tRNA), which interacts with mRNA and the ribosome in a process termed translation.

BOX 5.1 Types of RNA Produced in Cells

mRNA—messenger RNA; codes for proteins
rRNA—ribosomal RNA; within ribosomes, catalyzes protein synthesis
tRNA—transfer RNA; adaptors between mRNA and amino acids in protein synthesis
snRNA—small nuclear RNA; splicing of pre-mRNA in the nucleus
snoRNA—small nucleolar RNA; processing of rRNA in the nucleolus
miRNA—micro RNA; regulates gene expression by blocking mRNA translation
siRNA—small interfering RNA; turns off gene expression through alteration in chromatin
piRNA—piwi-interacting RNA; protects the nucleus from transposable elements (like prions)
lncRNA—long noncoding RNA; regulates diverse processes including X-chromosome inactivation in females

Transcription

Transcription is the process whereby mRNA is synthesized from a single-stranded DNA template. The process is similar in some respects to DNA replication. Double-stranded DNA must be separated in the region of the gene to be copied, and specific enzyme complexes (DNA-dependent RNA polymerases) orchestrate the production of the mRNA polymer. Only one of the DNA strands contains the desired gene sequence and serves as the template for the synthesis of mRNA. This strand is called the *sense strand.* The other strand is termed the *nonsense* or *antisense strand* and is not transcribed into an RNA message. There are three types of RNA polymerases in human cells (types I, II, III), and each transcribes different categories of genes. Polymerase II transcribes protein-encoding genes and is the focus of this section.

Some genes are continuously active in certain cells, whereas others are carefully regulated in response to cellular needs and environmental

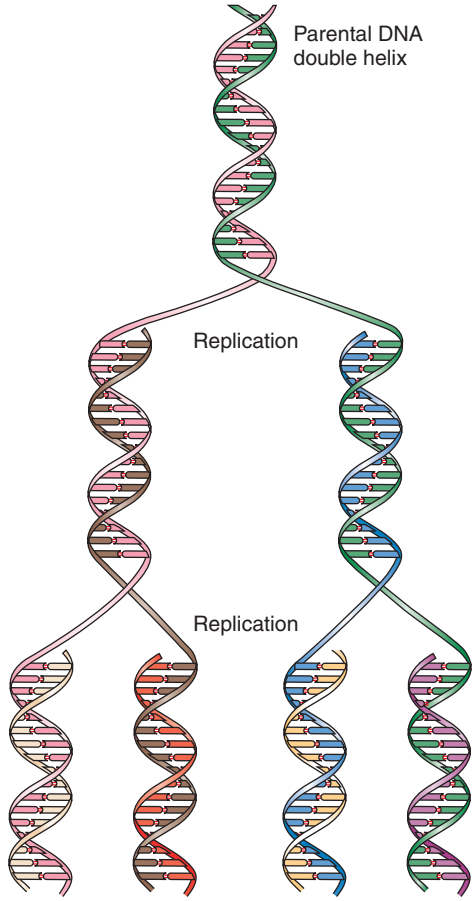


FIG 5.6 DNA replication is semiconservative. Each of the new DNA double helices contains one newly synthesized strand and one original strand.

signals. Special sequences of DNA near a desired gene may enhance or inhibit its rate of transcription. In general, a gene is transcribed when the RNA polymerase–enzyme complex binds to a promoter region just upstream of the gene's start point. This binding event requires the cooperative function of numerous DNA-binding proteins. Once bound at the promoter, the RNA polymerase directs the separation of the DNA double helix and catalyzes the synthesis of the RNA message by matching the appropriate RNA bases to the DNA template (Fig. 5.7). The RNA message is directly complementary to the DNA sequence, except that uracil replaces thymine.

In higher organisms, the DNA template for a particular protein is littered with stretches of bases that must be removed from the original RNA transcript (pre-mRNA) before it can be translated into a protein. These intervening segments, called introns**,** are removed in the nucleus by a complex splicing process, resulting in an mRNA sequence that contains only the wanted segments, called exons**.** Introns range from 10 to 100,000 nucleotides in length. On average, 90% of a gene is composed of introns and only 10% remains in the final mRNA transcript; thus a single gene may contain dozens of introns that must be precisely removed. The function of introns remains largely a mystery, although they are believed to be important in the evolution of new genetic information and in gene regulation. Many of these intron sequences are conserved across species, which implies an important function. The presence of introns also allows alternate splicing so that one gene can be used to code for more than one protein. The removal of introns and splicing of the RNA transcript are mediated by a group of small RNA molecules located in specialized areas of the nucleus called the

TABLE 5.1 RNA Codons for the Different Amino Acids and for Start and Stop

Amino Acids	RNA Codons					
Alanine	GCU	GCC	GCA	GCG		
Arginine	CGU	CGC	CGA	CGG	AGA	AGG
Asparagine	AAU	AAC				
Aspartic acid	GAU	GAC				
Cysteine	UGU	UGC				
Glutamic acid	GAA	GAG				
Glutamine	CAA	CAG				
Glycine	GGU	GGC	GGA	GGG		
Histidine	CAU	CAC				
Isoleucine	AUU	AUC	AUA			
Leucine	CUU	CUC	CUA	CUG	UUA	UUG
Lysine	AAA	AAG				
Methionine	AUG					
Phenylalanine	UUU	UUC				
Proline	CCU	CCC	CCA	CCG		
Serine	UCU	UCC	UCA	UCG	AGC	AGU
Threonine	ACU	ACC	ACA	ACG		
Tryptophan	UGG					
Tyrosine	UAU	UAC				
Valine	GUU	GUC	GUA	GUG		
Start (CI)	AUG					
Stop (CT)	UAA	UAG	UGA			

CI, Chain initiation; *CT,* chain termination.

spliceosomes. The *snRNAs,* or small nuclear RNAs, cause the introns to loop out like a lariat, bringing the adjacent exons close together, followed by cutting and splicing. The snRNAs are combined with proteins to form molecules called *small nuclear ribonucleoproteins* (snRNPs) that form the core of the spliceosome. The snRNPs prevent escape through the nuclear envelope until all the necessary splicing has been accomplished. An estimated 98% of pre-mRNA transcripts can be spliced in different ways to increase the number of different protein forms produced by a single gene.

The processed mRNA is finally transported to the cell cytoplasm through pores in the nuclear membrane that contain complexes that inspect the mRNA for certain structural characteristics that distinguish it from RNA debris. The mRNA then directs the synthesis of a protein in cooperation with tRNA and the ribosomes. Each mRNA may serve as a template for thousands of copies of protein before it is degraded.

Translation

Translation is the process whereby mRNA is used to direct the synthesis of a protein. The mRNA is read in linear fashion from one end to the other, with each set of three nucleotides serving as a codon for a particular amino acid. The codons in the mRNA do not directly recognize the amino acids. Intermediary molecules, or "translators," are required. These intermediaries are the tRNA molecules. A schematic drawing of a tRNA molecule is shown in Fig. 5.8, illustrating its L-shaped, three-dimensional structure. A codon reading area (anticodon) is located at one end and an amino acid attachment at the other. A group of specialized enzymes that have a binding pocket for a particular amino acid and a reading pocket for the anticodon are needed to attach the correct amino acid to its appropriate tRNA. The anticodon is formed by a sequence of three nucleotides. Recognition between the mRNA codon and the tRNA anticodon is accomplished by the same kind of complementary base pairing as was described for DNA. The complex machinery of the

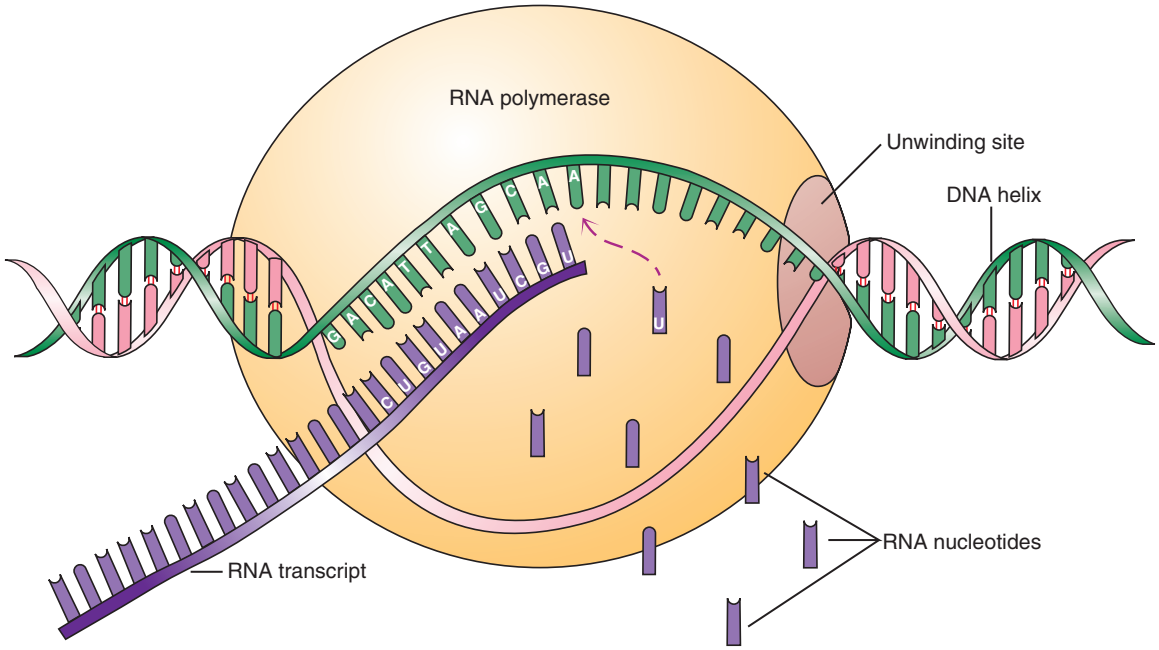


FIG 5.7 A moving RNA polymerase complex unwinds the DNA helix ahead of it while rewinding the DNA behind. One strand of the DNA serves as the template for the formation of mRNA.

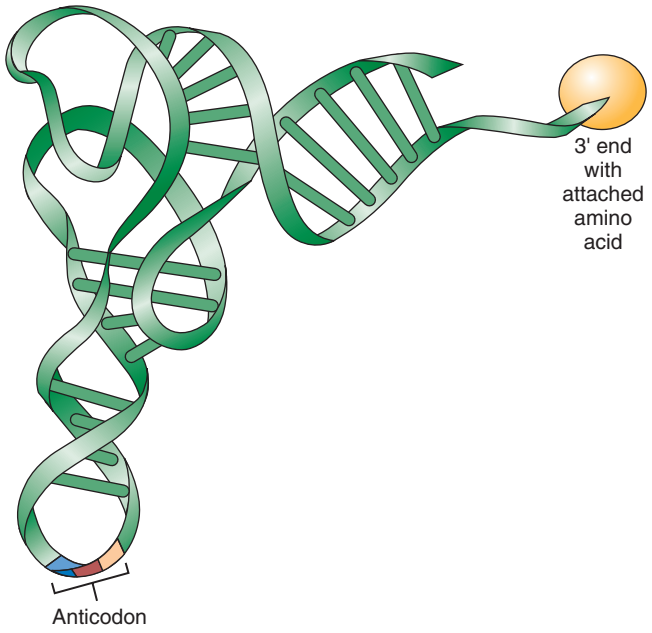


FIG 5.8 Schematic drawing of a transfer RNA (tRNA) molecule. Each tRNA binds a specific amino acid, which corresponds with the three-base sequence at the anticodon end.

ribosome is needed to align the tRNA on the mRNA and to catalyze the peptide bonds that hold the amino acids together. Ribosomes are large complexes of protein and RNA. Each ribosome is composed of two subunits that are first assembled in a special part of the nucleus called the *nucleolus* and then transported through the nuclear pores to the cytoplasm. The smaller subunit binds the mRNA and the tRNA, whereas the larger subunit catalyzes the formation of peptide bonds between the incoming amino acids. The ribosome must first find the appropriate starting place on the mRNA to set the correct reading frame for the codon triplets. Then the ribosome moves along the mRNA, translating the nucleotide sequence into an amino acid sequence, one codon at a time (Fig. 5.9). The newly synthesized protein chain is released from the ribosome when a "stop codon" signaling the end of the message is reached. The new protein is typically bound by "chaperone" proteins that help it fold into its final three-dimensional shape. Amino acids belong to one of three groups—polar, nonpolar, or charged— which affects how the protein is processed and folded into its final structure (Fig. 5.10).

KEY POINTS

- Genes are the basic units of inheritance and are composed of DNA located on chromosomes. Genes direct the daily activities of the cell by controlling the production of proteins. Less than 1.5% of DNA forms genes that code for proteins. Some DNA codes for RNA transcripts that perform a variety of functions within the nucleus, but no function is known for the majority of the genomic DNA.
- The structure of DNA can be envisioned as a twisted ladder, with the sugar–phosphate bonds as the sides of the ladder and the four nucleotide bases (adenosine [A], cytosine [C], guanine [G], and thymine [T]) forming the rungs. The nucleotides form complementary base pairs: C with G and A with T.
- The DNA double helix must separate into single strands to provide a template for synthesizing new identical DNA strands that can be passed on to daughter cells during cell division. DNA replication is accomplished by the enzyme complex DNA polymerase. DNA synthesis has extremely high fidelity.
- A linear sequence of DNA that codes for a protein is called a *protein-encoding gene.* During transcription, genes provide a template for the synthesis of mRNA by the enzyme complex RNA polymerase.
- After a complex process of cutting and splicing of the pre-mRNA transcript within the nucleus, the mRNA is transported to the cytoplasm and translated into a protein. Each nucleotide triplet (codon) in the mRNA codes for a particular amino acid. Protein synthesis is accomplished by ribosomes, which match the mRNA codon with the correct tRNA anticodon and then catalyze the peptide bond to link amino acids together into a linear protein.

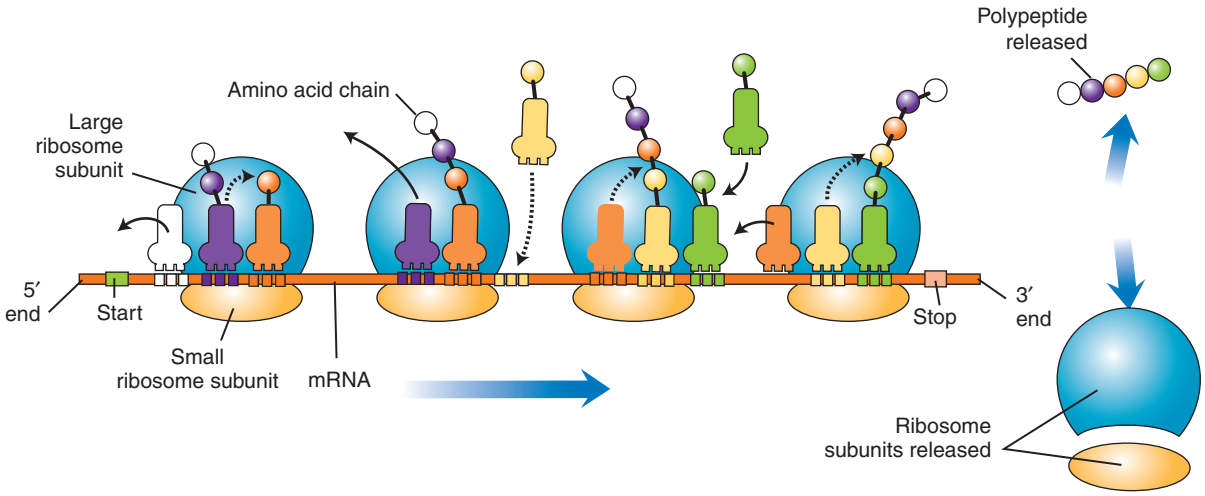


FIG 5.9 Synthesis of a protein by the ribosomes attached to an mRNA molecule. Ribosomes attach near the start codon and catalyze the formation of the peptide chain. The mRNA strand is read in groups of three nucleotides (codons) until the stop codon is reached and the peptide is released. Several ribosomes may translate a single mRNA into multiple copies of the protein.

REGULATION OF THE GENOME

The genome contains the genetic information of the cell and ultimately determines its form and function. All the various cells in a multicellular organism contain the same genes, and differences in cell type are thought to be the result of differences in DNA *expression*. To maintain the cell's phenotype, some genes must be actively transcribed, whereas others remain quiescent. In addition, the cell must be able to change the expression of certain genes to respond and adapt to changes in the cellular environment. At any one time, a cell expresses 30% to 60% of its approximately 28,000 genes (19,000 protein-coding genes and 9000 RNA genes). There is evidence that gene expression can be regulated at each of the steps in the pathway from DNA to RNA to protein synthesis. The proteins made by a cell can be controlled in the following ways: (1) regulating the rate and timing of gene transcription; (2) controlling the way the mRNA is spliced; (3) selecting the mRNAs that are transported to the cytoplasm; (4) selecting the mRNAs that are translated by ribosomes; (5) selectively destroying certain mRNAs in the cytoplasm; or (6) selectively controlling the activity of the proteins after they have been produced.

For a majority of genes, the most important regulators of expression are the transcriptional controls. Approximately 10% of the cell's protein-coding genes are devoted to making transcription-regulating proteins that are able to enhance or inhibit gene expression. These transcription regulators recognize and bind only particular DNA sequences and thus are specific to the genes they regulate. The approximately 2000 different genes that code for gene regulatory proteins work in combination to control numerous genes. The ability to regulate gene expression allows the cell to alter its structure and function in response to signals from its environment.

Transcriptional Controls

The gene regulatory proteins described in the preceding paragraphs are thought to control gene transcription by binding in the major groove of the DNA double helix near the promoter sequence of the gene to be regulated. Binding of the regulatory proteins may either enhance or inhibit RNA polymerase binding and subsequent transcription of the gene. This is sometimes referred to as "turning on" or "turning off" a gene. The DNA-binding proteins are able to recognize their specific binding sites because of small variations in structure of the external surface of the DNA double helix and do not require separation of the strands to bind. These regulatory DNA-binding proteins can be categorized either as positive controls that activate transcription *(activators)* or as negative controls that inhibit transcription *(repressors)*.

In humans, the strategies for gene regulation are complex. Gene regulatory proteins often bind DNA segments far from the gene being regulated, and binding of several gene regulatory proteins in combination is often necessary. A critical step in initiating gene transcription in human cells is the assembly of general transcription factors at the promoter region. General transcription factors are a group of DNA-binding proteins necessary for RNA polymerase activity, and initiation of transcription does not occur without them. Regulatory gene activator proteins help to collect the transcription factors at the promoter of the correct gene by first recognizing and binding to a specific DNA sequence and then coordinating the assembly of the transcription factors (Fig. 5.11).

Inhibition of transcription is achieved by gene repressor proteins, which also recognize and bind specific DNA sequences but inhibit the assembly of transcription factors at the site. Repressor proteins exert their effects through complex mechanisms, such as compacting the DNA to make it difficult to pry open, interfering with activator proteins, and binding up or inhibiting transcription factors. Inappropriate transcription of genes in a particular cell may have dire consequences for the cell or for the organism as a whole and is therefore a carefully regulated process. The presence, position, and activity of gene regulatory proteins may be regulated by various signaling cascades within the cell. Many of these signaling cascades are triggered by changes in the cell's environment, which then alter gene transcription (see Chapter 3). This process is highly complex, with numerous signaling pathways often converging on a particular gene regulatory system. Even after the mRNA transcript is produced, it may not be allowed to reach the ribosome for translation. Small RNA molecules called *micro RNA (miRNA)* and *small interfering RNA (siRNA)* can anneal to complementary segments of the mRNA within the nucleus. In some cases, these small RNAs regulate gene splicing, but in other cases they "silence" the gene by preventing the mRNA from being translated into a protein.

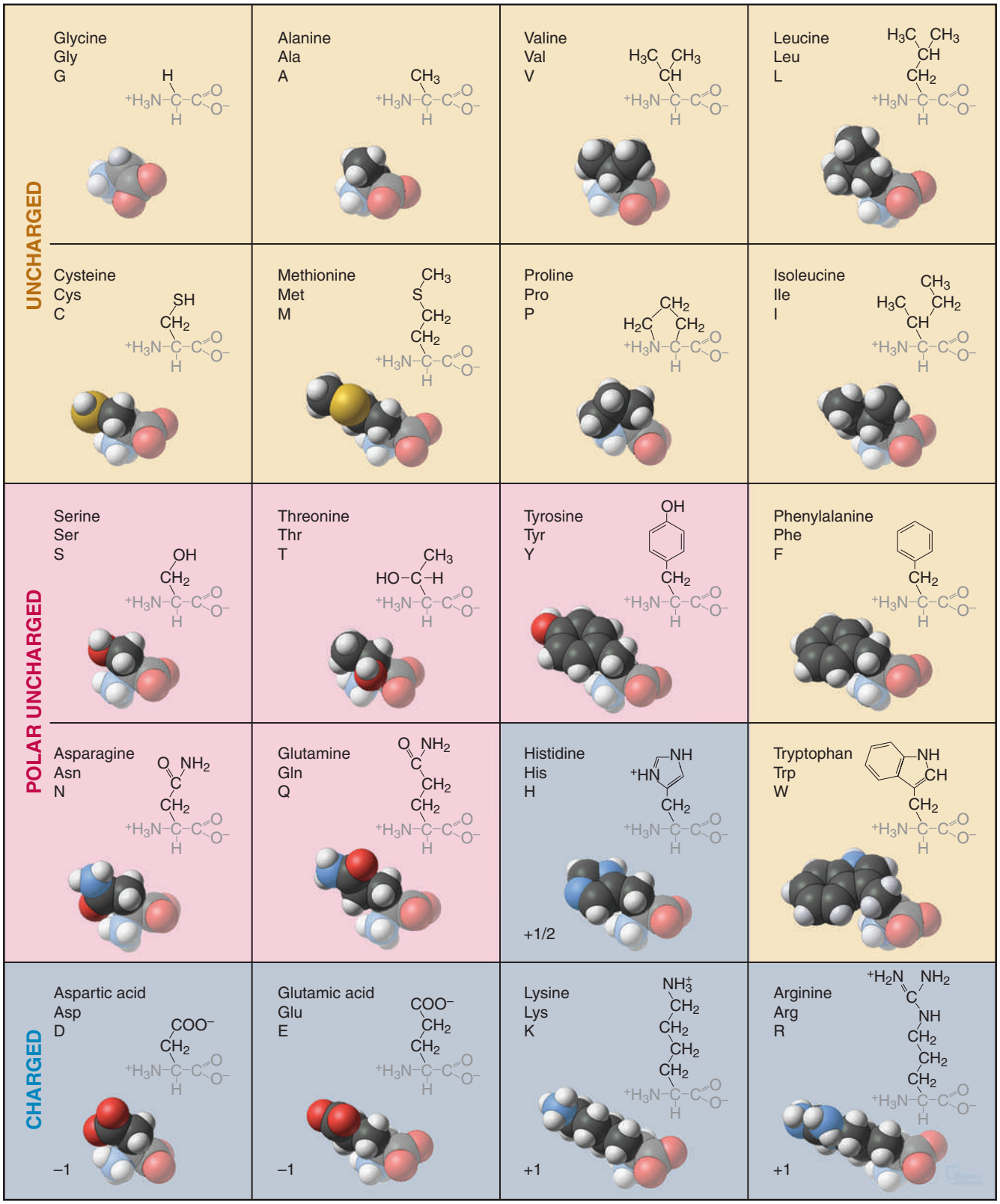


FIG 5.10 The 20 amino acids that form proteins have different chemical structures that affect their solubility in lipids and water. Nonpolar amino acids tend to locate in the lipid bilayer or in the interior of globular proteins, whereas polar and charged amino acids interact well with water. (From Pollard T, Earnshaw W: *Cell biology,* 2007, Philadelphia, Saunders.)

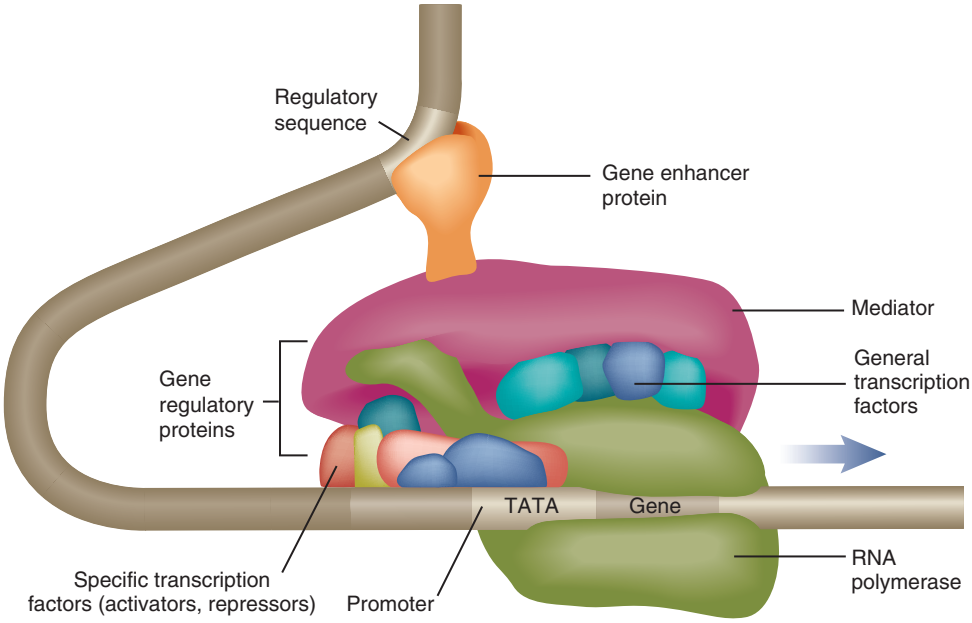


FIG 5.11 Gene activator proteins coordinate the assembly of general transcription factors at the promoter region of the gene to be transcribed. RNA polymerase is unable to bind and begin transcription until the requisite transcription factors are in place. Mediator and general transcription factors are the same for all RNA polymerase transcribed genes, but regulators are specific.

KEY POINTS

- All the cells in an individual have essentially the same DNA; however, cells differ greatly in structure and function. This occurs because genes are selectively expressed in particular cells.
- Gene expression can be regulated at any step in the pathway from DNA to RNA to protein synthesis. The most important regulators are transcriptional controls.
- A critical step for initiation of gene transcription is the assembly of general transcription factors at the promoter region of the gene.
- The actions of general transcription factors and RNA polymerase are controlled by a large number of regulatory proteins that specifically bind to DNA. The presence of certain DNA-binding proteins at specific sites can activate or repress the transcription of a particular gene in response to signals in the cell's environment.
- A number of small RNA molecules function to regulate mRNA transcription and processing in the nucleus.

DIFFERENTIATION OF TISSUES

Cell Diversification and Cell Memory

The cells of a multicellular organism tend to specialize to perform particular functions in coordination with other cells and tissues of the body. Cells not only must become different during development, but also must remain different in the adult, after the original cues for cell diversification have disappeared. The differences among cell types are ultimately the result of the differentiating influences experienced in the embryo. Differences are maintained because the cells retain the effects of those past influences and pass the memory on to their descendants. When a skin cell divides to replace lost skin cells, the daughter cells are also skin cells; when a liver cell divides, its daughter cells are liver cells; and so on. The behavior of cells of higher organisms is governed by their genome and their present environment, as well as by their developmental history.

There is substantial evidence that the differences in tissue structure and function in a particular organism are not due to deletions or additions to the genes. All the cells of an organism contain essentially the same genes. It is the expression of a relatively few tissue-specific genes that results in differences among cell types. The exact mechanisms leading to the stable expression of tissue-specific genes in particular cell types are partly unknown; however, differences in DNA packaging and the combination of gene regulatory proteins passed on during cell division are thought to be important. The DNA in human cells is extensively packaged, so that 40 inches of linear DNA can be compacted to fit into the cell nucleus. However, different regions of chromosomes can be more or less condensed (see Fig. 5.4).

Some regions of DNA, called heterochromatin, are so condensed that they are not open to transcription. It is thought that the pattern of packaging as well as the DNA-binding proteins that regulate it are transmitted to progeny when a cell divides such that the pattern of gene expression is maintained as the cell's developmental memory. An example of this mechanism is the inactivation of one of the X chromosomes in females. In mammals, all female cells contain two X chromosomes (XX), whereas male cells contain an X and a Y chromosome (XY). One of the X chromosomes in females is permanently inactivated early in development by condensed packaging. This apparently occurs to prevent a double dose of the X gene products. Which of the two X chromosomes is inactivated in a particular cell is a random event. However, the same X chromosome will be inactive in all of the cell's progeny. Methylation of cytosine nucleotides (Fig. 5.12) in specific areas of DNA with CG sequences also makes the area resistant to transcription. The pattern of methylation-inactivated genes in a particular cell type is "remembered" in subsequent generations of cells and may explain, in part, how differentiated tissues remain differentiated in the adult.

Mechanisms of Development

Embryonic development is associated with selective gene expression that controls four essential processes to enable a single cell to develop into a complex organism: (1) cell proliferation, (2) cell specialization,

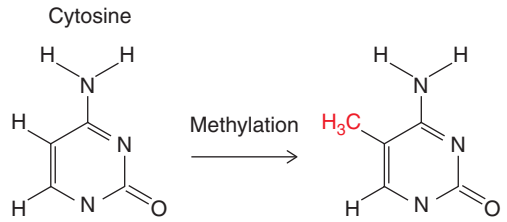


FIG 5.12 Methylation of cytosine is a common way that gene transcription is repressed.

(3) cell-to-cell interactions, and (4) cell movement and migration. Each time a cell divides, it must retain memory of the developmental events that have preceded the division so that it can progress along a developmental pathway toward becoming a differentiated tissue. Cells have a genetic memory: the genes a cell expresses and the way it behaves depend on the cell's past, as well as its present environment. There is no overall controlling center; each cell must make its own developmental decisions.

Two major classes of proteins are particularly important for multicellular development: (1) transmembrane proteins in the cell surface that participate in cell adhesion and communication and (2) DNA-binding proteins that regulate gene transcription. Differences between cells in an embryo are a necessary prelude to development of a multicellular organism and arise in various ways. Very early in embryonic development, cells begin to divide asymmetrically so that daughter cells are not identical—those on the outside of the group of cells receive different environmental cues than those on the inside, which are surrounded by other cells. These simple differences in cell-to-cell adhesion may alter the transcription of a set of genes. The altered genetic expression will then be passed on to daughter cells in the next cell division, making them diverge further from the original cell. Subsequently, the cell will respond differently to environmental influences, which further alter the cell's structure and function. Thus cells become committed to a developmental pathway over the course of many cell divisions that transmit the history of previous exposures through sequential changes in gene expression.

Continued interactions with nearby cells, chemical gradients, and extracellular matrix components provide clues to guide the cell to its appropriate form and location in the developing organism. Chemicals that control the patterning of fields of nearby tissue are termed morphogens**.** For example, cells in the head region may specialize to secrete a "position signal" for other cells. The morphogen is progressively degraded as it diffuses through the neighboring tissue, such that it has higher concentration close to the source. A particular cell will have information regarding its proximity to the head region based on the surrounding concentration of the chemical. Morphogens are thought to be effective only over small distances. Thus the gross distinctions between head and tail, for example, must be made very early in the embryo, and morphogens can provide only a general pattern for future development. Successive levels of detail can be provided later by other positional signals.

The organization of molecules surrounding the cell surface also provides positional information. The extracellular matrix is composed of a large meshwork of molecules that is produced locally by cells in the area. Some common components include the proteins collagen and elastin; long polysaccharide chains called *glycosaminoglycans;* and a variety of peptides, growth factors, and hormones. The extracellular matrix is highly organized, with components binding to each other and to the cell membrane in specific ways. The extracellular matrix is thought to be important in cell development through its ability to screen or modulate the transport of molecules, such as growth factors, to the cell membrane and through direct contacts with the cell membrane that effect changes in cell structure and function.

The extracellular matrix surrounding the cells in different locations provides positional information to cells that must migrate to their final destination. In vertebrates, connective tissue cells appear to provide much of this positional information. While the migratory cell travels through the connective tissue, it continually samples the surroundings, searching for cues to guide it. Certain chemokines (CXCL12) and complementary receptors (CXCR4) on migratory cells are an important signaling system to help cells localize to certain areas. Migratory cells with other specific cell surface receptors may interact differentially with the extracellular matrix in different areas. In this way the migratory cell can be guided along particular paths and induced to settle in particular areas. Once the migratory cell has settled, local extracellular matrix molecules may further affect the cell's growth rate, differentiation, and likelihood of survival.

Interactions between the extracellular matrix and nearby cells are mediated primarily by binding proteins called integrins**.** Integrins are transmembrane proteins that tie the cell's cytoskeleton to particular matrix structures (see Chapter 3). They enable the cytoskeleton and extracellular matrix to communicate across the plasma membrane in specific ways. In addition to inducing cells to bind in a particular location, integrins have been shown to activate intracellular signaling pathways, which may influence cell behavior in numerous ways (e.g., cell shape, polarity, metabolism, development, and differentiation).

The steps leading to the development of differentiated tissues in a multicellular organism are such that, once differentiated, a cell type generally does not revert to earlier forms. Some cells in a tissue are terminally differentiated and have limited capacity to change form or replicate. Tissues also maintain less differentiated stem cells that are able to proliferate depending on environmental cues. Some stem cells, located mainly in the bone marrow, are quite similar to embryonic stem cells and can be recruited into tissues where they proliferate and differentiate into tissue cells. The ability of these multipotent or pluripotent stem cells to survive and differentiate correctly in their adopted home depends on making complex cell-to-cell and cell-to-matrix connections. In the absence of an appropriate environment, the cells will undergo apoptosis and die (see Chapter 4).

Differentiated Tissues

The more than 200 different cell types in the adult human are generally classified into four major tissue categories: epithelium, connective tissue, muscle, and nerve. Tissue types and some of their subtypes are summarized in Table 5.2. Most of the organ systems of the body are combinations of these four tissue types mixed in a highly organized and cooperative manner.

Epithelial Tissue

Epithelial cells cover the majority of the external surfaces of the body and line the glands, blood vessels, and internal surfaces. Epithelial cells adopt a variety of shapes and functions, depending on their locations. For example, the *stratified* epithelium that composes the epidermis of the skin is several layers thick and is primarily protective in function. New epithelial skin cells are formed from stem cells in the deepest part of the epidermis, where it contacts the basal lamina. As cells mature, they move outward toward the surface until they become keratinized and finally flake away (Fig. 5.13). Keratin is a tough protective protein that is present in large quantities in the outer skin layers of flattened, dead epithelial cells. The epidermis in humans is completely replaced about once per month, but turnover can occur more rapidly after injury to the skin.

TABLE 5.2 Major Categories and Location of Body Tissues

Tissue Type	Locations
Epithelial Tissue	
Simple squamous	Lining of blood vessels, pulmonary alveoli, Bowman capsule
Simple cuboidal	Thyroid, sweat, and salivary glands; kidney tubules
Simple columnar	Lining of intestine, glandular ducts
Pseudostratified (mixed cell shapes)	Male urethra, respiratory tract passages
Stratified squamous	Skin, mucous membranes
Stratified columnar	Epiglottis, anus, parts of pharynx
Stratified transitional (layers of different cell shapes)	Bladder
Connective Tissue	
Loose	Widespread locations, dermis of skin, adipose tissue, organs
Dense/supportive	Cartilage, bone, tendons, joints, fascia surrounding muscles
Hematopoietic	Bone marrow, lymph tissue, plasma
Muscle Tissue	
Skeletal	Voluntary muscles of body
Cardiac	Heart (myocardium)
Smooth	Intestine, blood vessels, bladder, uterus, airways
Myoepithelial	Mammary, sweat, and salivary glands
Nervous Tissue	
Neurons	Central and peripheral nerves
Neuroglia	Primarily central nervous system

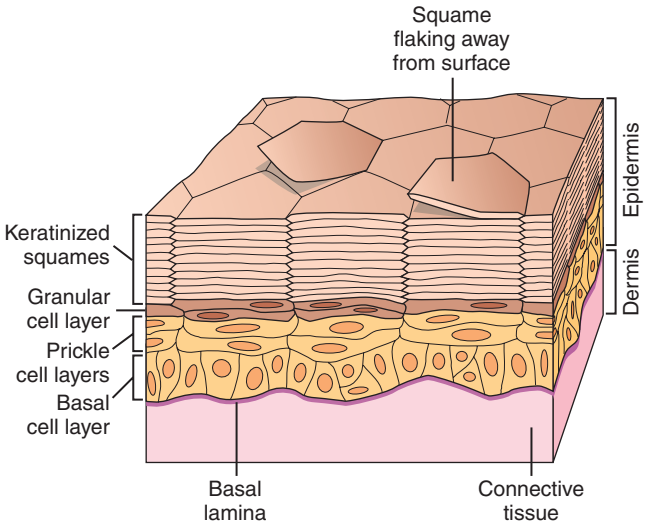


FIG 5.13 Organization of epidermal skin layers, showing the flattened keratinized outer layer. Epithelial cells are continually produced by stem cells at the basal lamina and then migrate to the surface.

In addition to stratified epithelium, the epithelium may be characterized as simple or pseudostratified according to the number and arrangement of cell layers (Fig. 5.14). *Simple* epithelium consists of a single layer of cells, all of which contact the basement membrane. Simple epithelium is found in the lining of blood vessels and body cavities, in many glands, and in the alveoli of the lungs. The simple epithelium that lines the blood vessels is called *endothelium.* Simple epithelium also forms the kidney tubules and lines the intestine, where absorption is its primary function. *Stratified* epithelium consists of two or more layers of epithelial cells and is found in mucous membranes, such as the mouth, and in the skin, as mentioned previously. Epithelium that appears to be more than one layer thick because of a mixture of cell shapes but is actually a single layer is called *pseudostratified* epithelium. The linings of the respiratory tract and some glands contain pseudostratified epithelium.

Epithelial cells may also be classified according to cell shape. The three basic cell shapes are squamous, cuboidal, and columnar. Squamous

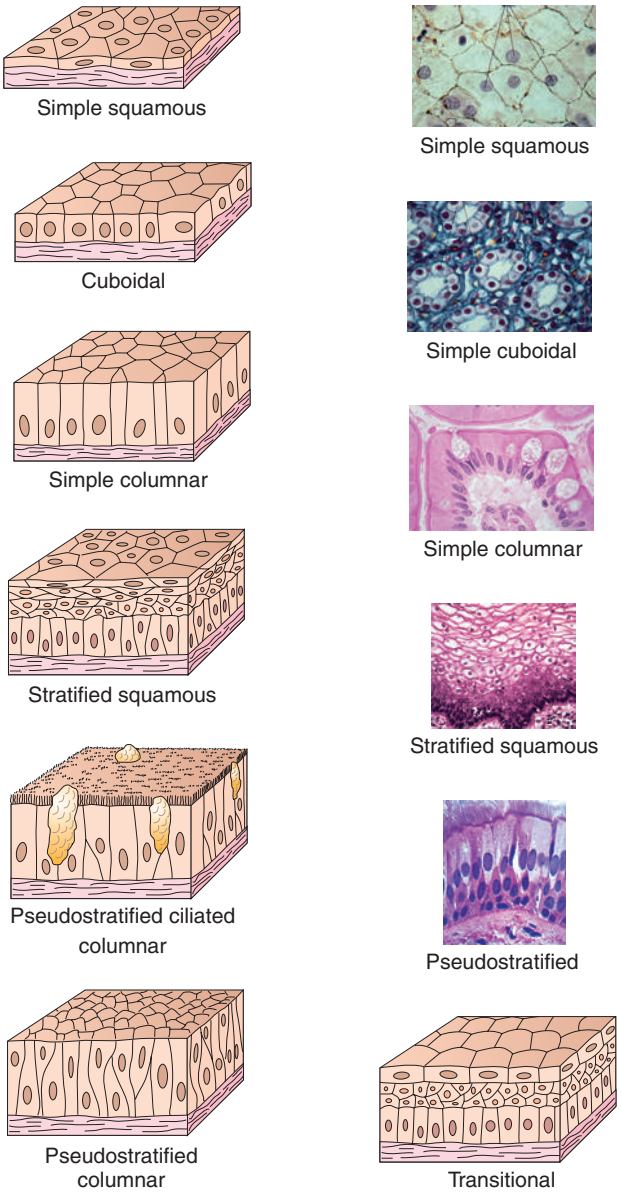


FIG 5.14 Various epithelial tissue shapes and layering.

cells are thin in comparison to their surface area and have a flattened appearance. Cuboidal cells are approximately equal in width and height, similar to a cube. Columnar cells are a bit taller than they are wide, resembling a rectangular column. Several classifications of epithelial tissue are given in Table 5.2, using both shape and layering as criteria.

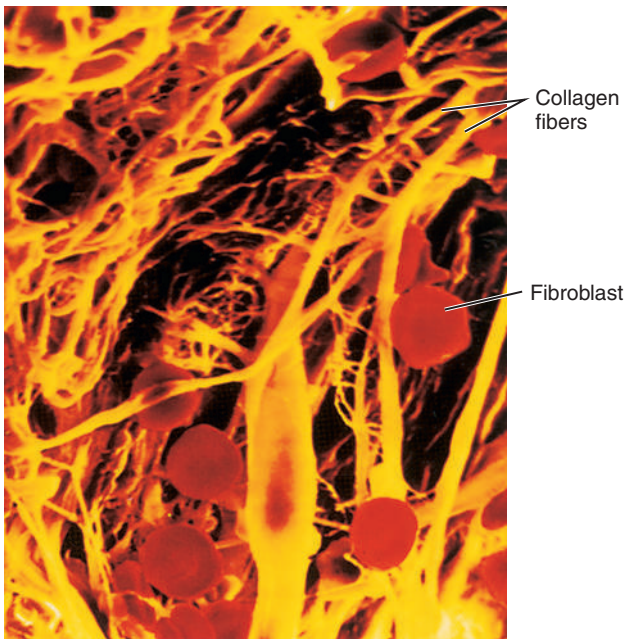
Connective Tissue

Connective tissue is the most abundant and diverse tissue in the body, including cell types as different as bone cells, fat cells, and blood cells. Connective tissue commonly functions as a scaffold on which other cells cluster to form organs, but it does much more than hold tissues together. Connective tissue cells often form an elaborate extracellular matrix, which is thought to be important in the maintenance of cell differentiation. Connective tissue cells play an important part in the support and repair of nearly every tissue and organ in the body. Three major classifications of connective tissue are commonly identified: loose connective tissue, dense or supportive tissue, and hematopoietic tissue.

Loose connective tissue appears unstructured, with a fair amount of space between fibers of the extracellular matrix. The matrix contains a number of cell types and an elaborate meshwork of protein and other molecules (Fig. 5.15). The primary protein constituents are collagen, elastin, and reticular fibers. Collagen is composed of tough, nonelastic bundles of protein fibers that are secreted by fibroblasts. It gives structural strength to skin, tendons, ligaments, and other tissues. The ability of a structure to withstand deforming and stretching forces is due, in large part, to elastin, which can return to its original length after being stretched, like a rubber band. Elastin is important to the function of structures such as the aorta, which must expand to accept the blood ejected from the heart during systole and bounce back to its original shape during diastole. Reticular fibers are short branching fibers that provide networks for the attachment of connective tissue to other cell types, such as epithelial cell attachments in glands, hematopoietic cells in bone marrow, and the parenchymal cells (functional cells) in organs. Cell types associated with loose connective tissue include the fibroblasts, mast cells, and adipocytes (fat cells).

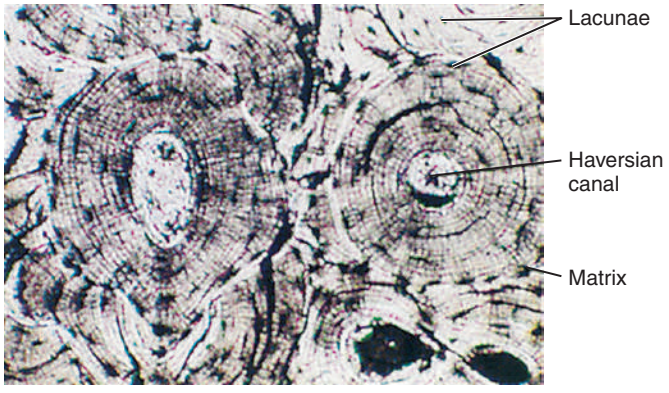
Dense or supportive connective tissue is rich in collagen, which gives strength to structures such as cartilage, tendon, bone, and ligaments. The collagen fibers are more organized and densely packed than fibers found in loose connective tissue. Cartilage cells, or chondrocytes, may be found in the trachea, joints, nose, ears, vertebral disks, organs, and the young skeleton. Once formed, the collagenous extracellular matrix structures require little maintenance and do not receive a blood supply. Bone is a highly dense form of connective tissue composed of a mixture of tough collagen fibers and solid calcium phosphate crystals in approximately equal proportions. Throughout the bone's hard extracellular matrix are channels and cavities occupied by living cells (osteocytes) (Fig. 5.16). These cells incessantly model and remodel their bony environment, responding to environmental signals. These osteocytes are of two kinds: the cells that erode old bone are called *osteoclasts,* whereas the cells that form new bone are called *osteoblasts.* Osteoblasts detect when a bone is subjected to a greater load stress and adapt by strengthening the bone mass. Conversely, when the load is removed, as during bed rest, the osteoclasts busily digest the bone, often resulting in some of the common complications of immobility. Osteoclasts, like macrophages, are derived from monocytes that are produced in the bone marrow. The monocytes travel via the bloodstream and collect at sites of bone resorption, where they fuse together to become osteoclasts. Osteocyte activity is essential for bone growth and the repair of bone injuries. (See Chapter 50 for a detailed description of the musculoskeletal system.)

The blood-forming organs of the body contain a specialized type of connective tissue called *hematopoietic* tissue. The blood cells include the red cell, or erythrocyte**,** which is specialized for the transport of oxygen; the platelet, or thrombocyte**,** which is important in blood coagulation; and a host of white cells, or leukocytes**,** which mediate immune function. Blood-forming tissue is located in the bone marrow, spleen, and lymphatic tissue. Hematopoietic cells are necessarily nomadic, traveling to distant areas of the body and sometimes settling in a particular organ, sometimes moving continuously (Fig. 5.17). Blood cells have a short life span in comparison to other cells and must continually be replenished. This is accomplished by the hematopoietic stem cells. Stem cells reside primarily in the bone marrow and are multipotent;



Loose connective tissue

FIG 5.15 Scanning electron micrograph of fibroblasts in loose connective tissue of a rat cornea. The matrix is composed primarily of collagen fibers (magnification ×440). (From Solomon EP: *Introduction to human anatomy and physiology,* ed 4, Philadelphia, 2016, Saunders, p 38.)



Bone

FIG 5.16 Photomicrograph of a section of compact bone showing circular networks formed by the action of osteoclasts and osteoblasts as they remodel the bone. The osteocytes occupy the lacunae and canals. (From Solomon EP: *Introduction to human anatomy and physiology,* ed 4, Philadelphia, 2016, Saunders, p 38.)

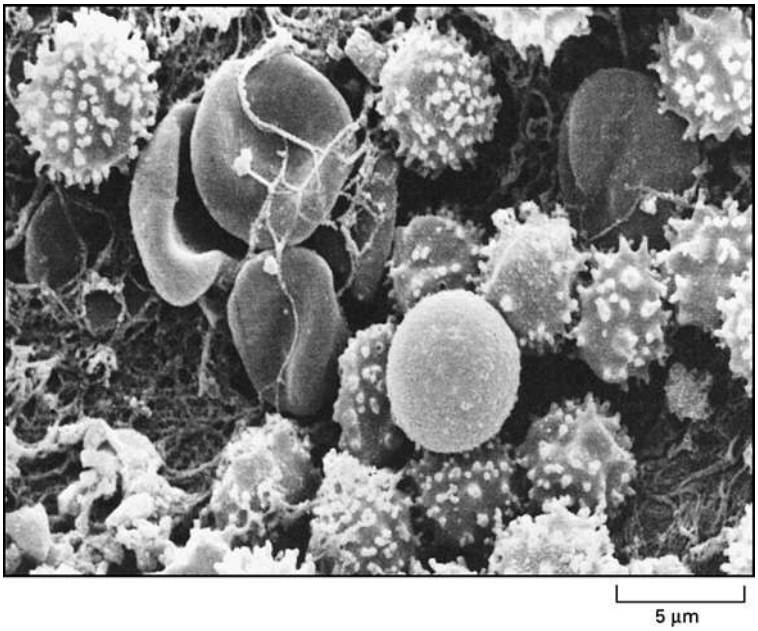


FIG 5.17 Scanning electron micrograph of red and white blood cells in the lumen of a blood vessel. Red blood cells are smooth and concave, whereas white blood cells are rough and rounded. (From Alberts B et al, editors: *Molecular biology of the cell,* ed 5, New York, 2008, Garland Science, p 1451. Courtesy Ray Moss.)

they may differentiate into any of the blood cell types. This results in a system that can respond quickly to the changing needs of the body.

Muscle Tissue

The term *muscle* refers to tissues that are specialized for contraction. Muscle cells, or myocytes, are usually long and thin and packed with the proteins actin and myosin, which constitute the contractile apparatus. In mammals, there are four main categories of muscle cells: skeletal, cardiac, smooth, and myoepithelial (Fig. 5.18). Contraction in all four types depends on the presence of intracellular free calcium and occurs because of interactions between actin and myosin filaments. Actin and myosin filaments differ among cell types with regard to amino acid sequence, arrangement within the cell, and the mechanisms that control contraction. The mechanism of muscle contraction has been called *the sliding filament hypothesis* or *cross-bridge theory.* These terms describe the interactions of the actin and myosin filaments while they form bonds and pull past each other, causing the muscle cell to shorten. Contraction is initiated by an increase in intracellular free calcium concentration and requires energy in the form of adenosine triphosphate. A detailed description of actin–myosin cross-bridging and the role of calcium, troponin, and tropomyosin can be found in Chapter 17.

Skeletal muscle is responsible for nearly all voluntary movements. Skeletal muscle cells fuse together to form long multinucleated fibers that can be huge, up to 0.5 meter (m) in length. Once fused and differentiated into mature skeletal muscle cells, they cannot enter the cell cycle and divide to produce new cells. Skeletal muscle stem cells (satellite cells) are retained in the muscle tissue and can proliferate in response to muscle damage. The actin and myosin proteins in skeletal muscle are aligned in orderly arrays, giving the tissue a striped appearance under the microscope, which in turn has led to the term *striated muscle.* Skeletal muscle contracts in response to stimulation from the motor neurons of the nervous system (see Chapter 50). As in other types of muscle, stimulation results in an increase in free calcium concentration within the cell. In skeletal muscle, the calcium originates from internal storage sites in the sarcoplasmic reticulum. Contraction is initiated when the calcium binds troponin, a regulatory protein attached to the actin filament. Because of the high energy requirements of contracting skeletal muscle, the cells are packed with energy-producing mitochondria.

Like skeletal muscle, *cardiac muscle* also has a striated appearance attributable to the systematic organization of its actin and myosin filaments. Cardiac muscle cells are linked by special structures, called *intercalated disks* and gap junctions, that cause the tissue to behave as a syncytium**:** all the cells contract synchronously. Cardiac muscle contracts in response to activation of pacemaker cells in the heart that have the special property of automaticity. *Automaticity* refers to the inherent ability of the cell to initiate an action potential without outside stimulation. The contractile mechanisms of cardiac muscle are similar to those of skeletal muscle, requiring free calcium to interact with troponin, resulting in the formation of actin–myosin cross-bridges. In cardiac muscle, some of the free calcium originates from the sarcoplasmic reticulum, but diffusion into the cell through channels in the cell membrane is also necessary. These membrane calcium channels represent an important difference from skeletal muscle, because they can be manipulated by drugs (calcium channel blockers) without disrupting skeletal muscle control. (Cardiac muscle is discussed in Chapter 17.)

Smooth muscle comprises a diverse group of tissues located in organs throughout the body. Smooth muscle generally is not under voluntary control and therefore is called *involuntary muscle.* Some types of smooth muscle are able to contract intrinsically, and most are influenced by the autonomic nervous system. Smooth muscle is found in blood vessels and in the walls of hollow organs, such as those of the gastrointestinal tract, uterus, and large airways.

The structure of smooth muscle differs considerably from that of skeletal and cardiac muscle, and therefore some classification schemes consider it to be a member of the connective tissue family. The actin and myosin filaments are less organized in smooth muscle, and the

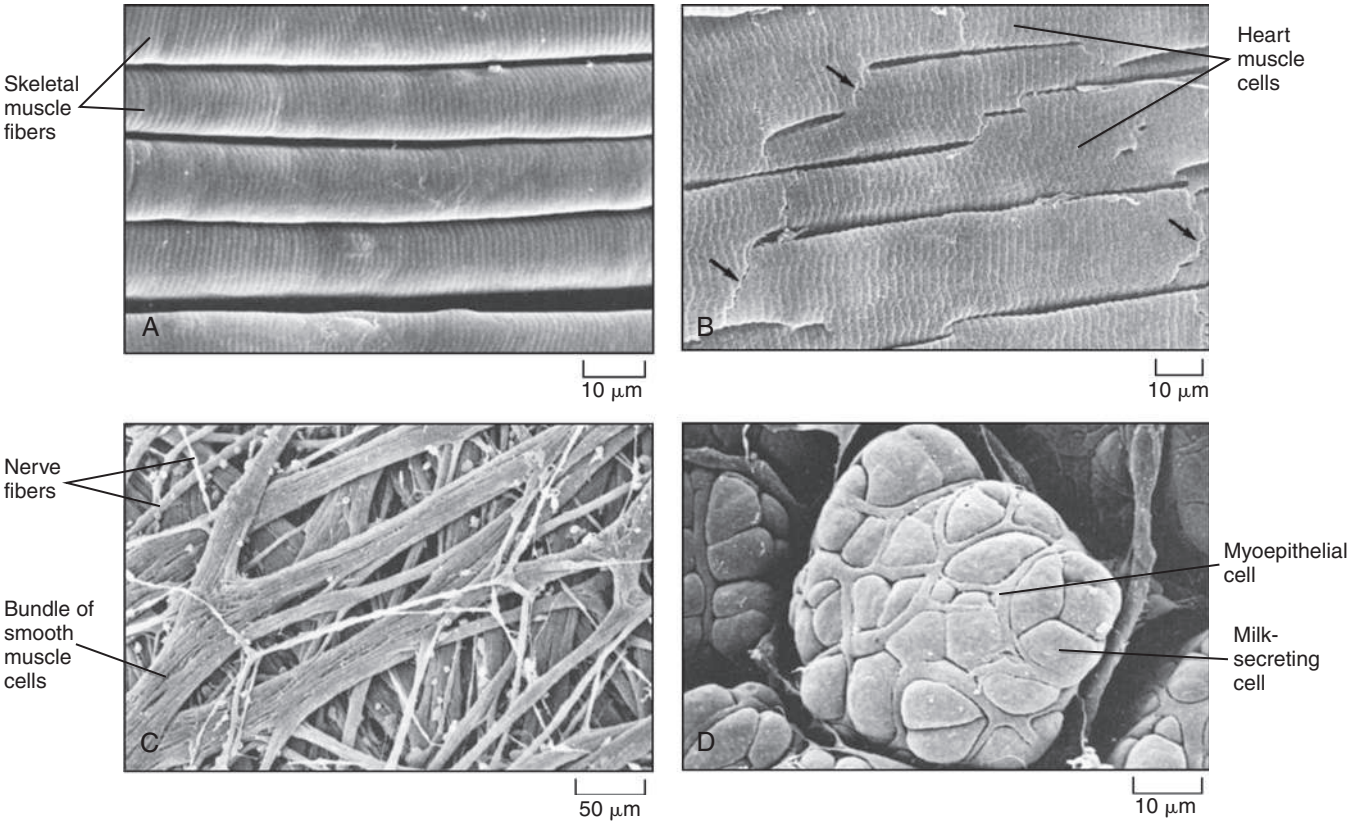


FIG 5.18 The four classes of muscle cells. **A,** Skeletal muscle. **B,** Heart (cardiac) muscle. **C,** Smooth muscle (bladder). **D,** Myoepithelial cells in a mammary gland. (A and C, From Alberts B et al, editors: *Molecular biology of the cell,* ed 6, New York, 2015, Garland Science, p 1233. A, Courtesy Junzo Deskati. B, From Fujiwara T: Cardiac muscle. In Canal ED, editor: *Handbook of microscopic anatomy,* Berlin, 1986, Springer-Verlag. C, Courtesy Satoshi Nakasiro. D, From Nagato T et al: A scanning electron microscope study of myoepithelial cells in exocrine glands, *Cell Tissue Res* 209:1–10, 1980.)

muscle does not have striations. Smooth muscle contraction tends to be slower and can be maintained indefinitely. This is critical to the function of blood vessels, which must maintain a degree of contraction or vascular tone to maintain the blood pressure. Smooth muscle has no troponin and uses the protein calmodulin as the calcium-binding regulatory protein. When calmodulin binds calcium ions in the cytoplasm, it activates the enzyme myosin light chain kinase (MLCK), which phosphorylates myosin and stimulates the rate of cross-bridge formation. Actin filaments are attached to structural proteins called *dense bodies* that pull in the sides of the muscle cell when actin–myosin cross-bridging causes the filaments to increasingly overlap (Fig. 5.19). Smooth muscle contraction is highly dependent on the diffusion of extracellular calcium into the cell through calcium channels in the plasma membrane (sarcolemma). Thus like cardiac muscle, smooth muscle can also be affected by drugs that alter the calcium channel's ability to conduct calcium. For example, calcium channel–blocking drugs are used to cause the smooth muscle in arterial blood vessels to relax as a treatment for high blood pressure.

Myoepithelial cells represent the fourth class of muscle cells. They are located in the ducts of some glands (e.g., mammary, sweat, and salivary). Unlike all other types of muscle, myoepithelial cells lie in the epithelium and are derived from embryonic ectoderm, whereas skeletal, cardiac, and smooth muscle are derived from embryonic mesoderm. Myoepithelial cells contract in response to specific stimuli (e.g., oxytocin in the mammary gland) and serve to expel the contents from the gland.

Nervous Tissue

Nervous tissue is widely distributed throughout the body, providing a rapid communication network between the central nervous system and various body parts. Nerve cells are specialized to generate and transmit electrical impulses rapidly. Like muscle, nerves are excitable; they respond to stimulation by altering their electrical potentials. This excitability is caused by the presence of voltage-sensitive ion channels located in the plasma membrane of the nerve cell. Movement of ions through these channels results in the production and propagation of action potentials along the length of the neuron. Neurons communicate their action potentials to other nerve and muscle cells through synapses. At the synapse, the presynaptic neuron releases a chemical neurotransmitter into the space between itself and the next neuron (synaptic cleft), where it diffuses across and interacts with receptors on the postsynaptic neuron.

A typical neuron is composed of three parts: a cell body, an axon, and one or more dendrites (Fig. 5.20). The cell body contains the nucleus and other cytoplasmic organelles. The axon is generally long (as long as 1 m) and may be encased in a myelin sheath. The axons usually conduct impulses away from the cell body, whereas the dendritic processes usually receive information and conduct impulses toward the cell body. Neurons are classified on the basis of the number of projections extending from the cell body. Neurons are terminally differentiated and incapable of replicating. However, neural stem cells are located in certain areas of the brain like the hippocampus and may replicate to form either neurons or glial cells in response to specific signals (see Chapter 43).

In addition to neurons, nervous tissue contains a variety of supportive cells, termed neuroglia ("nerve glue"), that nourish, protect, insulate, and clean up debris in the central nervous system. These include the astrocytes, oligodendroglia, ependymal cells, and microglia. (See Chapter 43 for a detailed description of nervous system anatomy and physiology.)

KEY POINTS

- The structure and function of cells are influenced by the genome and environment as well as by developmental history.
- Embryonic development is associated with selective gene expression that controls four essential processes to enable a single cell to develop into a complex organism: (1) cell proliferation; (2) cell specialization; (3) cell-to-cell interactions; and (4) cell movement and migration.
- Many terminally differentiated cell types are unable to divide. Some tissues, such as skin and bone marrow, maintain large numbers of stem cells, which have great capacity to proliferate. Tissues are able to recruit multipotent stem cells from the bone marrow that have the capacity to become differentiated tissue cells if given appropriate survival and developmental conditions. Some differentiated cells, such as liver cells, maintain the ability to divide even though they are differentiated.
- Different cell types in the adult human are classified into four major categories: epithelium (e.g., skin, glands, endothelium); connective tissue (e.g., bone, cartilage, fat, blood); muscle (e.g., skeletal, cardiac, smooth); and nervous tissue (e.g., neuronal, glial).

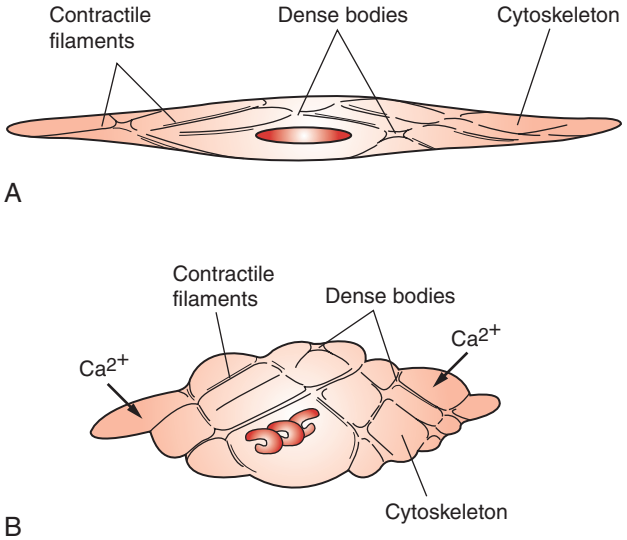


FIG 5.19 Schematic drawing of a smooth muscle cell when relaxed **(A)** and contracted **(B).** Contraction begins with the entry of Ca^{2+} into the cell through L-type voltage-gated calcium channels. Ca^{2+} is also released from the sarcoplasmic reticulum. The calcium ions bind to cytoplasmic calmodulin to form a complex that activates myosin light chain kinase (MLCK). The kinase attaches a phosphate to the myosin head area, which stimulates its cycling activity. The myosin binds to actin filaments and tugs on them with each cross-bridge cycle. While the myosin and actin filaments pull closer together and overlap more, the muscle cell shortens. The actin filaments are attached to dense bodies that are analogous in function to the Z-disk protein in cardiac and skeletal muscle. Smooth muscle can maintain long-term actin–myosin cross-bridges that maintain a level of tone.

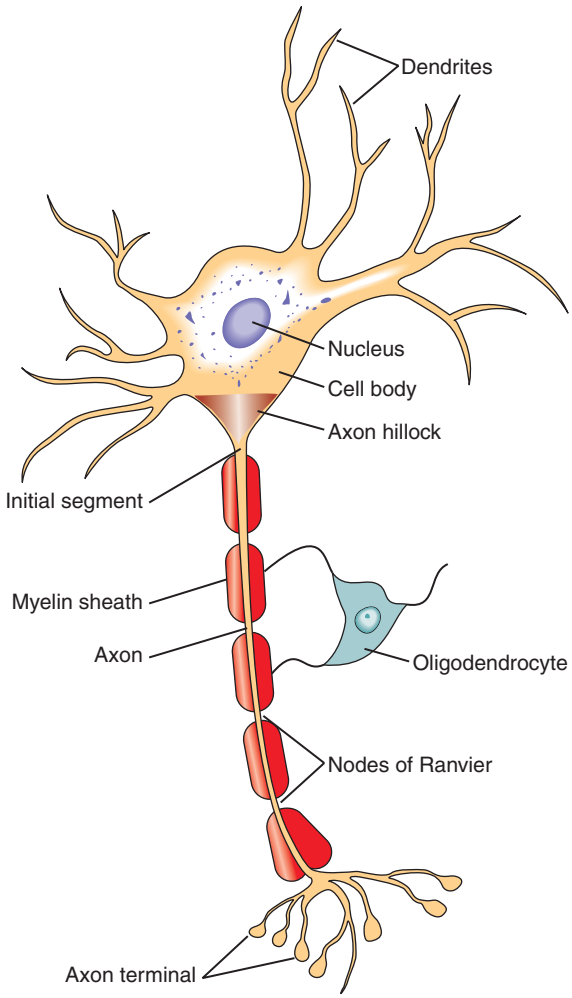


FIG 5.20 Diagram of a typical neuron showing the cell body, axon, and dendrites. Neurons have many shapes and sizes.

SUMMARY

The development, differentiation, and daily activities of a cell are directed by its genome. Genes are sequences of nucleotides that provide the template for the production of RNA or cellular proteins. In large part, the kinds and amounts of cellular proteins determine cell structure and function. All the cells of the body possess essentially the same DNA, but through complex processes of differentiation they become specialized to perform particular functions. Different sets of genes are active in different cell types. The four major classes of differentiated tissues are epithelial, connective, muscle, and nerve. These four tissues interdependently form the functioning systems of the body.

RESOURCES

Genome Structure and Regulation

Alberts B, et al: Control of gene expression. In Alberts B, et al, editors: *Molecular biology of the cell*, ed 6, New York, 2015, Garland Science, pp 369–438.

Alberts B, et al: DNA chromosomes, and genomes. In Alberts B, et al, editors: *Molecular biology of the cell*, ed 6, New York, 2015, Garland Science, pp 173–236.

Alberts B, et al: DNA replication, repair and recombination. In Alberts B, et al, editors: *Molecular biology of the cell*, ed 6, New York, 2015, Garland Science, pp 237–298.

Alberts B, et al: How cells read the genome: From DNA to protein. In Alberts B, et al, editors: *Molecular biology of the cell*, ed 6, New York, 2015, Garland Science, pp 299–368.

Crick FHC: The genetic code: III. *Sci Am* 215(4):55–62, 1966.

Frisch L: *The genetic code. Cold Spring Harbor symposia on quantitative biology*, Cold Spring Harbor, NY, 1966, Cold Spring Harbor Laboratory.

Grunberg S, Hahn S: Structural insights into transcription initiation by RNA Polymerase II. *Trends Biochem Sci* 38:603–611, 2013.

Meisenberg G, Simmons WH: The human genome. In Meisenberg G, Simmons WH, editors: *Principles of medical biochemistry*, ed 3, Philadelphia, 2012, Saunders, pp 93–117.

Steitz TA: A structural understanding of the dynamic ribosome machine. *Nature* 9:242–253, 2008.

Thomas MC, Chiang CM: The general transcription machinery and general transcription factors. *Crit Rev Biochem Mol Biol* 41:105–178, 2006.

Watson JD, Crick FHC: Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* 171:737–738, 1953.

Differentiation of Tissues

Alberts B, et al: Development of multicellular organisms. In Alberts B, et al, editors: *Molecular biology of the cell*, ed 6, New York, 2015, Garland Science, pp 1145–1216.

Alberts B, et al: Stem cells and tissue renewal. In Alberts B, et al, editors: *Molecular biology of the cell*, ed 6, New York, 2015, Garland Science, pp 1217–1262.

Bell O, Tiwari VK, Thomä NH, Schübeler D: Determinants and dynamics of genome accessibility. *Nat Rev Genet* 12(8):554–564, 2011.

Gardiner NJ: Integrins and the extracellular matrix: key mediators of development and regeneration of the sensory nervous system. *Dev Neurobiol* 71(11):1054–1072, 2011.

Gurdon JB: The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. *J Embryol Exp Morphol* 10:622–640, 1962.

6

Genetic and Developmental Disorders

Linda D. Ward

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- How are genes transmitted from parent to offspring?
- How is pedigree analysis used to determine whether a trait is inherited in an autosomal-dominant, autosomal-recessive, or X-linked pattern?
- How might abnormal meiosis lead to alterations in chromosome number or structure?
- What are the inheritance patterns and general clinical features of some common genetic disorders?
- What is the role of the environment in the development of congenital disorders?
- What methods of genetic testing are available?

CHAPTER OUTLINE

Principles of Inheritance, 95
DNA Mutation and Repair, 96
GENETIC DISORDERS, 98
Chromosomal Abnormalities, 99
Aberrant Number of Chromosomes, 99
Abnormal Chromosome Structure, 100
Examples of Autosomal Chromosome Disorders, 100
Trisomy 21 (Down Syndrome), 100
Trisomy 18 (Edwards Syndrome) and Trisomy 13 (Patau Syndrome), 101
Cri du Chat Syndrome, 101
Examples of Sex Chromosome Disorders, 101
Klinefelter Syndrome, 101
Turner Syndrome, 102
Multiple X Females and Double Y Males, 102
Mendelian Single-Gene Disorders, 102
Autosomal-Dominant Disorders, 103
Marfan Syndrome, 103
Huntington Disease, 104
Autosomal-Recessive Disorders, 104
Albinism, 105
Phenylketonuria, 106
Cystic Fibrosis, 106
Sex-Linked (X-Linked) Disorders, 106
Hemophilia A, 107
Nonmendelian Single-Gene Disorders, 107
Anticipation, 109
Mitochondrial Gene Mutations, 109
Genomic Imprinting, 110
Polygenic and Multifactorial Disorders, 110
Environmentally Induced Congenital Disorders, 111
Periods of Fetal Vulnerability, 111
Teratogenic Agents, 111
Chemicals and Drugs, 111
Infectious Agents, 112
Radiation, 112
Other Disorders of Infancy, 113
Diagnosis, Counseling, and Gene Therapy, 113
Prenatal Diagnosis and Counseling, 113
Genetic Analysis and Therapy, 114
Recombinant DNA Technology, 114

Geneticists and parents alike have marveled at the development of a recognizable human baby, with eyes and ears, toes and fingers, from its simple beginning as a single cell containing one set of genes. Considering the enormous list of potentially disastrous genetic and environmental influences, the birth of a healthy, normal child does indeed seem like a miracle. Although the risk of bearing a child with mental or physical defects is small for most parents, it is real and is often a source of worry during the prenatal period. Disorders that are present at birth are called *congenital,* whether the cause is genetic, environmental, or both. Some congenital disorders are associated with structural defects attributable to errors in fetal development and are called *congenital malformations*. It is estimated that 2% to 3% of newborns have a major malformation of cosmetic or functional significance.

Malformations are frequently associated with genetic causes. Chromosomal abnormalities are thought to cause approximately 6% of congenital abnormalities, and single-gene disorders are implicated in

7% to 8% of cases. It has been estimated that most people harbor five to eight defective genes that are recessive and therefore of little consequence to their health. In addition, there are many known and unknown environmental hazards to which the parent and fetus may be exposed. Environmental agents that can adversely affect the developing embryo or fetus are called *teratogens* and include radiation, pathogens, chemicals, and certain drugs. In half of cases, a clear explanation for a malformation cannot be found, and most congenital malformations are thought to occur due to multifactorial (genetic + environmental) causes. Some inherited genetic disorders do not become apparent until later in childhood or adulthood and therefore are not considered congenital. In this chapter, the general principles of inheritance; genetic and environmental causes of congenital disorders; and the principles of diagnosis, counseling, and gene therapy are described.

PRINCIPLES OF INHERITANCE

"Whom does the baby look like?" is frequently asked of new parents. It is common knowledge that traits tend to run in families, but Gregor Mendel, a nineteenth-century monk turned geneticist, was the first to notice that certain traits in pea plants were transmitted in a predictable way from parent to offspring. The same is true for humans. Height, weight, skin color, eye color, and hair color are some of the physical traits that characterize an individual; other inherited traits contribute to risk for disease. *Phenotype* refers to the physical and biochemical traits or attributes of an individual that are outwardly apparent. A person's phenotype reflects his or her unique genetic makeup, or genotype, along with environmental factors that may affect gene expression. In some instances, a mutation in a single gene may be sufficient to cause a particular characteristic, including a disease, particularly when the gene encodes a critical protein. In the case of such single-gene (monogenic) diseases, Mendel's principles of inheritance are useful to predict the risk of passing the disease to offspring. More often, traits (including risk for disease) are determined by multiple genes and the effects of an individual's environment. Inheritance of these complex or multifactorial conditions is much more difficult to predict.

Human DNA is organized into 46 chromosomes which are diploid, which means they occur as 23 pairs. One member of each chromosome pair is inherited from the mother and the other from the father. Paired (homologous) chromosomes usually look identical under the microscope but differ in DNA sequence. Before cell division, chromosomes look like Xs of varying sizes and shapes. The X-shaped chromosome is really made up of two identical linear chromosome units, called *chromatids*, which separate during meiosis. The point at the middle of the X at which the two sister chromatids are united is the centromere (Fig. 6.1). Chromosomes are characterized on the basis of their total size, placement of the centromere, and characteristic banding patterns when exposed to certain stains (Fig. 6.2).

Of the 23 pairs of chromosomes, 22 are homologous and are called *autosomes*. The remaining pair consists of the sex chromosomes and differs in males and females. Females receive an X chromosome from each parent, whereas males receive an X chromosome from their mother and a Y chromosome from their father. A person's genotype is a result of the union of 23 maternal and 23 paternal chromosomes at conception. Sexual reproduction allows the genomes of the parents to mix, producing offspring that differ genetically from one another and from their parents. This source of genetic variability is advantageous to the species because it allows for adaptation and evolution in a changing environment.

For the two germ cells (i.e., egg and sperm) to combine to form a cell with the normal complement of 46 chromosomes (23 pairs), each germ cell must contribute half of the total. *Meiosis* refers to a special form of cell division that results in germ cells that are haploid; they have half of the normal number of chromosomes. In contrast to mitosis (see Chapter 3), meiosis involves two divisions of chromosomal DNA. A comparison of meiotic and mitotic cell division is shown in Fig. 6.3.

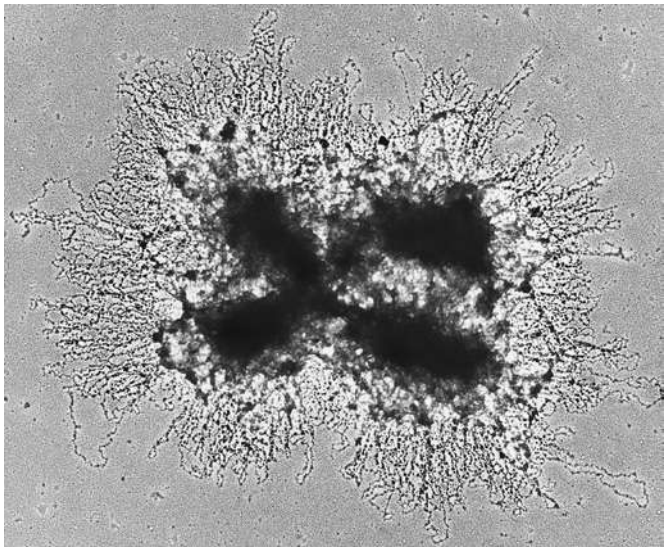


FIG 6.1 Scanning electron micrograph of a chromosome showing the two sister chromatids attached at the centromere. Sister chromatids separate during meiosis with one chromatid being distributed to each daughter cell. (From Pollard T et al: *Cell biology,* ed 2, Philadelphia, 2007, Saunders, p. 224.)

During the first phase of meiosis, pairs of homologous chromosomes with duplicated sister chromatids come in close contact. Portions of the homologous chromosomes are exchanged in a process called *crossing over* (Fig. 6.4). This results in a mixing of the maternal and paternal genes of the cell to form a new combination of genes within the chromosomes. Genetic recombination is very precise, such that genes are exchanged intact and not interrupted in the middle. On average, each homologous pair of chromosomes has one to three crossover events occurring during the first meiotic division. The first cellular division of meiosis results in two cells, each with 46 chromosomes. These two cells undergo a second division in which the sister chromatids are pulled apart (similar to normal mitosis), resulting in four cells, each having only 23 chromosomes. Each of the germ cells has a different combination of genes that, when passed on through sexual reproduction, will form a new, genetically unique individual.

The genes that code for a particular gene product, such as an enzyme, are located at a particular position (locus) on the chromosome. Gene loci are described by their position on the long arm (q arm) or the short arm (p arm) of the chromatid. For example, the gene locus 2p13 is located on the short arm of chromosome 2 at region 1, band 3 (Fig. 6.5). Genes come in several forms, called *alleles.* Normally, a person has two alleles for each gene, one received from each parent. If both alleles are functionally identical, the individual is said to be homozygous for that gene. If two different alleles are present, the individual is heterozygous. Sometimes, a particular allele can cause a trait to be expressed independent of the gene's homolog; such an allele is said to be dominant. The homologous allele whose effect is masked by a dominant allele is said to be recessive. A recessive trait is apparent only if *both* alleles for the trait are recessive (homozygous). It is important to remember that not all gene pairs occur in dominant and recessive forms.

Because transmission of single-gene (monogenic) traits from parent to offspring follows Mendel's rules, inheritance can be demonstrated using a Punnett square (Fig. 6.6). In a Punnett square, alleles for a gene

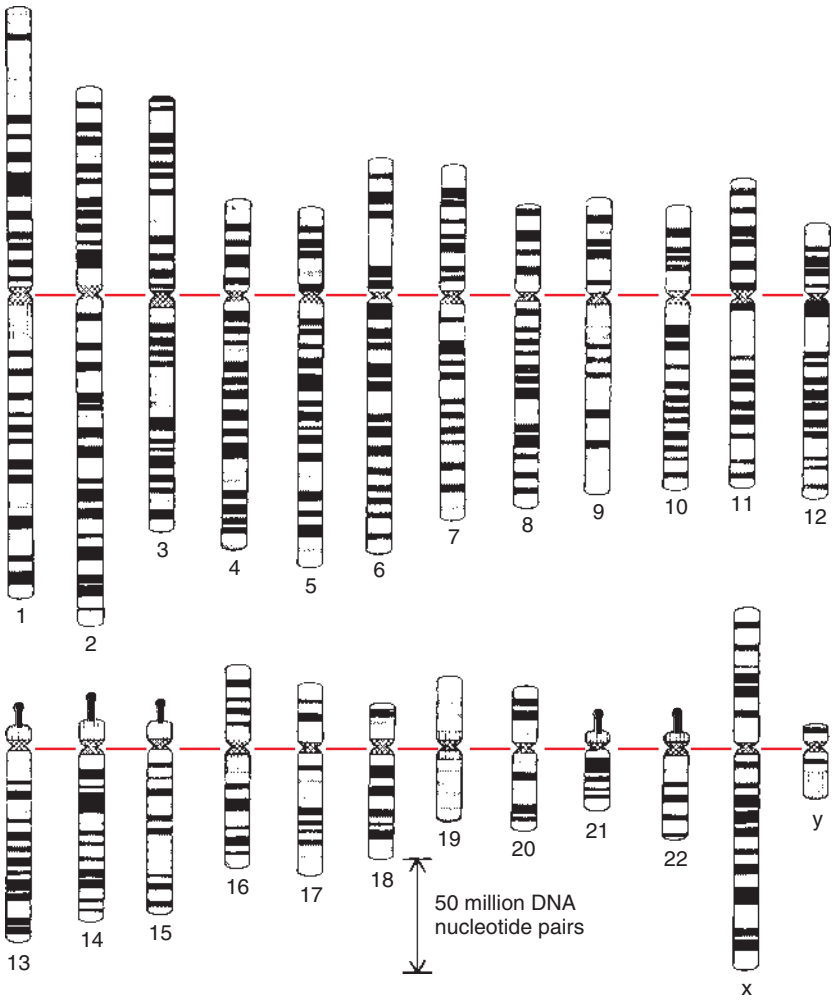


FIG 6.2 A standard map of the banding pattern of each of the 23 chromosomes of the human. Somatic cells contain two copies of each chromosome. The centromere region is marked by the line. (From Alberts B et al, editors: *Molecular biology of the cell,* ed 6, New York, 2015, Garland Science, p. 181.)

are represented by capital and lowercase letters. In the case of conditions with dominant and recessive alleles, a capital letter is used to depict the dominant allele, and a lowercase letter represents the recessive allele. Punnett squares are based on the mendelian principle that gene inheritance is random. Thus, if both parents are heterozygous for a dominant trait (Aa), the offspring will have a 25% probability of being AA, a 50% probability of being Aa, and a 25% probability of being aa. Persons having the AA and Aa genotypes may *express* the trait in a similar manner. If, however, the A allele is associated with a disease, being homozygous (AA) may be lethal. Achondroplasia and osteogenesis imperfecta are examples of mongenic conditions in which homozygosity for the disease allele is most often lethal. A dominant trait or condition will be absent in the aa genotype. Many genetic diseases are carried on a recessive allele and are manifested only in the homozygous (aa) genotype (e.g., cystic fibrosis, phenylketonuria [PKU]). Persons who are heterozygous (Aa) for conditions transmitted in a recessive pattern are said to be carriers because they are able to pass the defective recessive gene to their offspring even though they do not exhibit the trait.

Some alleles, perhaps even most alleles, are not clearly dominant or recessive and result in a blending, or *codominant* expression, of the trait. Blood type, for example, has three distinct alleles: A, B, and O. The A and B alleles may both be expressed, resulting in the AB blood type. Most traits result from the interaction of several genes and are called *polygenic*. Polygenic traits are heritable, but predicting their occurrence is more difficult than with single-gene traits. Polygenic traits are often affected by environmental factors (such as dietary intake, activity level, stress, and environmental exposures), which affect the ultimate expression of those genes. Such traits are said to be multifactorial; examples include height, weight, and blood pressure. Most common diseases, such as heart disease, asthma, diabetes, and cancer, are multifactorial as well.

DNA Mutation and Repair

The term *mutation* refers to a permanent change in DNA structure. Mutations can occur spontaneously, most often when copying errors occur during DNA replication, or they can be caused by exposure to mutagens such as radiation, chemicals, or viruses. Regardless of the cause, DNA changes rarely result in permanent alterations (mutations). The stability of the genes, and thus the low mutation rate, depend on efficient DNA repair mechanisms.

There are a variety of cellular DNA repair mechanisms. Most require the presence of a normal complementary DNA template to correctly repair the damaged strand of DNA. Single-stranded breaks (loss of bases from only one DNA strand) are therefore readily repaired. Double-stranded breaks, involving both strands of complementary DNA, may result in permanent loss of genetic information at the break point when

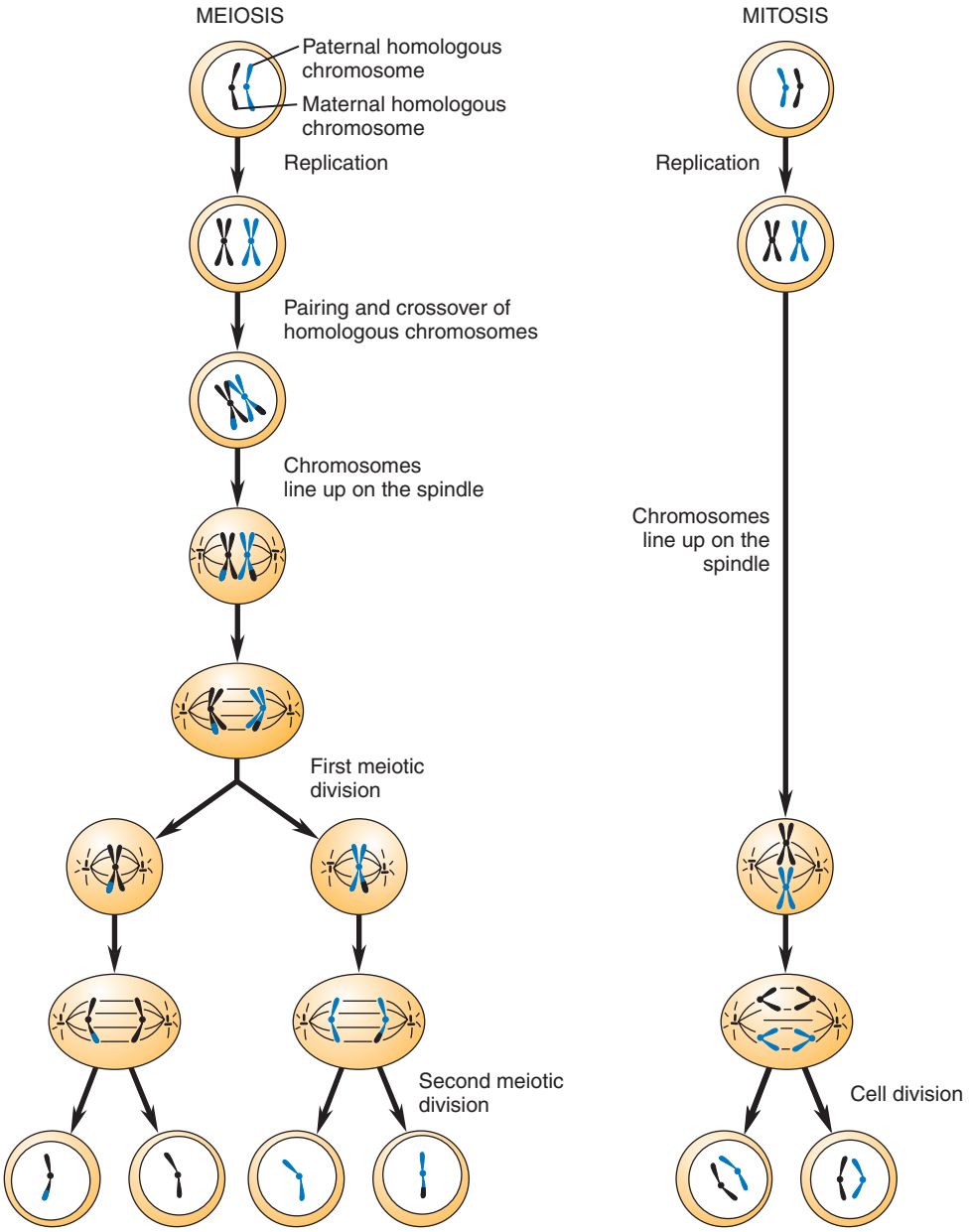


FIG 6.3 Comparison of meiosis and normal mitotic cell division, showing only one homologous chromosome pair. In meiosis, the homologous chromosomes form a pair and exchange sections of DNA in a process called *crossing over*. Two nuclear divisions are required in meiosis to form the haploid germ cells.

the broken strands are reunited. Different types of DNA damage are detected and repaired by different enzyme systems. The steps in one type of DNA repair are shown in Fig. 6.7.

Genetic mutations are generally of two types: mutations that *do* or *do not* change the "reading frame" of the gene. Recall that a sequence of three DNA bases (a codon) is required to code for each amino acid and that amino acids are connected together in a long chain to make a protein. Some mutations alter a codon so that it encodes a different amino acid; in that case, the resulting protein may not function as expected. In some cases the inclusion of the wrong amino acid in the protein is of no clinical significance; however, sickle cell anemia and beta thalassemia are examples of diseases that occur when a single-base (point) mutation causes substitution of a single amino acid. A frameshift mutation occurs when one or more bases are added or deleted and change the "reading frame" of the DNA sequence. The DNA sequence is normally "read" in groups of three bases, with no spaces between codons. If base pairs are added or deleted, all of the downstream codons in that gene may be changed, resulting in a protein with a greatly altered amino acid sequence (Fig. 6.8). Larger mutations may also occur and alter chromosome structure through loss, gain, or translocation of chromosome segments. These processes are discussed later in the chapter.

KEY POINTS

- Human DNA is organized into 46 chromosomes (23 pairs). Paired chromosomes look similar under the microscope but differ in DNA sequence. One member of each pair is inherited from the mother and the other from the father.
- Twenty-two pairs of chromosomes are autosomes. The remaining pair, the sex chromosomes, confers maleness (XY) or femaleness (XX).

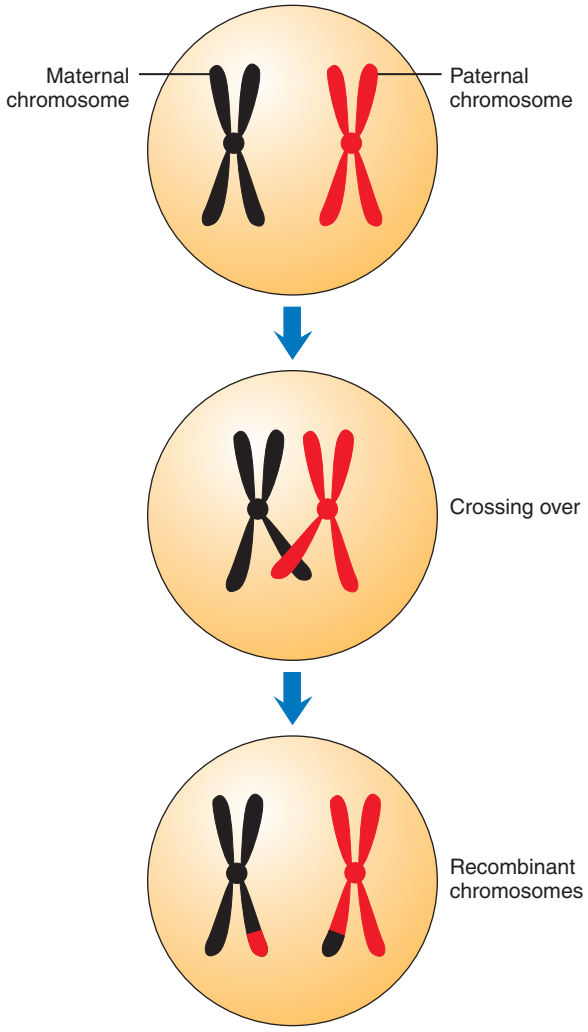


FIG 6.4 Crossing over during meiotic prophase I results in a reassortment of genes between homologous chromosomes.

- During meiotic cell division, the chromosomes are distributed to daughter cells. Meiosis results in four daughter cells, each having half the normal number of chromosomes (23 chromosomes).
- Genes that code for a particular trait may come in several forms or alleles. Genotype refers to the particular set of alleles an individual receives. Phenotype refers to an individual's observable attributes. People with different genotypes may have similar phenotypes.
- Some traits involve only one gene locus and are called *single-gene traits.* The transmission of these traits from parent to offspring follows predictable patterns. The expression of single-gene traits is determined by whether the gene is dominant or recessive. Most traits result from the interaction of several genes. These polygenic traits do not follow predictable patterns of inheritance.

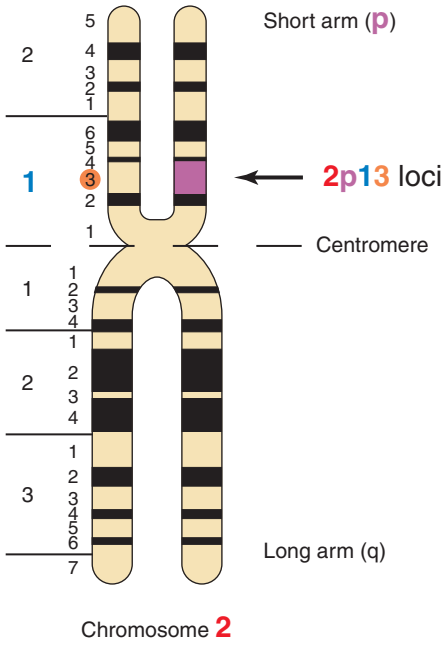


FIG 6.5 Metaphase chromosome showing location of centromere and long and short arms of the chromatids. Gene loci are described by the chromosome number, location on short (p) or long (q) arm, region, and band.

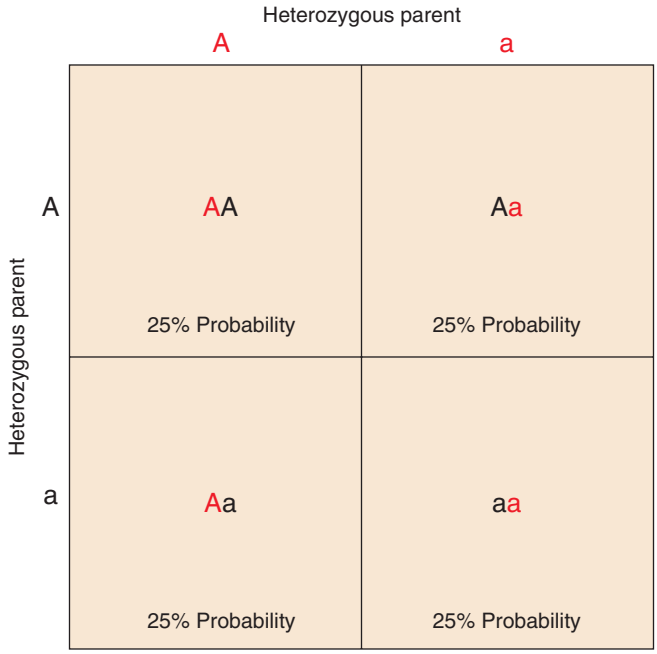


FIG 6.6 A Punnett square shows the distribution of parental genes to their offspring. This example shows the mating of two heterozygous individuals. A, Dominant gene; a, recessive gene.

GENETIC DISORDERS

Genetic disorders may be apparent at birth or may not be clinically evident until much later in life. The majority of genetic disorders are inherited from the affected individual's parents; however, new (de novo) mutations sometimes occur during gamete formation or arise during fetal development. Genetic disorders encountered clinically are only a small percentage of those that occur and represent the less extreme aberrations that permit live birth.

Disorders that are genetic in origin traditionally have been divided into three groups: (1) chromosomal aberrations, (2) mendelian single-gene disorders, and (3) polygenic or multifactorial disorders. A fourth group encompasses a number of single-gene defects that do not follow classic mendelian patterns of inheritance. This group includes triplet

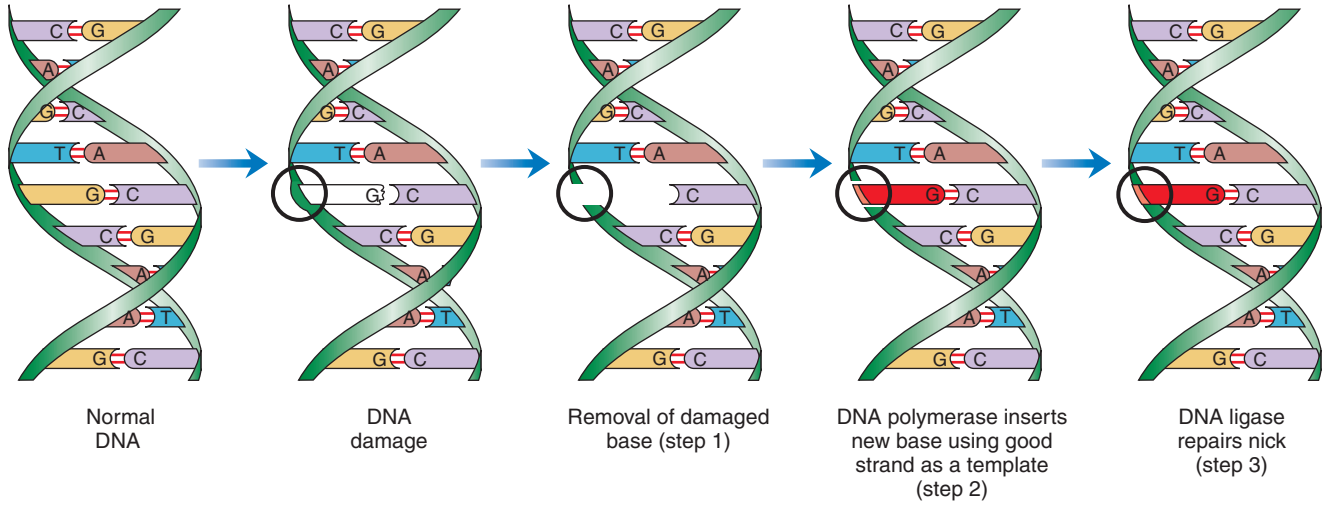


FIG 6.7 Steps of DNA repair. In step 1 the damaged section is removed; in steps 2 and 3 the original DNA sequence is restored.

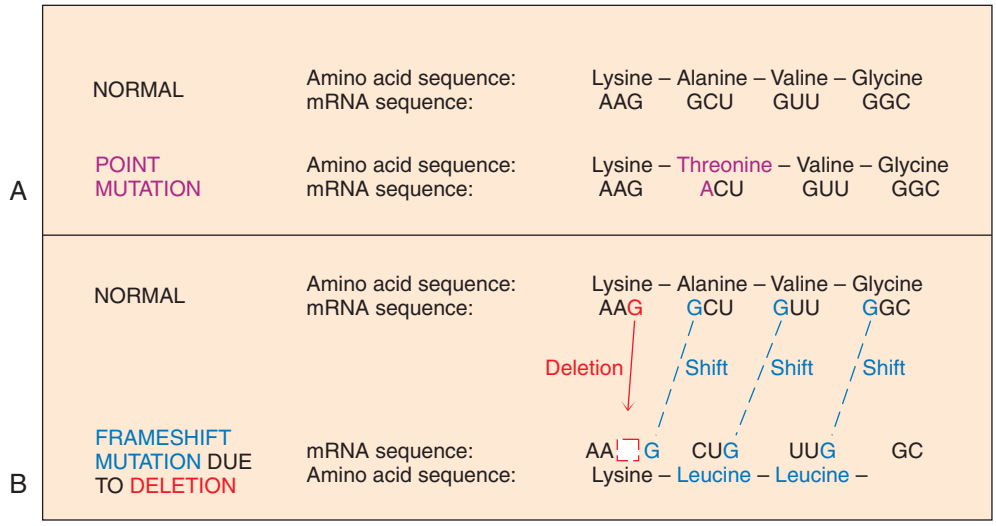


FIG 6.8 Schematic illustration of mutations that alter the messenger RNA sequence and the resulting protein amino acid sequence. **A,** Point mutation alters one amino acid. **B,** Frameshift mutation alters all downstream amino acids.

repeat (trinucleotide) mutations, mitochondrial gene mutations, and mutations influenced by genomic imprinting. General principles of transmission and selected examples are included for each of the four groups.

CHROMOSOMAL ABNORMALITIES

Chromosomal defects are generally due to an abnormal number of chromosomes or alterations in the structure of one or more chromosomes. Errors in the separation of chromosomes during meiosis may result in abnormal numbers of chromosomes. Extra or missing chromosomes and large structural defects (involving more than about 5 million base pairs) can be detected using traditional genetic testing methods such as karyotyping. These sorts of chromosomal abnormalities occur commonly but are usually lethal. Although they are found in at least half of spontaneous pregnancy losses (miscarriages), they have been found in only 0.5% to 1% of live-born newborn infants. In recent years, the development of molecular testing methods has allowed detection of much smaller structural defects, increasing the detection rate of chromosomal abnormalities.

Aberrant Number of Chromosomes

The union of a human of sperm and egg results in a fertilized egg (zygote) with the full complement of 46 chromosomes: 22 pairs of autosomes and 2 sex chromosomes (euploid). *Aneuploidy* refers to an abnormal number of chromosomes—in humans, either more or less than 46. Aneuploidy is most commonly caused by nondisjunction. *Nondisjunction* means that paired homologous chromosomes fail to separate normally during one of the meiotic divisions (Fig. 6.9). The resulting germ cells then have an abnormal number of chromosomes; for example, one germ cell may have only 22 chromosomes and the other has 24 chromosomes. When one of these abnormal germ cells combines with a normal germ cell containing 23 chromosomes, the resulting zygote will either be deficient by one chromosome (45) or have an extra chromosome (47). Another cause of aneuploidy is anaphase lag, in which one chromosome lags behind and is therefore left out of

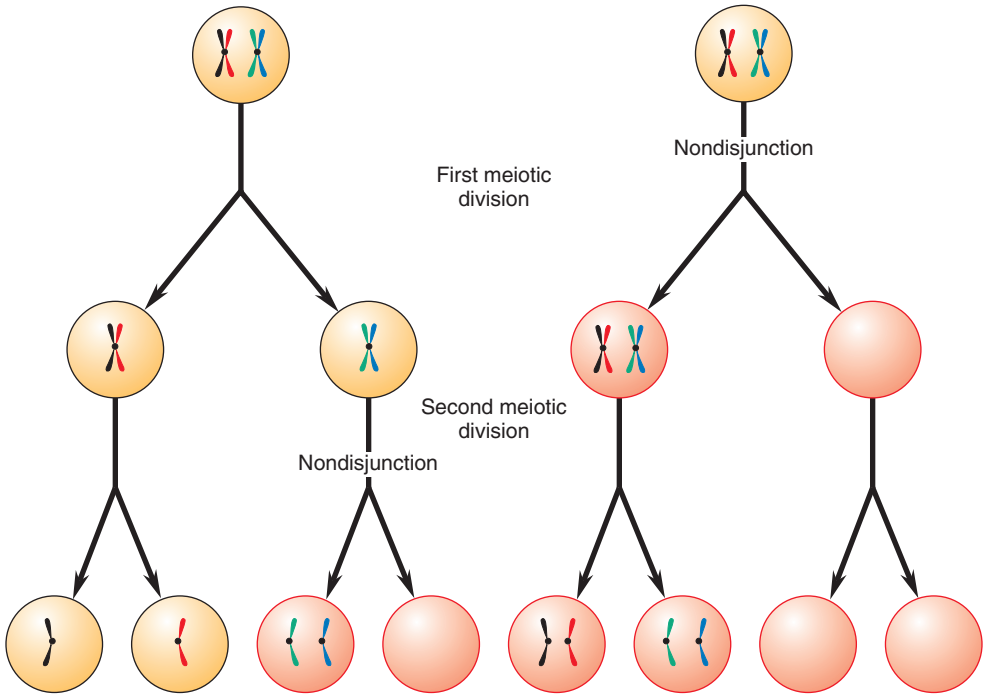


FIG 6.9 Mechanism of nondisjunction leading to aneuploidy. For simplicity, only one pair of chromosomes is shown.

the newly formed cell nucleus. This results in one daughter cell with the normal number of chromosomes and one with a deficiency of one chromosome, a condition called *monosomy*. *Polysomy* refers to the condition of having too many chromosomes.

Aneuploidy is thought to occur rather frequently during human gametogenesis and has been associated with advanced maternal age, abnormalities in parental chromosome structure, and abnormalities in crossing over. Most often, aneuploid zygotes are nonviable. In particular, monosomy involving an autosome is not usually compatible with life. Autosomal polysomy involving a single extra copy of a chromosome carrying a small number of genes may result in a viable fetus. Severe disability, however, nearly always results (e.g., trisomy 21—Down syndrome). Disorders involving extra or missing sex chromosomes are more common and less debilitating.

Abnormal Chromosome Structure

Alterations in chromosome structure are usually due to breakage and loss or rearrangement of chromosome fragments during meiosis or mitosis. During meiosis, the homologous chromosomes normally pair up and exchange genetic alleles in a process called *crossing over.* Normal crossing over involves precise gene exchange between homologues, with no net gain or loss of DNA. When the normal process of crossing over goes awry, portions of chromosomes may be lost, attached end-for-end, or attached to the wrong chromosome. Mitosis also presents opportunities for chromosomal breakage and rearrangement. The severity of chromosomal rearrangement ranges from insignificant to lethal, depending on the number and importance of the genes involved. The common types of chromosomal rearrangements are translocations, inversions, deletions, and duplications (Fig. 6.10).

Chromosomal *translocations* result from the exchange of pieces of DNA between nonhomologous chromosomes. If no genetic material is lost, as in a balanced or reciprocal translocation, the individual may have no symptoms or disorder. However, an individual with a balanced translocation is at increased risk of producing abnormal gametes. The exchange of a long chromatid arm for a short one results in the formation of one very large chromosome and one very small chromosome (see Fig. 6.10). This is called a *robertsonian translocation* and is responsible for a rare hereditary form of Down syndrome, discussed later in the chapter. Isochromosomes occur when the sister chromatids separate incorrectly at the centromere such that the two identical short arms remain together, as do the two long arms.

Inversion refers to the removal and end-for-end reinsertion of a section of chromosome (see Fig. 6.10). Like balanced translocations, inversions involve no net loss or gain of genetic material and are often without consequence to the individual. Difficulties result, however, when homologous chromosomes attempt to pair up during meiosis. The chromosome with an inverted section may not pair up properly, resulting in duplications or loss of genes at the time of crossing over. Thus the offspring of an individual harboring an inversion may be affected.

Loss of chromosomal material is called *deletion.* Deletions may result when a break occurs in the arm of a single chromosome, creating a fragment of DNA with no centromere. The fragment is then lost at the next cell division. Chromosomal deletions have been associated with some forms of cancer, including retinoblastoma (see Chapter 7). Deletions at both ends of a chromatid may cause the free ends to attach to one another, forming a ring chromosome.

In contrast to a deletion, where genes are lost, *duplication* results in extra copies of a portion of DNA. The consequences of duplications are generally less severe than those from loss of genetic material.

Examples of Autosomal Chromosome Disorders

Trisomy 21 (Down Syndrome)

Trisomy 21 is a chromosomal disorder in which individuals have an extra copy of chromosome 21. It is the most common of the chromosomal disorders and a leading cause of mental disability, estimated to occur in about 1 in 700 live births. The incidence varies among populations and over time, however, and reflects trends in maternal age, prenatal diagnosis, and termination of affected pregnancies. The syndrome was

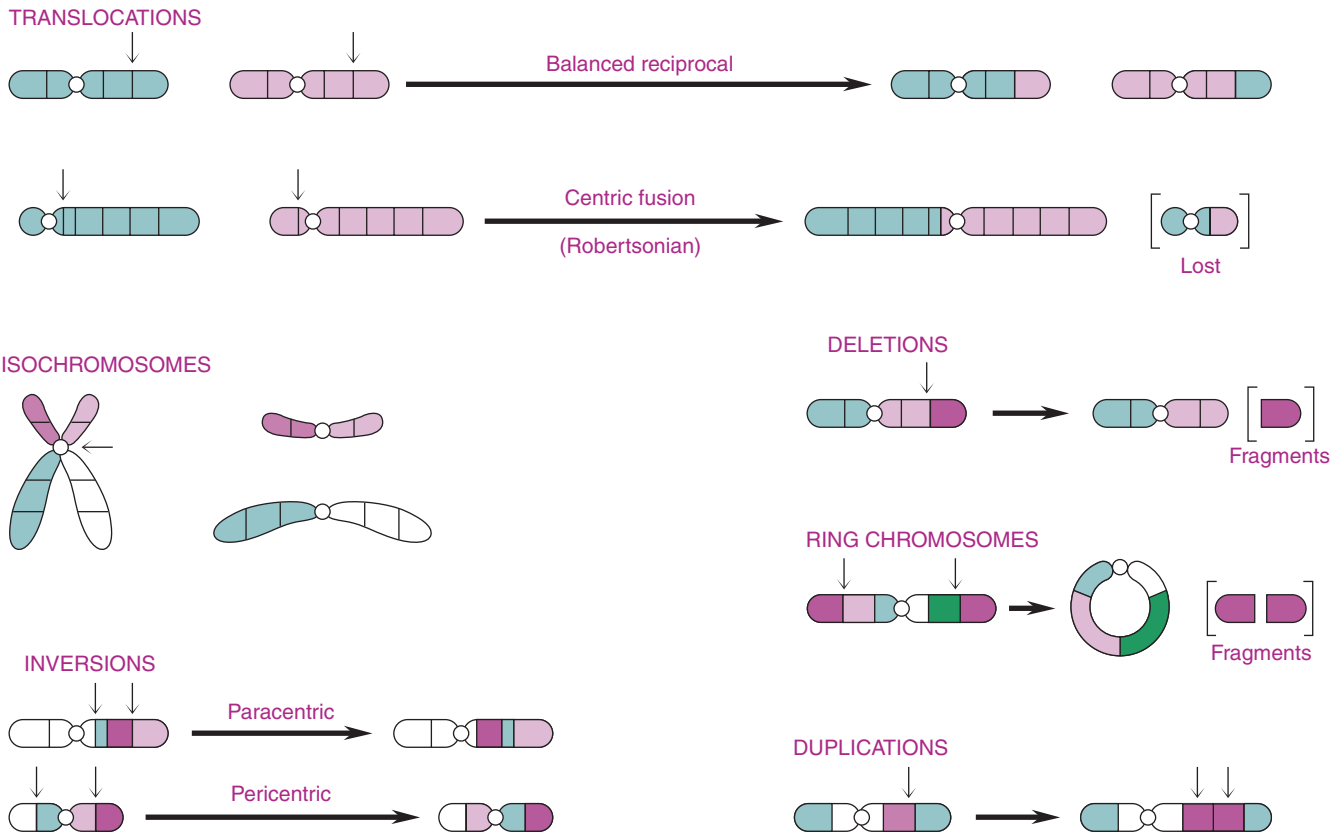


FIG 6.10 Types of chromosomal rearrangement. (Adapted from Kumar V et al: *Robbins and Cotran pathologic basis of disease,* ed 8, Philadelphia, 2010, Saunders, p. 160.)

first described by Langdon Down in 1866 and is characterized by intellectual disability, protruding tongue, low-set ears, epicanthal folds, poor muscle tone, and short stature (Fig. 6.11). Children with Down syndrome often are afflicted with congenital heart deformities and an increased susceptibility to respiratory tract infections, leukemia, and early-onset Alzheimer disease. The precise causes of these signs and symptoms are poorly understood, but the increased number of genes (or gene dose) is implicated, rather than errors in DNA sequence.

In 95% of cases, the extra chromosome 21 is thought to be of maternal origin, and the incidence of trisomy 21 is clearly associated with advanced maternal age. Table 6.1 demonstrates a rise in the incidence of Down syndrome from maternal age 20 to 50 years. The reason for increased susceptibility of the ovum to nondisjunction with age remains unknown. A less common form of Down syndrome (occurring in about 4% of cases) is due to a chromosomal translocation of the long arm of chromosome 21 to another chromosome; this is an example of a robertsonian translocation. This form of Down syndrome is not associated with increased maternal age but is passed from parent to offspring. Testing for translocations in newborns with Down syndrome is recommended to determine recurrence risk for families.

Trisomy 18 (Edwards Syndrome) and Trisomy 13 (Patau Syndrome)

Trisomy of chromosome 18 or 13 occurs much less commonly than trisomy 21 and is more severe. Most affected pregnancies are lost before term, and live-born infants usually do not survive more than a few days or weeks. It is not insignificant that the few trisomy conditions compatible with life involve chromosomes containing the smallest numbers of genes. Trisomies involving chromosomes 8, 9, and 22 also have been described but are extremely rare.

Cri du Chat Syndrome

Deletion of part of the short arm of chromosome 5 results in a syndrome characterized by severe mental retardation, round face, and congenital heart anomalies. The syndrome was so named because of the characteristic cry of the affected infant, which is caused by laryngeal malformation and resembles a cat crying. Some children afflicted with this syndrome survive to adulthood, and they generally thrive better than those with the trisomies.

Examples of Sex Chromosome Disorders

Klinefelter Syndrome

Aneuploidies involving the sex chromosomes X or Y are much better tolerated than autosomal aneuploidies. Klinefelter syndrome occurs when a male has one or more extra copies of the X chromosome. The incidence of Klinefelter syndrome is about 1 in 600 live-born males, making it the most common sex chromosome abnormality. Most individuals with Klinefelter syndrome have a single extra copy of the X chromosome (an XXY genotype); however, individuals with more than one extra X (XXXY and XXXXY) have also been described. The presence of the Y chromosome determines the sex of these individuals to be male; however, the extra X chromosomes result in abnormal sexual development and feminization. The condition may be diagnosed at puberty, when failure to develop secondary sex characteristics may become apparent. Associated symptoms reflect low testosterone levels. Hypogonadism (small testicles) is universal in Klinefelter syndrome, most often associated with infertility. Tall stature with long arms and legs, a feminine hair distribution, gynecomastia (breast enlargement), and marginally impaired intelligence are common (Fig. 6.12). Testosterone therapy can achieve a dramatic reduction in the feminine characteristics associated with Klinefelter syndrome.

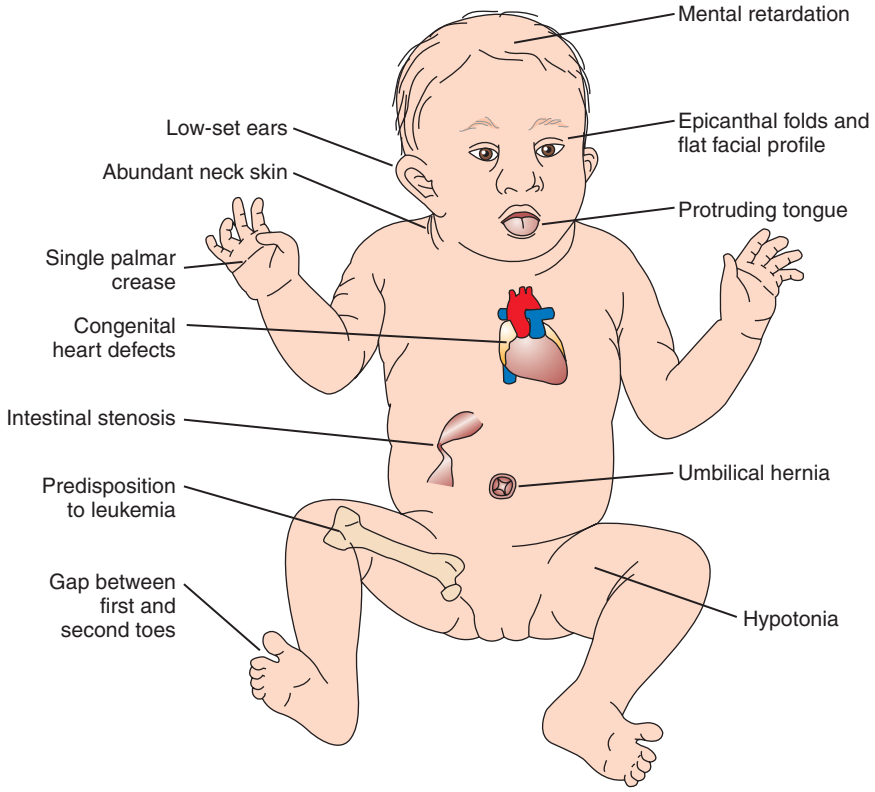


FIG 6.11 Typical clinical manifestations of trisomy 21 (Down syndrome).

TABLE 6.1 Frequency of Trisomy 21 (Down Syndrome) in Relation to Maternal Age

Age of Mother at Birth (Year)	Frequency of Trisomy 21 at Birth
20	1/1470
25	1/1333
30	1/935
35	1/353
37	1/200
39	1/112
41	1/68
43	1/46
45	1/36
50	1/26

Data from Morris JK et al: Comparison of models of maternal age-specific risk for Down syndrome live births, *Prenat Diagn* 23:252–258, 2003.

Turner Syndrome

Also known as *monosomy X*, Turner syndrome is associated with the presence of only one normal X chromosome and no Y chromosome. The absence of the Y chromosome results in a female phenotype; however, the ovaries fail to develop or fail prematurely. In some cases of Turner syndrome, the second X chromosome is not entirely missing but is structurally abnormal. In the majority of cases, the missing or damaged X chromosome is of paternal origin. Most fetuses with monosomy X are lost during pregnancy, and the incidence is about 1 in 2500 live female births. Principal characteristics of Turner syndrome include short stature, webbing of the neck, a wide chest, lymphedema of the hands and feet at birth, and failure to develop secondary sexual characteristics. Many girls with Turner syndrome also have congenital heart defects (Fig. 6.13).

Multiple X Females and Double Y Males

A relatively common disorder of the sex chromosomes is the presence of an extra copy of the X chromosome in females (XXX) or of the Y chromosome in males (XYY). Most individuals appear normal and have few physical problems, although IQ may be marginally impaired.

MENDELIAN SINGLE-GENE DISORDERS

In contrast to the chromosomal disorders described earlier, mendelian disorders result from mutations in single genes. The affected genes may code for enzymes, structural proteins, regulatory proteins, or regulatory RNA molecules that, due to the gene mutation, do not function as expected. An individual normally has two copies or alleles of each gene (one allele from each parent). A recessive disorder is expressed only when the individual is homozygous for the altered gene; that is, the individual has two recessive copies. Dominant disorders are expressed despite the presence or configuration of the other allele of the pair. Mendelian disorders are generally classified according to the *location* of the defective gene (autosomal or sex chromosome) and the *mode of transmission* (dominant or recessive). The great majority of mendelian disorders are familial (attributable to mutated genes inherited from the parents), but 15% to 20% represent new mutations. The occurrence of new mutations ranges widely. For example, new mutations for Huntington disease are rare, whereas more than 80% of individuals with achondroplasia have new mutations. A genetic pedigree may be used to trace the transmission of the disease through a family. The pedigree (Fig. 6.14) graphically displays family relationships and family members

FIG 6.12 Typical clinical manifestations of Klinefelter syndrome. (From Moore KL, Persuad TVN: *The developing human: clinically oriented embryology,* ed 8, Philadelphia, 2007, Saunders, p. 466.)

affected by diseases or conditions and is a useful tool in determining the pattern of inheritance as recessive, dominant, or sex-linked. Mendelian genetics is based on the principle that single genes are randomly and independently transmitted to offspring such that there is a 50:50 chance of receiving one or the other of a parent's alleles for a particular gene. It is important to note that there are many exceptions to these rules, but they generally are useful in predicting transmission patterns for a number of single-gene disorders. Thousands of single-gene traits and disorders have been identified. A comprehensive database of the chromosomal location and sequence of these single-gene traits and disorders, called *Online Mendelian Inheritance in Man,* can be accessed at http://www.ncbi.nlm.nih.gov/omim.

Autosomal-Dominant Disorders

Autosomal-dominant disorders occur when a mutation in a gene located on one of the autosomes is sufficient to cause the disorder despite the presence of a normal allele for the same gene. Autosomal-dominant disorders follow predictable patterns of inheritance (Fig. 6.15), which may be summarized as follows:

- Males and females are equally affected.
- Affected individuals usually have an affected parent.
- Unaffected individuals do not transmit the disease.
- Offspring of an affected individual (with an unaffected mate) have a 1 in 2 chance of inheriting the disease.
- The rare mating of two individuals, each carrying one copy of the defective gene (heterozygous), results in a 3 in 4 chance of producing an affected offspring. For some disorders, having two copies of the defective gene is lethal.

The list of known autosomal-dominant disorders is long. Many are described in later chapters as they relate to system pathophysiology. A partial list is presented in Table 6.2. In general, autosomal-dominant

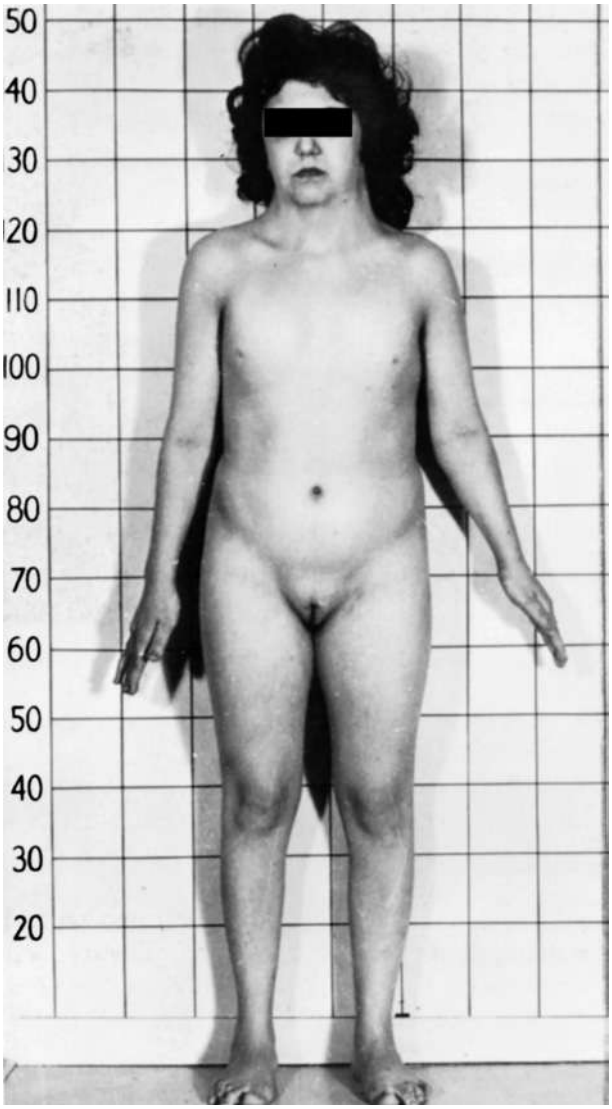


FIG 6.13 Typical clinical manifestations of Turner syndrome. (From Connor JM, Ferguson-Smith MA: *Essential medical genetics,* ed 5, London, 1997, Blackwell Scientific, p. 123.)

disorders involve key structural proteins or regulatory proteins, such as membrane receptors. Marfan syndrome and Huntington disease are commonly cited examples of autosomal-dominant disorders and are briefly described here.

Marfan Syndrome

Marfan syndrome is a disorder of the connective tissues of the body. Individuals with Marfan syndrome are typically tall and slender with long, thin arms and legs (Fig. 6.16). Because of the long, thin fingers, this syndrome has also been called *arachnodactyly* ("spider fingers"). It is commonly suggested that President Abraham Lincoln may have had this disorder. Although skeletal and joint deformities can be problematic, cardiovascular lesions are the most life threatening. The medial layer of blood vessels, particularly the aorta, tends to be weak and susceptible to dilation and rupture. Dysfunction of the heart valves may occur from poor connective tissue support. Marfan syndrome has been traced to hundreds of different mutations in the fibrillin 1 *(FBN1)* gene on chromosome 15. Fibrillin 1 is a glycoprotein that provides structural support in microfibil-rich connective tissues. Marfan syndrome

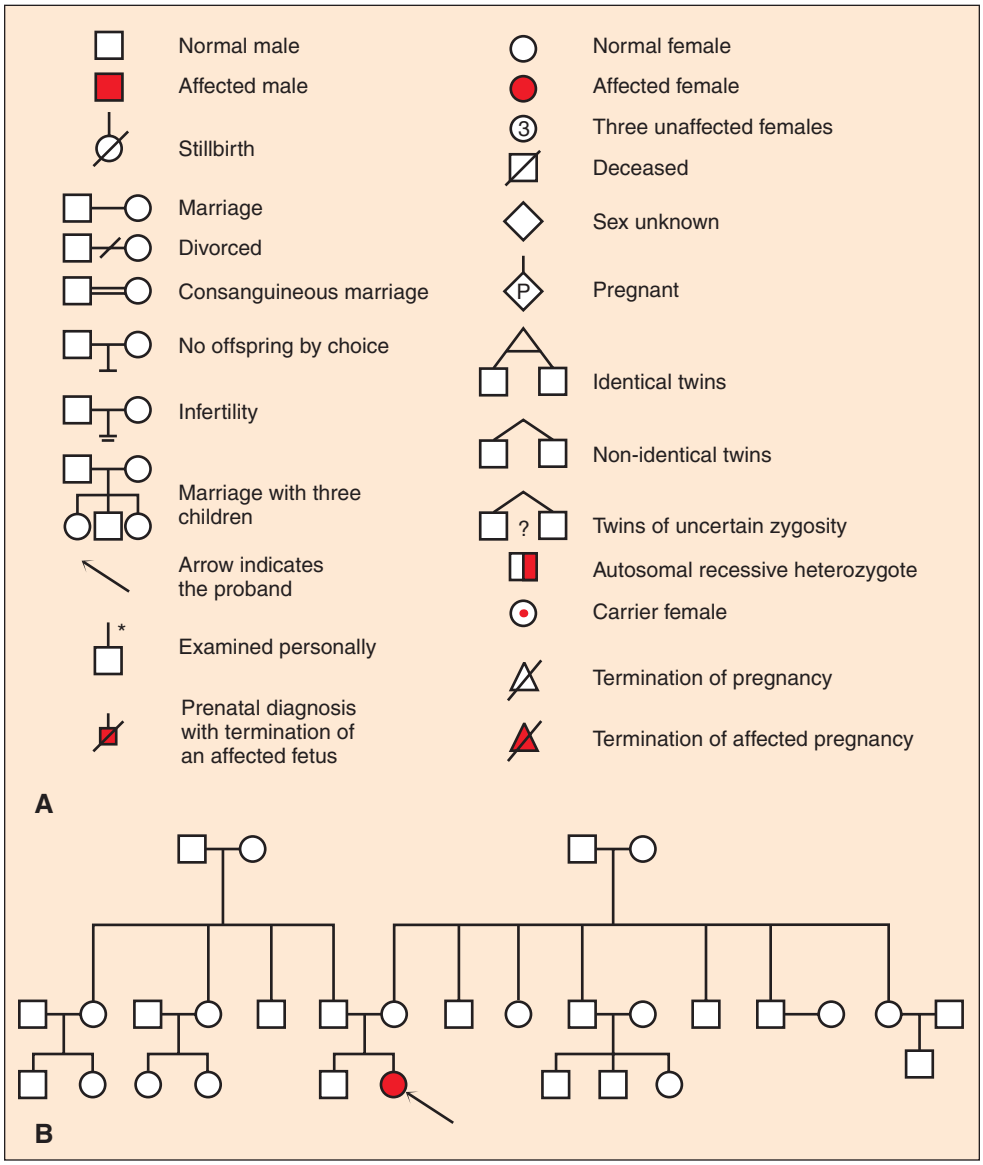


FIG 6.14 **A,** Common symbols for pedigree analysis. **B,** Typical family pedigree chart.

occurs when mutations in the fibrillin 1 gene cause abnormal fibrillin or low levels of fibrillin to be produced, leading to weakened connective tissues.

Huntington Disease

Huntington disease is an autosomal-dominant disease that primarily affects neurologic function. The symptoms of mental deterioration and involuntary movements of the arms and legs do not appear until approximately age 40 years. The disease was formerly called *Huntington chorea* (from the Greek *khoreia,* meaning "dance") because of the uncontrolled movements of the limbs. The delayed onset of symptoms means that the disease may be transmitted to offspring before the parent is aware that he or she harbors the defective gene. The prevalence rate is about 1 in 20,000 persons.

The gene abnormality in Huntington disease has been localized to chromosome 4, where an abnormally large number of triplet repeats (CAG) has been noted. Triplet repeats of more than 40 are reliably associated with development of the disease, and the greater the number of triplet repeats, the earlier the onset of symptoms. The Huntington disease protein (huntingtin) has a long segment of glutamine amino acids that are coded by the CAG triplet repeat. Extra CAG repeats cause extra glutamines in the huntingtin protein, and by a mechanism that is poorly understood, the abnormally long polyglutamine repeat in the protein causes the degeneration of nerve cells in specific brain regions. Clinical manifestations and pathophysiology of Huntington disease are discussed in Chapter 45. Other disorders are also caused by expanded triple repeats; these disorders usually affect the neurologic system and are collectively called *trinucleotide repeat disorders.*

Autosomal-Recessive Disorders

Like autosomal-dominant diseases, autosomal-recessive disorders occur due to a mutation of a gene located on one of the autosomes. Autosomal-recessive disorders, however, occur only when both alleles of a particular gene are mutated. Predictable patterns of inheritance (Fig. 6.17) are summarized as follows:

- Males and females are equally affected.
- In most cases, the disease is not apparent in the parents or relatives of the affected individual, but both parents are carriers of the mutant recessive gene.
- Unaffected individuals may transmit the disease to offspring.

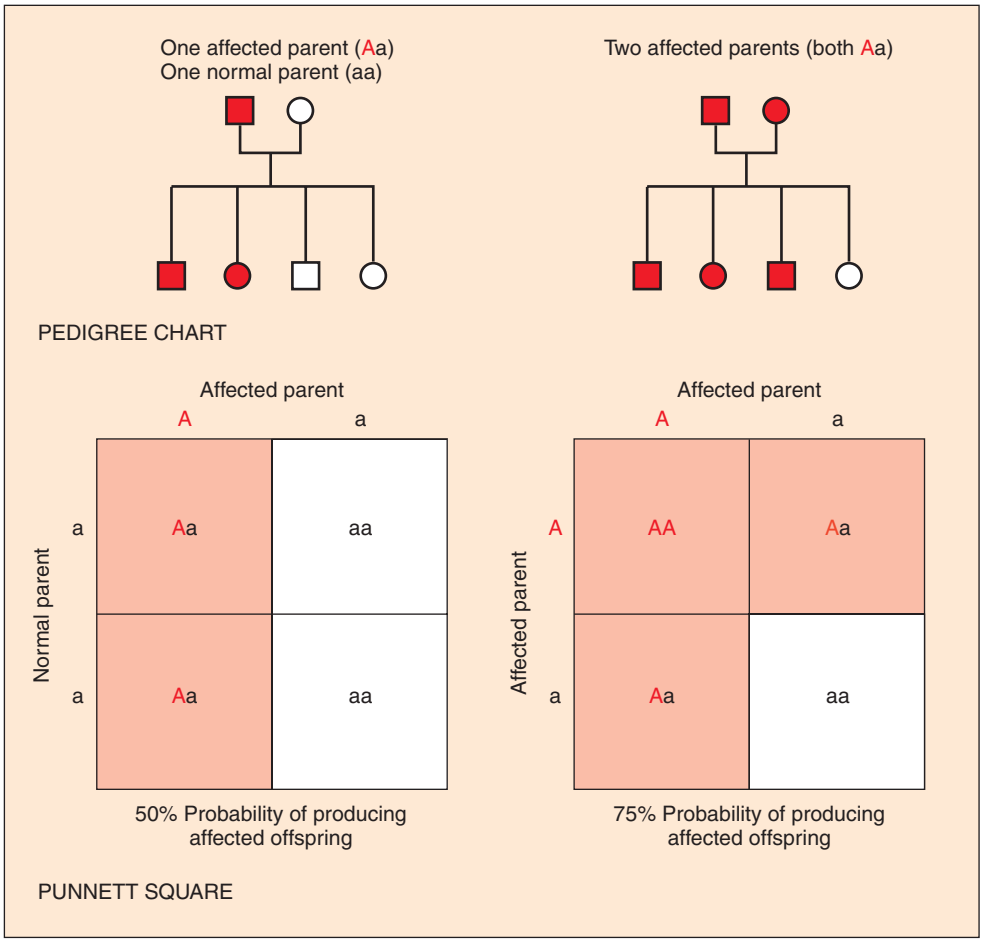


FIG 6.15 Typical pattern of inheritance of an autosomal-dominant trait (e.g., Marfan syndrome). **A,** Pedigree chart. **B,** Punnett square.

TABLE 6.2 Autosomal-Dominant Disorders

System	Disorder
Nervous	Huntington disease
	Neurofibromatosis
	Myotonic dystrophy
	Tuberous sclerosis
Urinary	Polycystic kidney disease
Gastrointestinal	Familial polyposis coli
Hematopoietic	Hereditary spherocytosis
	Von Willebrand disease
Skeletal	Marfan syndrome
	Ehlers–Danlos syndrome (some variants)
	Osteogenesis imperfect
	Achondroplasia
Metabolic	Familial hypercholesterolemia
	Acute intermittent porphyria

From Kumar V et al: *Robbins & Cotran pathologic basis of disease,* ed 9, Philadelphia, 2015, Saunders, p. 141.

- The mating of two carriers (heterozygous) results in a 1 in 4 chance of producing an affected offspring and a 2 in 4 chance of producing an offspring who carries the disease.

It is estimated that nearly everyone carries several mutated recessive genes, which in the homozygous state would cause disease. The odds of two people who share the same recessive mutation parenting a child is relatively low, unless those two people are genetically related. Related individuals are much more likely to carry the *same* recessive genes. Because recessive diseases are only expressed when both alleles for a particular gene are mutated (homozygous), they are often associated with consanguinity—the mating of related individuals. The closer the biological relationship, the greater the proportion of shared genes and the greater the risk of producing affected offspring.

Recessive disorders often involve abnormal enzymatic function when the gene for a particular enzyme is absent or present in a mutated, nonfunctional form. Most often, enzyme deficiencies are not apparent in heterozygotes who carry one normal gene because the normal gene produces enough of the necessary enzyme. In the homozygous state, however, neither gene encoding the enzyme is functional, resulting in an enzyme deficiency. Table 6.3 includes a partial list of the large number of autosomal-recessive disorders that have been identified. Many of these diseases reflect inability to metabolize nutrients (inborn errors of metabolism) or to synthesize cellular components because of enzyme deficiencies. Albinism, PKU, and cystic fibrosis are described here as representative examples. Other disorders are described in the discussions of system pathophysiology in later chapters.

Albinism

Albinism refers to a lack of pigmentation of the hair, skin, and/or eyes. There are several types of albinism; all types involve disruption of melanin synthesis. Albinism is currently classified according to the

FIG 6.16 Clinical manifestations of Marfan syndrome. Skeletal deformities such as pectus excavatum and abnormal curvature of the thoracic spine are common findings. (From Turnpenny P: *Emory's elements of medical genetics,* ed 14, Philadelphia, 2012, Churchill Livingstone, p. 301.)

TABLE 6.3 Autosomal-Recessive Disorders

System	Disorder
Metabolic	Cystic fibrosis
	Phenylketonuria
	Galactosemia
	Homocystinuria
	Lysosomal storage disease
	α_1-Antitrypsin deficiency
	Wilson disease
	Hemochromatosis
	Glycogen storage diseases
Hematopoietic	Sickle cell anemia
	Thalassemias
Endocrine	Congenital adrenal hyperplasia
Skeletal	Ehlers–Danlos syndrome (some variants)
	Alkaptonuria
Nervous	Neurogenic muscular atrophies
	Friedreich ataxia
	Spinal muscular atrophy

From Kumar V et al: *Robbins & Cotran pathologic basis of disease,* ed 9, Philadelphia, 2015, Saunders, p. 141.

affected gene. For example, oculocutaneous albinism type 1 is associated with profound hypopigmentation of the skin and hair and reduced pigmentation of the iris and retina. Affected individuals are at risk for sunburn and skin cancer and generally exhibit impaired vision, nystagmus (involuntary eye movements), and photosensitivity.

Phenylketonuria

PKU results from an inability to metabolize the amino acid phenylalanine because of lack of the enzyme phenylalanine hydroxylase. It is one of several enzyme deficiencies that are often referred to as *inborn errors of metabolism.* The symptoms of the disorder are caused by accumulation of dietary phenylalanine in the body, which primarily affects the nervous system. Children with PKU tend to be overly irritable and tremulous and have slowly developing mental retardation. Excess phenylalanine is excreted in the urine in the form of phenylketones, hence the name phenylketonuria. Infants typically have a musty odor because of excess phenylalanine by-products in the sweat and urine. The enzyme deficiency can be detected soon after birth and managed with a low-phenylalanine diet. Because treatment must be instituted very early to prevent mental retardation, routine screening for PKU is performed at birth.

Cystic Fibrosis

Cystic fibrosis is one of the most common single-gene disorders. About 1 in 20 Caucasian Americans harbor the defective gene, and the incidence of cystic fibrosis is approximately 1 in 2500 live births. The clinical abnormalities associated with cystic fibrosis have been traced to a defect in a membrane transporter for chloride ions in epithelial cells. The alteration in chloride transport is associated with production of abnormally thick secretions in glandular tissues. The lung bronchioles and pancreatic ducts are primarily affected, often resulting in progressive destruction of these organs (see Chapter 22).

The cystic fibrosis gene was isolated in 1989 and mapped to chromosome 7. Designated *CFTR (for cystic fibrosis transmembrane conductance regulator),* the gene encodes a protein chloride channel of the same name. More than 1900 different mutations of this gene have been identified, all of which cause a defect in chloride transport across the cell membrane; however, the severity of chloride channel dysfunction varies widely with different mutations. The most common mutation, accounting for about 70% of cystic fibrosis cases, involves a deletion of three nucleotides that normally code for a phenylalanine at position 508 in the CFTR protein; this mutation is designated ΔF508. The absence of this single amino acid causes the protein to fold abnormally, preventing its release from the endoplasmic reticulum, where it is eventually degraded. A schematic of the normal CFTR protein is shown in Fig. 6.18. The protein belongs to the family of ABC transporters that bind and hydrolyze adenosine triphosphate (ATP) (see Chapter 3).

The discovery and characterization of the cystic fibrosis gene and CFTR protein have made it possible to envision effective gene therapy for this disorder. Clinical trials have been published; however, overcoming host immune responses has represented a persistent barrier. Some mutation-specific drugs have been effective; for example, a novel drug, ivacaftor, was approved in 2012 for the treatment of cystic fibrosis in patients who carry a fairly rare mutation, G551D, on at least one *CFTR* allele. Reliable genetic screening for mutations known to cause cystic fibrosis is readily available and included as part of newborn screening in many states, making early management possible.

Sex-Linked (X-Linked) Disorders

Sex-linked disorders occur because of a mutation of the sex chromosomes. Disorders linked to the Y chromosome are extremely rare, and for that reason the terms *sex-linked* and *X-linked* are often used interchangeably. Nearly all X-linked disorders are recessive, so females express the X-linked disease only in the rare instance in which both X chromosomes carry the defective gene. Males, however, do not have the safety margin of two X chromosomes and express the disease if their one and only X chromosome is abnormal. X-linked disorders follow predictable patterns

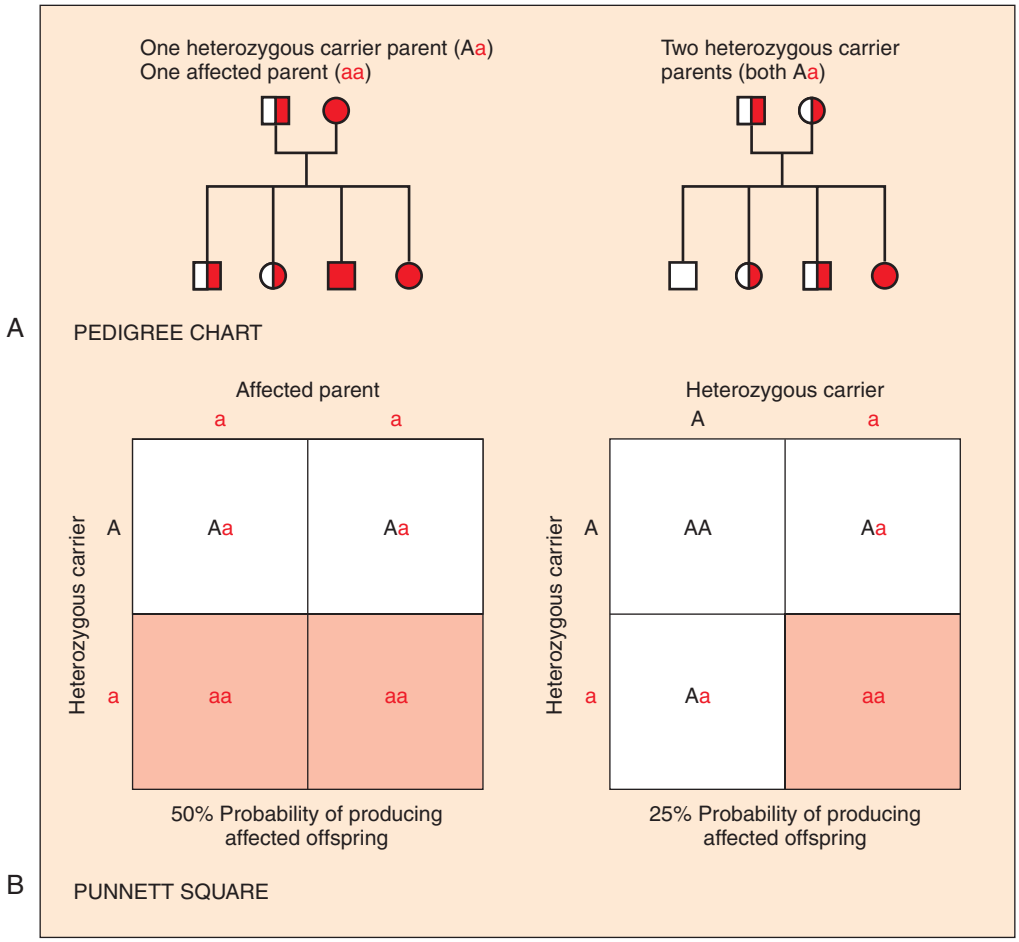


FIG 6.17 Typical pattern of inheritance of an autosomal-recessive trait (e.g., cystic fibrosis, sickle cell anemia). **A,** Pedigree chart. **B,** Punnett square.

of inheritance (Fig. 6.19), which are dependent on the sex of the offspring, and may be summarized as follows:

- Affected individuals are almost always male.
- Affected fathers transmit the defective gene to none of their sons but to all of their daughters.
- Unaffected males do not carry the defective gene.
- A carrier female has a 1 in 2 chance of producing an affected son and a 1 in 2 chance of producing a carrier daughter.
- Females are affected only in the rare homozygous state that may occur from the mating of an affected or carrier mother and an affected father.

Several X-linked recessive disorders have been identified, as presented in Table 6.4. A well-known example of an X-linked disease is hemophilia A.

Hemophilia A

Hemophilia A is a bleeding disorder associated with a deficiency of factor VIII, a protein necessary for blood clotting. Individuals afflicted with hemophilia A bleed easily and profusely from seemingly minor injuries (see Chapter 14). The transmission of hemophilia A in the European royal families constitutes one of the best-known pedigrees available (Fig. 6.20). Queen Victoria of England was the first known carrier of the disease, and a number of her male descendants were affected.

TABLE 6.4 X-Linked Recessive Disorders

System	Disorder
Musculoskeletal	Duchenne muscular dystrophy
Blood	Hemophilias A and B
	Chronic granulomatous disease
	Glucose-6-phosphate dehydrogenase deficiency
Immune	Agammaglobulinemia
	Wiskott–Aldrich syndrome
Metabolic	Diabetes insipidus
	Lesch–Nyhan syndrome
Nervous	Fragile X syndrome

From Kumar V et al: *Robbins & Cotran pathologic basis of disease,* ed 9, Philadelphia, 2015, Saunders, p. 142.

NONMENDELIAN SINGLE-GENE DISORDERS

Transmission of certain single-gene disorders may deviate from the classic mendelian principles of random and independent assortment. Three such categories have been described: (1) disorders caused by expanded triplet repeat mutations, such as fragile X syndrome; (2) disorders attributable to mitochondrial DNA mutations; and (3) disorders associated with genomic imprinting.

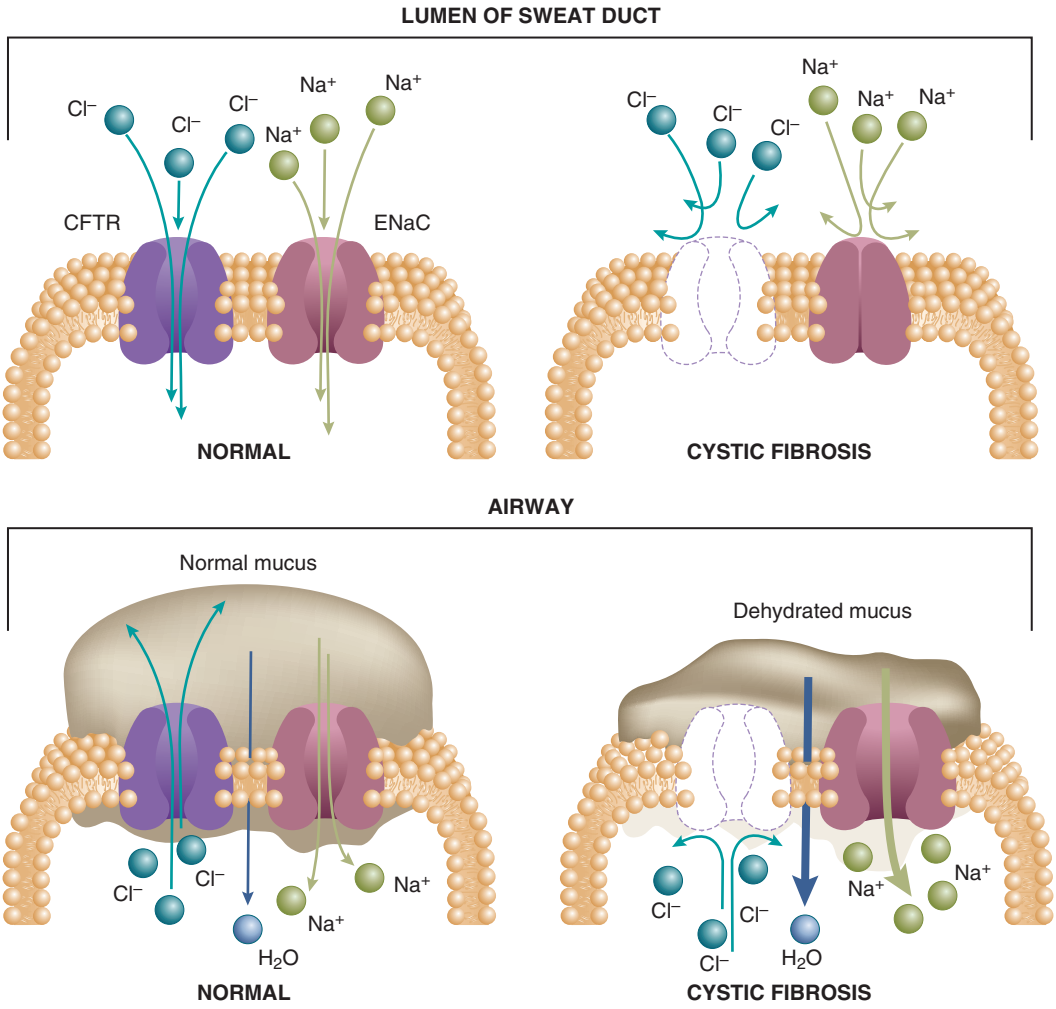


FIG 6.18 Schematic illustration of the cystic fibrosis transmembrane conductance regulator (CFTR) located in an epithelial cell. CFTR is a transmembrane protein that transports chloride from the cytoplasm into the lumen of the bronchiole. Mutations in the CFTR transporter gene are believed to cause the thick secretions typical of cystic fibrosis.

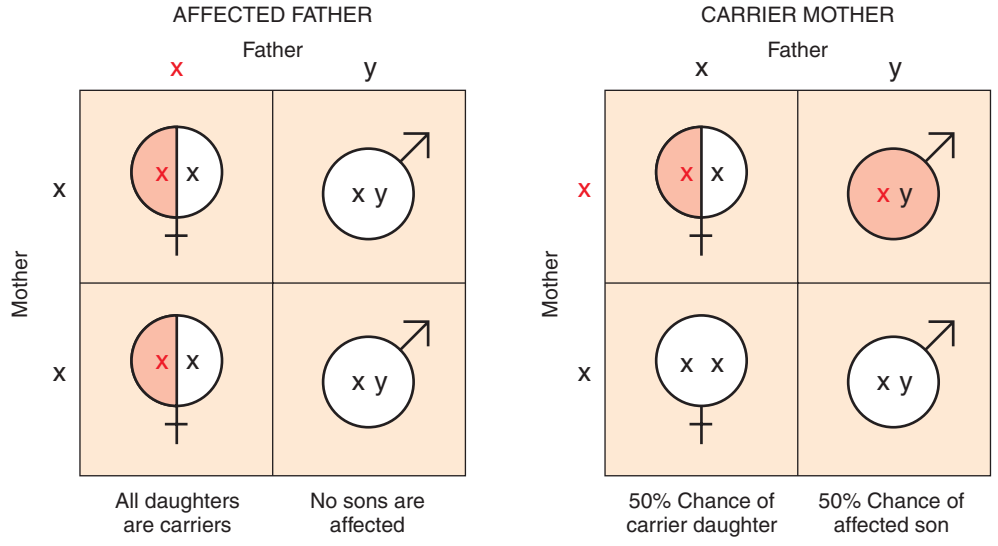


FIG 6.19 Typical inheritance pattern for X-linked disorders. The risk of disease varies according to the gender of the offspring.

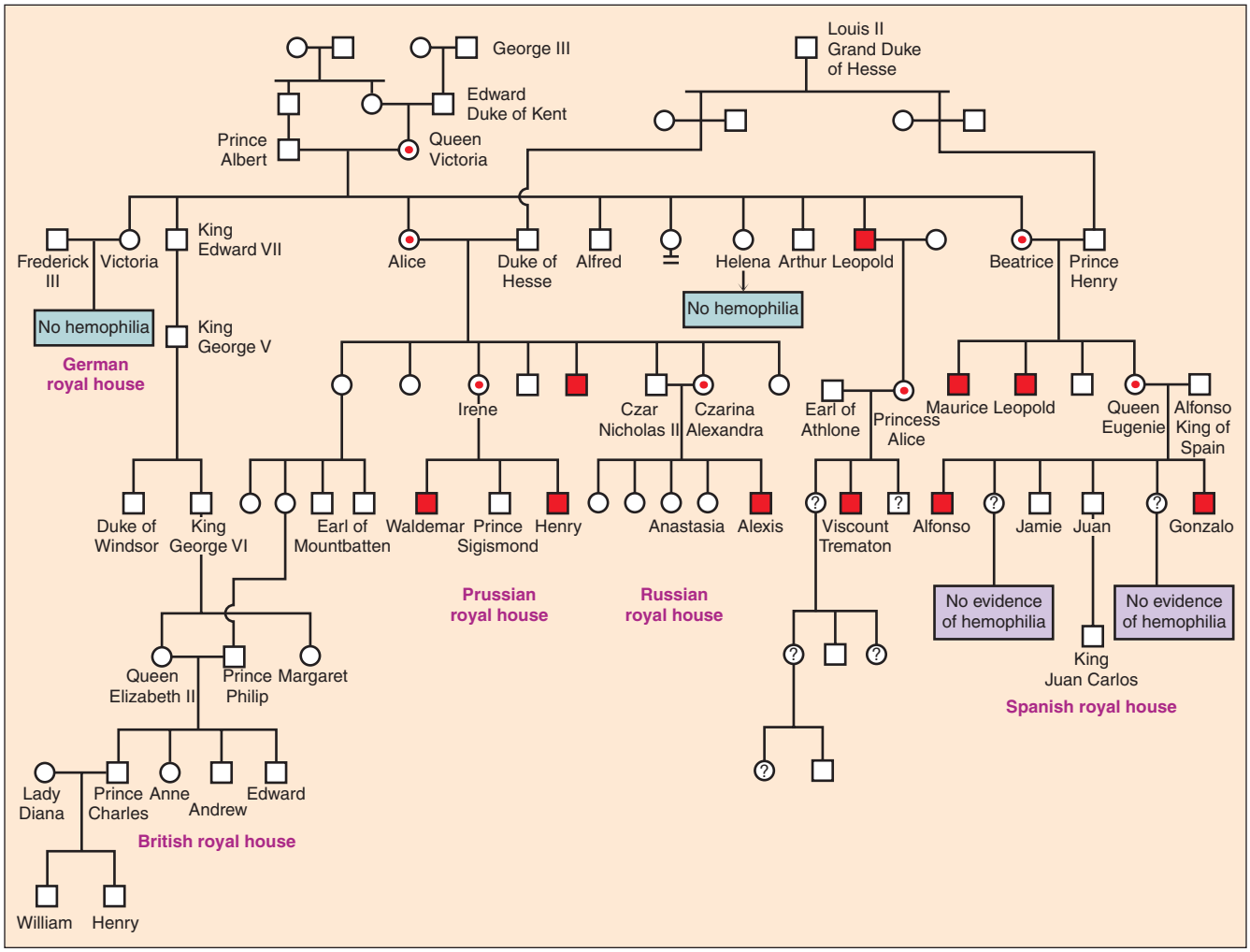


FIG 6.20 Pedigree chart for the transmission of the X-linked disease hemophilia A in the royal families of Europe.

Anticipation

Some genetic disorders deviate from typical mendelian inheritance patterns by becoming more severe and/or having earlier onset as the disorder is transmitted to future generations. This phenomenon, called *anticipation*, is most commonly seen in trinucleotide repeat disorders. Fragile X syndrome is a prototypical example of a trinucleotide repeat disorder that displays anticipation. Fragile X syndrome is the most common cause of familial mental retardation, exhibiting a prevalence rate of about 1 in 4000 males (1 in 8000 females). Affected individuals have an elongated (or expanded) repeat sequence composed of repeating CGG triplets in the *FMR1* gene on chromosome X. The protein normally produced by *FMR1* is crucial to the development and function of cerebral neurons. Normal individuals have an average of 29 CGG repeats at this gene locus. Persons with fragile X syndrome have significantly more: 200 to 2000 triplet repeats constitute a full mutation and produce the syndrome phenotype. The long repeat sequence may be visible on cytogenetic studies (karyotypes) as a characteristic narrowed segment on the long arm of the X chromosome. Persons who have an intermediate number of repeats (60 to 200) are said to have a *premutation.* Although they themselves do not have features of fragile X syndrome, they are at significant risk for producing affected offspring. The premutation is unstable and predisposed to expansion during oogenesis, but much less so during spermatogenesis. Anticipation, that is, expansion of the premutation as it is passed through the female lineage, results in disease manifestation in offspring at an earlier age or with increasing severity with each succeeding generation. Because the mutation is on the X chromosome, males with fragile X syndrome tend to be more severely affected, apparently because the presence of a second X chromosome in females moderates the clinical symptoms. Other trinucleotide repeat disorders, including Huntington disease and myoclonic dystrophy, also display anticipation.

Mitochondrial Gene Mutations

Nearly all genes are located in the cell nucleus; however, mitochondria in the cell cytoplasm have their own DNA (mitochondrial DNA or mtDNA), which includes several genes critical to energy production. At fertilization, mitochondria in the sperm (which are located near the tail) are lost, so mitochondrial genes are of maternal origin. Mothers transmit mitochondrial DNA to both sons and daughters, but only daughters transmit the mitochondrial genes to their offspring. Mitochondrial DNA is much more prone to mutation than nuclear DNA. Mitochondrial DNA codes for enzymes involved in oxidative phosphorylation and electron transport chain reactions, and mutations tend to cause dysfunction in tissues with high utilization of ATP such as nerve, muscle, kidney, and liver cells.

Genomic Imprinting

The concept of genomic imprinting challenges the long-held belief that the parental origin of a gene does not make any difference to the cells that inherit the gene. Genomic imprinting is a process whereby maternal and paternal genes are marked differentially by chemical tags that alter gene expression, such as silencing the gene. This chemical marking of genes is an example of an epigenetic effect, which is an alteration to DNA that does not involve a change in the sequence of a gene. Epigenetic effects, which are thought to play a critical role in gene expression, are both heritable and reversible.

Genomic imprinting can be illustrated by considering two very different syndromes, which at first glance appeared to be a result of the same chromosomal defect. Prader–Willi syndrome and Angelman syndrome both result from a fairly large deletion at the same location on chromosome 15 that includes several genes. Prader–Willi syndrome is characterized by mental retardation, short stature, obesity, poor muscle tone, and hypogonadism. Patients with Angelman syndrome are also mentally retarded, but they have ataxia and seizures and tend to laugh inappropriately. The fact that two different syndromes result from the same mutation was puzzling until it was discovered that Prader–Willi is associated with a deletion on the paternally derived chromosome 15, whereas the Angelman syndrome is associated the same deletion on the maternally derived chromosome 15. Both the "Angelman gene" and the "Prader–Willi gene" are normally present in the region of the deletion. The "Angelman gene" is normally active on a maternal chromosome and silent (due to imprinting) on the paternal chromosome, whereas the "Prader–Willi gene" is normally active on only the paternal chromosome and silent on the maternal chromosome. Thus normal cells have only one functional copy of each gene. The disorders arise when there is a deletion of chromosome 15 involving those genes, and a different syndrome arises according to whether the deletion is on the maternal or paternal chromosome (Fig. 6.21). These findings imply that the cell is not blind to the parental origin of chromosomes and that homologous chromosomes may be marked and function differently within the cell.

POLYGENIC AND MULTIFACTORIAL DISORDERS

Most human traits develop in response to more than one gene; such traits are called *polygenic*. Environmental influence also has a role in gene expression; traits and disorders influenced by multiple genes as well as environmental factors are called *multifactorial*. Examples of multifactorial traits are height, weight, and intelligence. Most common health conditions and disorders, including obesity, hypertension, diabetes and depression, are also multifactorial. Multifactorial traits do not follow clear-cut modes of inheritance but do tend to "run in families." Characteristics determined by multifactorial inheritance tend to have a range of expression in the population and demonstrate a "bell curve" distribution. They are thought to be produced by the interaction of several genes, each contributing a small additive effect and modulated by environmental influences, such as diet, activity, stress, and environmental exposures. Most multifactorial disorders present a range of severity, although a few disorders are either present or absent. In the latter case, it may be that a certain threshold number of defective genes must be inherited before the disease is expressed.

It is extremely difficult to predict the risk of occurrence of multifactorial disorders based on family history. Based on observing inheritance in the population, however, empirical estimates have been derived. Recurrence risk is specific for each disorder and varies among different populations, but in general the risk is much less than that found in mendelian disorders. As an example, neural tube defects, which are

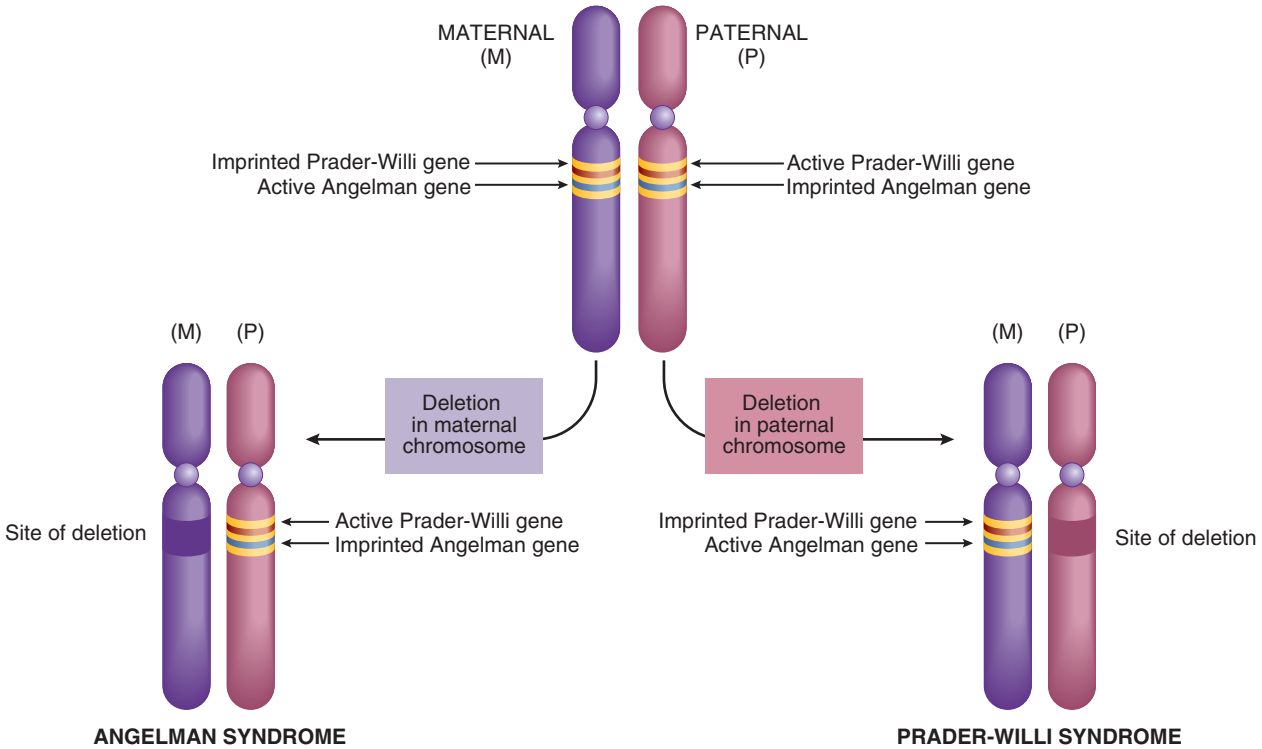


FIG 6.21 Angelman and Prader–Willi syndromes are examples of genetic imprinting, where the location of a mutation on the maternal or paternal homologous chromosome produces a different outcome.

multifactorial, occur in 2% to 3% of siblings of affected individuals. Based on such a recurrence risk, parents of one affected child have a 2% to 3% chance of bearing a second affected child.

In contrast to single-gene and chromosomal abnormalities, which are rare, multifactorial disorders are very common. High blood pressure, atherosclerosis, cancer, diabetes mellitus, cleft lip, and several forms of congenital heart defects are transmitted by multifactorial inheritance. This list is destined to grow as knowledge of the role of genetic mechanisms in cellular function and disease expands.

KEY POINTS

- Genetic disorders are of three general types: chromosomal aberrations, single-gene disorders, and polygenic/multifactorial disorders.
- Chromosome disorders result from an abnormality in number or structure. The presence of only one chromosome of a homologous pair is termed *monosomy* (e.g., Turner syndrome), and the presence of an excessive number of chromosomes is called *polysomy* (e.g., Down syndrome). Abnormal rearrangement of portions of the chromosomes (translocation, inversion, deletion, duplication) can result in loss or unusual expression of genes.
- Single-gene disorders result from mutations that alter the nucleotide sequence of one particular gene. Mendelian disorders are transmitted predictably in autosomal-dominant (e.g., Huntington disease), autosomal-recessive (e.g., cystic fibrosis), and sex-linked (e.g., hemophilia) patterns.
- Some single-gene disorders have unusual transmission patterns, which violate Mendel's laws. These include expanded triplet repeat mutations, mitochondrial DNA mutations, and genomic imprinting disorders.
- Multifactorial disorders are very common and result from the interaction of multiple genes and environmental influences. Disorders such as high blood pressure, cancer, and diabetes are multifactorial.

TABLE 6.5 Causes of Congenital Malformations in Humans

Cause	Malformed Live Births (%)
Genetic	
Chromosomal aberrations	10–15
Mendelian inheritance	2–10
Environmental	
Maternal/placental infections (e.g., rubella, toxoplasmosis, syphilis, cytomegalovirus, human immunodeficiency virus)	2–3
Maternal disease states (e.g., diabetes, phenylketonuria, endocrinopathies)	6–8
Drugs and chemicals (e.g., alcohol, androgens, folic acid antagonists, phenytoin, thalidomide, warfarin, 13-*cis*-retinoic acid)	≈1
Irradiation	≈1
Multifactorial	20–25
Unknown	40–60

Adapted from Stevenson RE et al, editors: *Human malformations and related anomalies,* New York, 1993, Oxford University Press, p. 115; Kumar V et al: *Robbins & Cotran pathologic basis of disease,* ed 9, Philadelphia, 2015, Saunders, p. 454.

ENVIRONMENTALLY INDUCED CONGENITAL DISORDERS

Adverse influences during intrauterine life are a significant cause of errors in fetal development that result in congenital malformations. The study of abnormal physical development (morphogenesis) is called *dysmorphology*. Most malformations are associated with genetic causes; however, exposure to certain environmental agents, such as chemicals, radiation, maternal disease states, and viruses, may adversely affect the developing fetus (Table 6.5). Agents that cause congenital malformation are called *teratogens*. Many substances, including some medications, are thought to have teratogenic potential, based on experiments in animals, but few are proved in humans. Exposure to a known teratogen may, but need not, result in a congenital malformation. Susceptibility to a teratogen depends on the amount of exposure, the developmental stage of the fetus when exposed, the prior condition of the mother, and the genetic predisposition of the fetus.

Periods of Fetal Vulnerability

The timing of the exposure to a teratogen greatly influences fetal susceptibility and the resulting type of malformation. The intrauterine development of humans can be divided into two stages: (1) the embryonic period, which extends from conception to 9 weeks of development, and (2) the fetal period, which continues until birth. Before the third week of gestation, exposure to a teratogen generally either damages so few cells that the embryo develops normally, or damages so many cells that the embryo cannot survive and spontaneous abortion occurs. Between the third and ninth weeks of gestation the embryo is very vulnerable to teratogenesis, with the fourth and fifth weeks being the time of peak susceptibility. Organ development (organogenesis) occurs during this period, and this complex process is very sensitive to injury, regardless of the cause. Each organ has a critical period during which it is most vulnerable to malformation (Fig. 6.22). Unfortunately, an embryo may be exposed to teratogens during the vulnerable period because the mother does not yet realize she is pregnant. During the fetal period, from 3 to 9 months, organs undergo further growth and maturation, and susceptibility to errors of development is significantly less. Fetal insults occurring after the third month are more likely to result in growth retardation or injury to normally formed organs.

Teratogenic Agents

The teratogenic potential of many agents is unknown. Several chemicals, some infections, and large doses of radiation are definitely associated with a higher risk of congenital disorders. In general, teratogens cause errors in morphogenesis by interfering with cell proliferation, migration, or differentiation. The specific mechanisms of action of most teratogens are unknown.

Chemicals and Drugs

The list of proven teratogenic chemicals and drugs includes thalidomide, alcohol, anticonvulsants, warfarin, folate antagonists, androgenic hormones, angiotensin-converting enzyme inhibitors, and organic mercury. Almost no drugs or chemicals are considered totally safe, and the current trend is to discourage pregnant women from using *any* drugs or chemicals. A classification system for determining relative risk of medications in pregnancy has been developed (Table 6.6). Two agents, thalidomide and alcohol, illustrate the teratogenic potential of chemicals.

In the 1960s, an increase in the incidence of congenital limb deformities was traced to maternal use of thalidomide, a tranquilizer, during

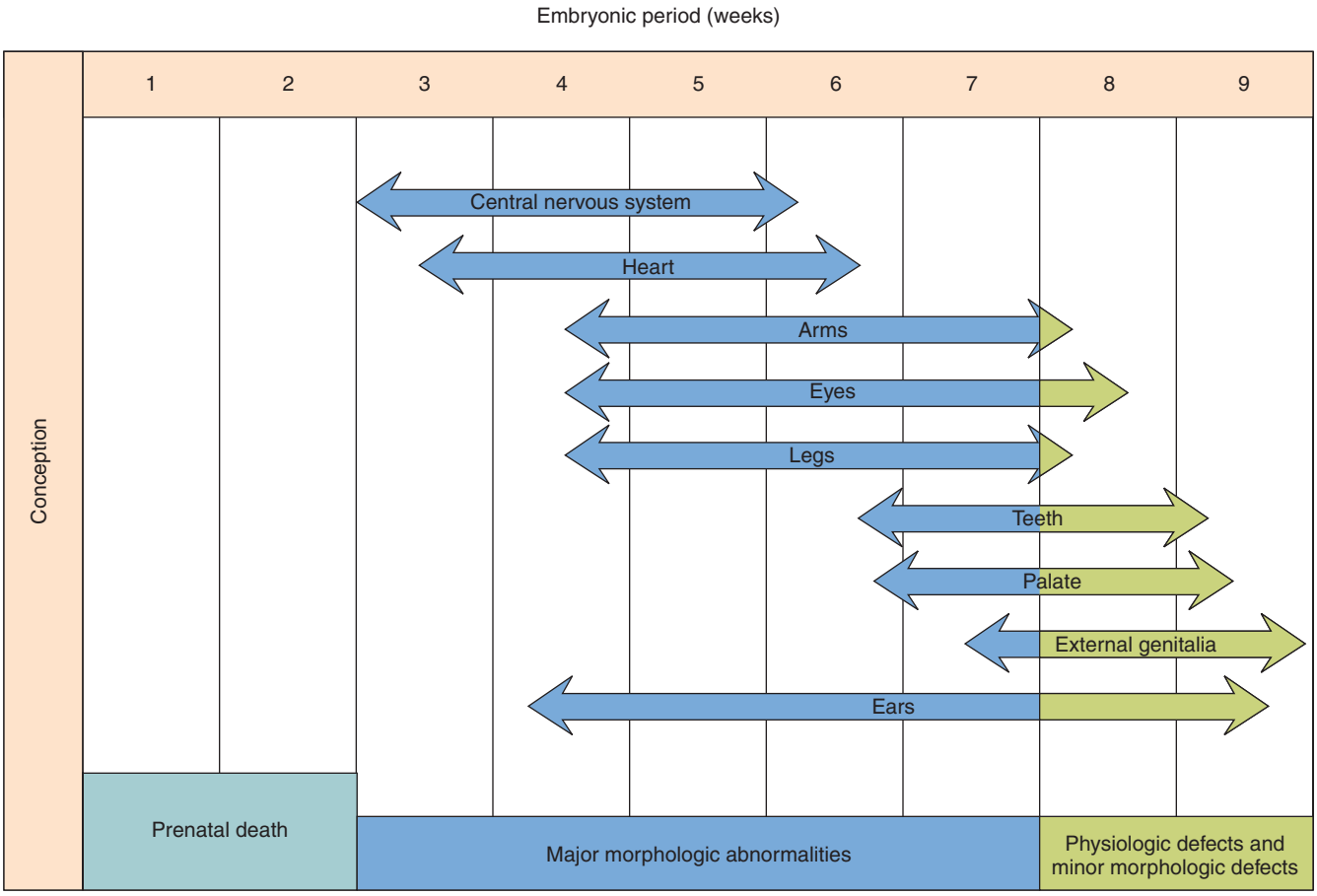


FIG 6.22 Vulnerable periods of fetal organ development.

early pregnancy. Exposure during the vulnerable period (20 to 35 days after conception) was associated with a very high risk of fetal malformation. Typically, the arms were short and flipperlike, although deformities ranged from mild abnormalities of the digits to complete absence of the limbs. Damage to other structures, particularly the ears and heart, also occurred. Thalidomide is one of the most potent teratogens known.

Prenatal exposure to alcohol is known to be associated with a wide spectrum of congenital conditions referred to as *fetal alcohol spectrum disorder* (FASD). Fetal alcohol syndrome (FAS) represents the severe end of the spectrum. It is estimated that between 0.2 and 1.5 of every 1000 newborns suffer from FAS. The prevalence of FASD is estimated to be much higher, reaching 1% and much higher in populations with high alcohol use among women of childbearing age. Affected infants display characteristic facial abnormalities and suffer from growth retardation, developmental delay, and learning and behavioral problems, with alterations to the developing brain resulting in the most devastating deficits. Multiple mechanisms for alcohol teratogenesis have been suggested. Data are insufficient to determine what, if any, level of alcohol intake during pregnancy is safe. It is clear that factors other than the absolute amount of alcohol intake during pregnancy are important in determining risk of FAS. Complete abstinence from alcohol during pregnancy is recommended.

Infectious Agents

A number of perinatal infections have been implicated in the development of congenital malformations. Certain viral infections appear to carry the greatest threat, although protozoa and bacteria have also been implicated. As with other teratogens, the gestational age of the fetus at the time of infection is critically important. Perhaps the best known viral teratogen is rubella. The risk period for rubella infection begins just before conception and extends to 20 weeks' gestation, after which the virus rarely crosses the placenta. Rubella-induced defects vary but typically include cataracts, deafness, and heart defects. Several other organisms cause a similar constellation of congenital defects; therefore the acronym TORCH was developed to alert clinicians to the potential teratogenicity of these infections. TORCH stands for **t**oxoplasmosis, **o**thers, **r**ubella, **c**ytomegalovirus, **h**erpes. The major features of the TORCH complex are shown in Fig. 6.23. The category of "others" includes several less frequently seen causes: hepatitis B, coxsackievirus B, mumps, poliovirus, zika virus, and others. All microorganisms of the TORCH complex are able to cross the placenta and infect the fetus.

Toxoplasmosis is a protozoal infection that can be contracted from ingestion of raw or undercooked meat and from contact with cat feces. Cytomegalovirus and herpes simplex virus are generally transmitted to the fetus by chronic carrier mothers. Cytomegalovirus and herpes simplex virus often colonize in the genital area of the mother. Infants who escape infection in utero may still acquire the virus as they pass through the birth canal (see Chapter 34).

Radiation

In addition to being mutagenic, radiation is teratogenic. The teratogenic potential of radiation became apparent from the increased incidence of congenital malformations in children born to women who underwent irradiation of the cervix for cancer and in the children of atomic bomb victims in World War II. It is not known if lower levels of radiation, such as those used in diagnostic x-rays, are teratogenic. It is generally

TABLE 6.6 Pregnancy Categories for Medication Administration

Category	Interpretation
A	Adequate and well-controlled studies in pregnant women have not shown an increased risk of fetal abnormalities to the fetus in any trimester of pregnancy.
B	Animal studies have revealed no evidence of harm to the fetus; however, there are no adequate and well-controlled studies in pregnant women. **OR** Animal studies have shown an adverse effect, but adequate and well-controlled studies in pregnant women have failed to demonstrate risk to the fetus in any trimester.
C	Animal studies have shown an adverse effect, and there are no adequate and well-controlled studies in pregnant women. **OR** No animal studies have been conducted, and there are no adequate and well-controlled studies in pregnant women.
D	Adequate and well-controlled or observational studies in pregnant women have demonstrated risk to fetus. However, benefits of therapy may outweigh potential risk. For example, the drug may be acceptable if needed in a life-threatening situation or serious disease for which safer drugs cannot be used or are ineffective.
X	Adequate and well-controlled or observational studies in animals or pregnant women have demonstrated positive evidence of fetal abnormalities or risks. Use of the product is contraindicated in women who are or may become pregnant.

recommended that pregnant women avoid diagnostic x-rays or be provided appropriate lead shielding if x-rays are required.

Other Disorders of Infancy

An infant may be afflicted with a variety of problems at birth that do not fall into the category of genetic or developmental malformations. These problems generally arise later in uterine life and often involve mechanical factors or problems with the health of the mother and placenta. For example, babies with low birth weight or immaturity at birth may have difficulty breathing and taking in adequate nutrition. Interruption of the placental oxygen supply because of maternal hemorrhage, sedation, or blood incompatibility may result in fetal brain injury. A difficult labor and delivery may result in a variety of injuries during the birth process. The details of these disorders of infancy and childhood may be found in specialized texts.

KEY POINTS

- Environmental factors that adversely affect the developing fetus are called *teratogens*. Exposure to teratogens is particularly dangerous during the third to ninth weeks of gestation.
- Known teratogens include chemicals and drugs, infections, and radiation. The teratogenic potential of many chemicals and drugs is unknown, so pregnant women are usually advised to avoid all drugs if possible.
- Of the infectious agents, viruses are the most teratogenic, particularly organisms of the TORCH variety (toxoplasmosis, others, rubella, cytomegalovirus, herpes virus).

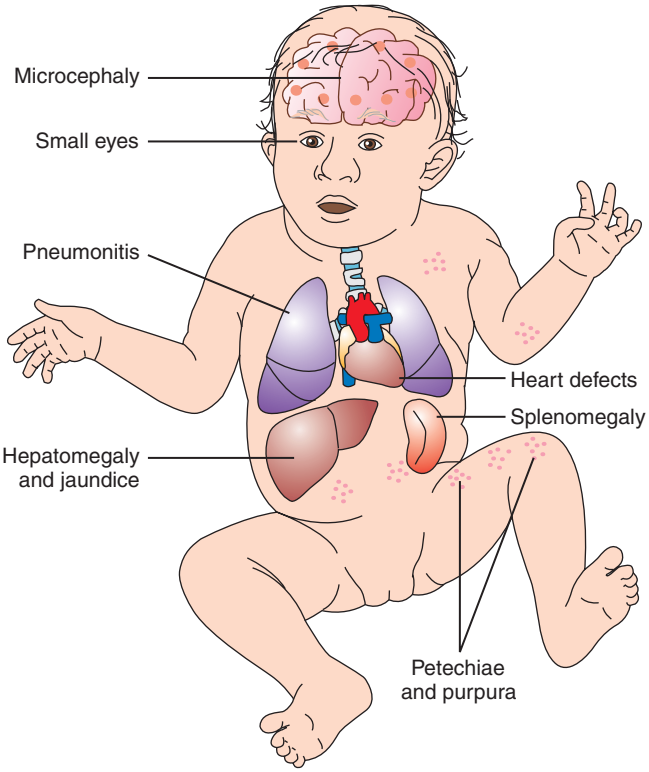


FIG 6.23 Major clinical findings in the TORCH (toxoplasmosis, others, rubella, cytomegalovirus, herpes) complex of infective congenital disorders.

DIAGNOSIS, COUNSELING, AND GENE THERAPY

In recent years, the ability to diagnose and manage genetic and developmental disorders has improved dramatically. Although pedigree analysis continues to be an important method for identifying at-risk individuals, for a number of disorders it is now possible to determine whether parents carry defective genes or if a particular fetus is afflicted. Currently, the ability to detect genetic mutations far exceeds the ability to offer definitive genetic treatment, triggering many ethical concerns. Unfortunately, many individuals at risk for transmitting recessive genetic diseases are not identified until the birth of an affected child. Genetic counseling and prenatal assessment then become extremely important in assisting the family in regard to future pregnancies.

Prenatal Diagnosis and Counseling

A number of conditions are associated with a higher risk of congenital anomalies and are indications for instituting counseling and prenatal diagnostic examination. These conditions include (1) a maternal age of 35 years or greater; (2) a history of previously bearing a child with a chromosomal disorder (such as trisomy 21); (3) a known family history of X-linked disorders; (4) a family history of inborn errors of metabolism; (5) the occurrence of neural tube anomalies in a previous pregnancy; and (6) awareness that the mother is a known carrier of a recessive genetic disorder. As diagnostic methods become more cost effective, general screening for other risk factors may be undertaken. Screening for genetic disorders may utilize ultrasound or maternal blood, but prenatal diagnostic testing usually requires invasive procedures such as amniocentesis, chorionic villus biopsy, or (rarely) umbilical cord blood sampling. Postnatal genetic analysis is usually done on peripheral blood samples of lymphocytes.

Ultrasound and amniocentesis are mainstays of prenatal diagnostic examination. Ultrasound is a noninvasive procedure that uses sound waves to produce a reflected image of the fetus. It is commonly used to determine gestational age, fetal position, and placental location. Ultrasound is also useful in detecting visible congenital anomalies such as spina bifida (neural tube defect); heart defects; and malformations of the face, head, body, and limbs.

Amniocentesis may be performed to determine genetic and developmental disorders not detectable by ultrasound. This invasive procedure is associated with a small but real risk of complications, including pregnancy loss. During amniocentesis, a needle is inserted through the abdomen or vagina and into the uterus. A sample of amniotic fluid containing skin cells shed by the fetus is removed for analysis. The amniotic fluid can be analyzed for abnormal levels of certain substances secreted by the fetus, such as α-fetoprotein, which may indicate neural tube defects. Fetal cells are also present in amniotic fluid and can be cultured and subjected to biochemical, chromosomal, and genetic analysis. Because amniocentesis is not recommended until at least 15 weeks' gestation, it does not provide information until relatively late in the pregnancy. Chorionic villus sampling (CVS) involves the removal of a piece of tissue directly from the chorion (the outer membrane of the fetal sac). It can be performed after about 9 weeks' gestation. Like amniocentesis, CVS is an invasive procedure that in rare instances results in pregnancy complication or loss.

Fetal cells obtained via either amniocentesis or CVS contain small amounts of DNA that can then be amplified into a larger quantity by a process called *polymerase chain reaction (PCR).* This DNA can then be analyzed by a variety of methods to determine the genetic sequence, the presence or absence of particular genes, or the presence of mutations. The fetal genome can be scanned for particular genes by exposing the processed DNA to microchips that are coated with thousands of genetic sequence probes in specific locations. Complementary base pairing between the probes fixed on the microchip and the DNA being tested is used to detect the presence of particular gene sequences.

In rare cases, a laparoscopic procedure may be performed to directly visualize the embryo or fetus. A fetoscope can be inserted through the cervix and into the uterus to identify and in some instances attempt to correct structural anomalies. The early diagnosis of congenital disorders allows a greater number of treatment options. Some disorders can be managed in utero; others may require early delivery, immediate surgery, or cesarean section to minimize fetal trauma. Early warning of fetal difficulties allows parents time to prepare emotionally for the birth of the child. In some instances, termination of the pregnancy may be the treatment of choice.

Genetic Analysis and Therapy

An exciting promise in the genome era of health care remains on the horizon—the treatment of genetic disease by replacing a defective gene with a normal, healthy gene. The concept is simple, and clinical trials have been conducted to treat a number of genetic disorders. The first federally approved gene therapy procedure was performed in 1990 to treat a child who suffered from a particular type of *severe combined immunodeficiency* (SCID) by introducing a functional gene for the enzyme adenosine deaminase. In the past, children who suffered from SCID had severely compromised immune systems and generally died from overwhelming infections unless their environment was strictly controlled. Use of gene therapy has shown some success in improving immune function and allowing these children to live in the outside world, although the safety of gene therapy is a continuing challenge.

Gene therapy has the potential for alleviating human suffering by curing genetic diseases, but it is accompanied by a number of moral and ethical dilemmas. Tampering with the human gene pool could have serious implications for human evolution. There is also the potential for using the technology to create "new and improved" human beings or human clones.

Recombinant DNA Technology

Over the past 30 years, DNA has gone from being the most difficult cellular molecule to study to being among the easiest. The great advances in molecular genetics during this time are due to plummeting costs of DNA sequencing and the development of recombinant DNA technologies. It is now possible to select a specific region of DNA, produce unlimited copies of it, determine its nucleotide sequence, use it to make unlimited quantities of a desired protein, or alter its DNA sequence at will (genetic engineering), and reinsert it into a living cell. Scientists can decipher the nucleotide sequence of an individual human genome, create DNA probes to explore an individual's genetic makeup for specific mutations, mass-produce therapeutic proteins and vaccines, and hope someday to cure genetic disorders by replacing mutated genes with normally functioning ones.

Recombinant DNA technology comprises a number of techniques, the most important of which are briefly described here:

- The long, difficult-to-handle DNA strands are more easily studied if cut into smaller pieces. This is accomplished by using restriction enzymes that cleave DNA at specific sites. The resulting pieces can then be separated by electrophoresis according to their size. A section of DNA can be collected and efficiently sequenced by automated means.
- Nucleic acid hybridization techniques take advantage of the natural tendency for DNA and RNA to find and bind to a complementary nucleotide sequence. A labeled piece of DNA or RNA can therefore be used to search for, or "probe" for, its complementary sequence among the many millions of sequences in a cell or cell extract. For example, in a fluorescence in situ hybridization (FISH) assay, a probe for a specific site on a chromosome is attached to a fluorescent label and incubated with a cell. The fluorescence is then examined to identify the location and number of copies of the particular chromosome sequence (Fig. 6.24). Without the hybridization technique, finding a desired gene among the 3 billion base pairs in the human genome could take many years of intense effort, like finding the proverbial needle in a haystack. Hybridization is also the method behind the "gene chip assay" technology. As an outcome of the Human Genome Project, the DNA sequences for the thousands of human genes and common mutations have been identified. Specific DNA sequence probes for these genes can be synthesized and attached to a fixed position on a plate (microchip). The DNA of interest can then be exposed to the chip, and the probes will preferentially bind to DNA segments having complementary bases. The position of the probe and the degree of DNA binding can then be analyzed by computer to produce a specific genetic analysis of gene mutations.
- DNA cloning is the technique used to produce many identical copies of a DNA sequence containing a gene of interest. The availability of large quantities of a purified gene sequence makes study and gene manipulation possible. A number of different techniques can be used to clone DNA. The PCR technique is very efficient if the DNA sequence is already partially known. Basically, the DNA sequence of interest is mixed with special DNA polymerases that use the DNA sequence as a template to produce double-stranded DNA. Each DNA thus produced can in turn act as a template for production of another DNA. Large quantities can be produced very rapidly by PCR. The DNA can also be cloned by inserting it into bacteria by use of a viral or plasmid vector. Bacteria that incorporate the desired gene are identified by hybridization with a labeled probe. The desired

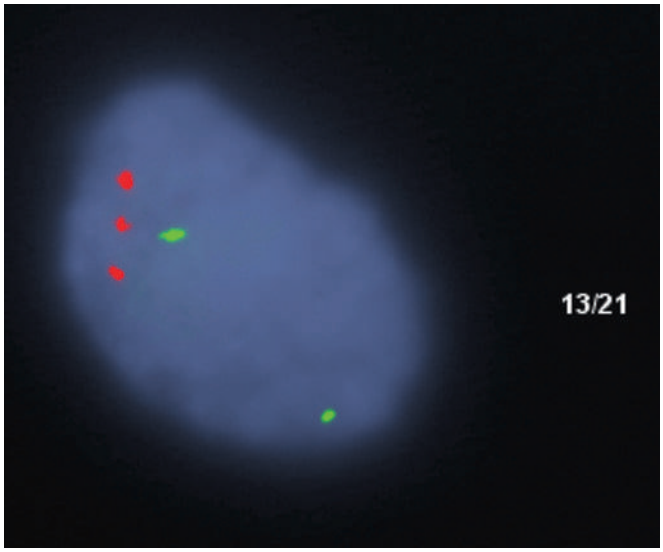


FIG 6.24 Fluorescence in situ hybridization assay showing an interphase nucleus. The red probe is hybridized to chromosome 21 and the green probe is hybridized to chromosome 13. Three copies of chromosome 21 are identified, confirming the diagnosis of trisomy 21. (From Kumar V et al: *Robbins & Cotran pathologic basis of disease,* ed 9, Philadelphia, 2015, Saunders, p. 161. Photograph courtesy Dr. Stuart Schwartz, Department of Pathology, University of Chicago, Chicago, IL.)

bacteria then are allowed to proliferate, making a copy of the DNA sequence along with their own genome with each cell division.

- Genetic engineering refers to a process whereby a gene of interest is altered from its original form. The altered (mutated) gene can be reintroduced into a cell to disclose its effect on cell function and thus elucidate the normal function of the original gene and its protein product. Genetic engineering has been applied to plants to increase their value as food crops. Genetically engineered bacteria are the source of more than 90 human proteins created in bacteria using recombinant DNA (rDNA) technology for pharmaceutical use; examples are human insulin, epoetin, and interferon.

Gene therapy relies heavily on these techniques to facilitate identification of genetic mutations, study of gene function, and development of methods to repair or replace mutated genes. Many more applications of rDNA technology will become apparent as research on the genetic basis of human function and disease proceeds.

KEY POINTS

- Risk factors that indicate the need for prenatal diagnostic examination and counseling include advanced maternal age (older than 35 years), a family history of genetic disorders, and the previous birth of a child with chromosomal or neural tube defects.
- Ultrasound, amniocentesis, and chorionic villus sampling (CVS) are the mainstays of prenatal assessment for genetic disorders. Technology for examining fetal DNA isolated from maternal blood has allowed some prenatal screening tests to be performed noninvasively.
- DNA sequences that are complementary to a gene of interest can be synthesized and used to probe a genome to determine whether and where the gene is present. These hybridization techniques make screening for genetic disorders relatively fast and simple.
- Gene therapy is the treatment of genetic disease by replacing defective genes with normal genes. Gene therapy is possible because of the advances attained in rDNA technology over the past 30 years.

SUMMARY

Genetic and developmental disorders are responsible for a number of congenital malformations. Congenital disorders are caused by genetic and environmental factors that disrupt normal fetal development. Genetic disorders are classified as (1) chromosomal alterations, including structural and numeric abnormalities; (2) mendelian disorders, including autosomal-dominant, autosomal-recessive, and X-linked disorders; (3) nonmendelian single-gene disorders, including triplet repeats, mitochondrial gene defects, and genetic imprinting disorders; and (4) polygenic or multifactorial disorders. Known environmental teratogens include radiation, infectious organisms, and various chemicals and drugs. The embryo is particularly susceptible to teratogens during the period of organogenesis, which extends from the third to the ninth week of gestation. Pedigree analysis, ultrasound, amniocentesis, and chorionic villus biopsy may provide helpful information regarding genetic risk and the prenatal condition of at-risk infants. DNA sequencing of normal and mutated genes has made it possible to efficiently screen for genetic disorders. DNA can be manipulated to produce human protein pharmaceuticals, and clinical trials to test gene therapies for a variety of genetic diseases have been conducted.

RESOURCES

Principles of Inheritance and Genetic Disorders

Barry PJ, Jones AM: New and emerging treatments for cystic fibrosis. *Drugs* 75:1165–1175, 2015. doi:10.1007/s40265-015-0424-8.

Bateson W: *Mendel's principles of heredity*, London, 1902, Cambridge University Press.

Down JHL: Observations on an ethnic classification of idiots. *Clin Lect Rep London Hosp* 3:259–262, 1866.

Jackson JM, Crider KS, Cragan JD, et al: Frequency of prenatal cytogenetic diagnosis and pregnancy outcomes by maternal race-ethnicity, and the effect on the prevalence of trisomy 21, Metropolitan Atlanta, 1996-2005. *Am J Med Genet* 164A:70–77, 2014. doi:10.1002/ajmg.a.36247.

Jorde LB, Carey JC, Bamshad MJ: *Medical genetics*, ed 5, Philadelphia, 2016, Elsevier.

Kinch MS: An overview of FDA-approved biologics medicines. *Drug Discov Today* 20(4):393–398, 2015.

Koolwal A, Singh A: Ivacaftor: a novel mutation modulating drug. *J Clin Diagn Res* 8(22):SE1–SE5, 2014.

Kumar V, Abbas AK, Aster JC: *Pathologic basis of disease*, ed 9, Philadelphia, 2015, Elsevier Saunders.

Kumar S, Tana A, Shankar A: Cystic fibrosis—what are the prospects for a cure? *Eur J Intern Med* 25:803–807, 2014. doi:10.1016/j.ejim.2014.09.018.

Lewis R: *Human genetics: Concepts and applications*, ed 11, New York, 2015, McGraw Hill.

Lewis RA: Oculocutaneous albinism Type 1. In Pagon RA, Adam MP, Ardinger HH, et al, editors: *GeneReviews [Internet]*, Seattle (WA), 2000 Jan 19 [Updated 2013 May 16]. University of Washington, Seattle, pp 1993–2015. Available from: http://www.ncbi.nlm.nih.gov/books/NBK1166/.

Macpherson JL, Rasko JEJ: Clinical potential of gene therapy: towards meeting the demand. *Intern Med* 44:224–233, 2014. doi:10.1111/imj.12366.

Senturias YSN: Fetal alcohol spectrum disorders: an overview for pediatric and adolescent care providers. *Curr Probl Pediatr Adolesc Health Care* 4(44):74–81, 2014. doi:10.1016/j.cppeds.2013.12.012.

Turnpenny P, Ellard S: *Emery's elements of medical genetics*, ed 14, Philadelphia, 2012, Churchill-Livingstone.

Waldorf KMA, McAdams RM: Influence of infection during pregnancy on fetal development. *Reproduction* 146(5):R151–R162, 2013. doi:10.1530/rep-13-0232.

Warby SC, Graham RK, Hayden MR: Huntington disease. In Pagon RA, Adam MP, Ardinger HH, et al, editors: *GeneReviews [Internet]*, Seattle (WA), 1998 Oct 23 [Updated 2014 Dec 11]. University of Washington, Seattle, pp 1993–2015. Available from: http://www.ncbi.nlm.nih.gov/books/NBK1305/.

7

Neoplasia

Jacquelyn L. Banasik

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- How do neoplastic cells differ from normal cells?
- In what ways do benign and malignant tumors differ?
- How might overexpression of proto-oncogenes lead to abnormal cellular proliferation?
- How might underexpression of tumor suppressor genes lead to abnormal cellular proliferation?
- What properties are gained during tumor progression that contribute to malignant behavior and metastasis?
- How are tumor grading and staging used to characterize cancers and guide the selection of cancer therapies?
- How might lifestyle and carcinogen exposure contribute to cancer risk?
- What effects does cancer have on the body?
- What options are available for cancer treatment?

CHAPTER OUTLINE

Benign Versus Malignant Growth, 118
Characteristics of Benign and Malignant Tumors, 118
Tumor Terminology, 118
The Malignant Phenotype, 118
Epidemiology and Cancer Risk Factors, 120
Tobacco Use, 120
Nutrition, 120
Antioxidants, 123
Genetic Mechanisms of Cancer, 123
Proto-Oncogenes, 124
Growth Factors (Mitogens), 124
Growth Factor Receptors, 124
Cytoplasmic Signaling Pathways, 124
Transcription Factors, 125
From Proto-Oncogene to Oncogene, 125
Tumor Suppressor Genes, 127
The Rb Gene, 127
The P53 Gene, 128
BRCA1 and BRCA2 Genes, 128
Multistep Nature of Carcinogenesis, 129
Initiation, 129
Promotion, 131
Progression, 131
Metastasis, 132
Patterns of Spread, 132
Angiogenesis, 133
Grading and Staging of Tumors, 134
Effects of Cancer on the Body, 136
Cancer Therapy, 137
Surgery, 137
Radiation Therapy, 138
Drug Therapy, 138
Immunotherapy, 138
Gene and Molecular Therapy, 138
Stem Cell Transplantation, 139

Neoplasia means "new growth." In common use, the term implies an *abnormality* of cellular growth and may be used interchangeably with the term *tumor.* It is no surprise that the discovery of a tumor in an individual can evoke feelings of disbelief, anger, and dread. Characterization of the tumor cells is of critical importance to determine whether the tumor is benign or malignant. The term *cancer* is applied only to malignant neoplasms. The diagnosis of a benign growth is received with great relief inasmuch as the tumor is generally easily cured. The diagnosis of a malignant cancer, on the other hand, may involve months of intensive and often uncomfortable treatment with uncertain outcomes. Cancer remains the second-leading cause of death in the United States for both men and women.

It is increasingly clear that cancer is associated with *altered expression of cellular genes* that normally regulate cell proliferation and differentiation. A unified theory of cancer causation has emerged, and new methods for cancer therapy continue to be developed. Cancer is a complex,

multifaceted disorder, with each individual cancer having some unique properties. A better understanding of the molecular characteristics of individual cancers is encouraging the development of specific therapies that target each cancer's weaknesses.

BENIGN VERSUS MALIGNANT GROWTH

Characteristics of Benign and Malignant Tumors

The terms *benign* and *malignant* refer to the overall consequences of a tumor to the host. Generally, malignant tumors have the potential to kill the host if left untreated, whereas benign tumors do not. This difference is not strict because some benign tumors may be located in critical areas. For example, a benign tumor may be life threatening if it causes pressure on the brain or blocks an airway or blood vessel. Histologic examination of a tumor is the primary mode for determining its benign or malignant nature. Certain tumor characteristics have historically been shown to indicate malignant potential. Important considerations include localization of the tumor and determination of the degree of tumor cell differentiation.

Benign tumors do not invade adjacent tissue or spread to distant sites. Many benign tumors are encapsulated by connective tissue, which is an indication of strictly local growth. Any evidence that tumor cells have penetrated local tissues (invasiveness), lymphatics, or blood vessels suggests a malignant nature with potential to spread to distant sites (metastasize).

As a general rule, benign cells more closely resemble their tissue type of origin (e.g., skin, liver) than do malignant cells. The degree of tissue-specific differentiation has traditionally been used to predict malignant potential. A lack of differentiated features in a cancer cell is called *anaplasia,* and a greater degree of anaplasia is correlated with a more aggressively malignant tumor. Anaplasia is indicated by variation in cell size and shape within the tumor, enlarged nuclei, abnormal mitoses, and bizarre-looking giant cells (Fig. 7.1). Regardless of histologic appearance, invasion of local tissue or evidence of metastasis to distant sites confirms the diagnosis of malignancy.

Other differences between benign and malignant tumors have been noted (Table 7.1). Benign tumors generally grow more slowly, have little vascularity, rarely have necrotic areas, and often retain functions similar to those of the tissue of origin. Conversely, malignant tumors often grow rapidly and may initiate vessel growth in the tumor. They frequently have necrotic areas and are dysfunctional.

Tumor Terminology

General rules for the naming of tumors have been developed to indicate the tissue of origin and the benign or malignant nature of the tumor. The suffix *-oma* is used to indicate a benign tumor, whereas *carcinoma* and *sarcoma* are used to indicate malignant tumors. Carcinoma refers to malignant tumors of epithelial origin and sarcoma to malignant tumors of mesenchymal (nerve, bone, muscle) origin. Thus a benign tumor of glandular tissue would be called an *adenoma,* but a malignant tumor of the same tissue would be called an *adenocarcinoma* (Table 7.2). Some notable exceptions to the rules are lymphomas, hepatomas, and melanomas, which are all highly malignant despite their *-oma* suffix. Leukemia refers to a malignant growth of white blood cells. The great majority of human cancers (90%) are carcinomas from malignant transformation of epithelial cells.

The Malignant Phenotype

Cells growing in normal tissue have predictable relationships with neighboring cells. In a particular tissue, the rate of cell proliferation is precisely matched to the rate of cell death. Normal cells require constant reassurance in the form of survival signals from their environment that their continued existence is desirable, and they proliferate only when space is available and appropriate mitogen-stimulating signals are present.

TABLE 7.1 General Characteristics of Benign and Malignant Tumors

Characteristic	Benign	Malignant
Histology	Typical of tissue of origin Few mitoses	Anaplastic, with abnormal cell size and shape Many mitoses
Growth rate	Slow	Rapid
Localization/ metastasis	Strictly local, often encapsulated/no metastasis	Infiltrative/frequent metastases
Tumor necrosis	Rare	Common
Recurrence after treatment	Rare	Common
Prognosis	Good, unless in critical area	Poor if untreated

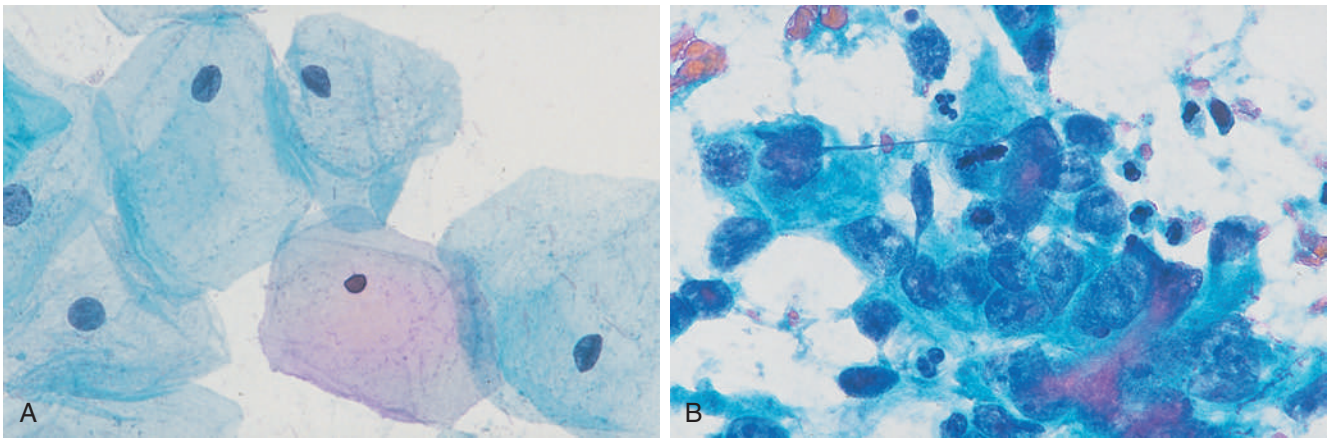


FIG 7.1 A, Normal Papanicolaou smear from the uterine cervix showing large, flat epithelial cells with small nuclei. **B,** Typical histologic appearance of anaplastic tumor cells showing variation in cell size and shape, with large, hyperchromic nuclei. (From Kumar V et al: *Robbins and Cotran pathologic basis of disease,* ed 9, Philadelphia, 2015, Saunders, p 1006. Courtesy Edmund S. Cibas, Brigham and Women's Hospital, Boston, MA.)

TABLE 7.2 Nomenclature for Neoplastic Diseases

Cell or Tissue of Origin	Benign	Malignant
Tumors of Epithelial Origin		
Squamous cells	Squamous cell papilloma	Squamous cell carcinoma
Basal cells	—	Basal cell carcinoma
Glandular or ductal epithelium	Adenoma	Adenocarcinoma
	Cystadenoma	Cystadenocarcinoma
Transitional cells	Transitional cell papilloma	Transitional cell carcinoma
Bile duct	Bile duct adenoma	Bile duct carcinoma (cholangiocarcinoma)
Liver cells	Hepatocellular adenoma	Hepatocellular carcinoma
Melanocytes	Nevus	Malignant melanoma
Renal epithelium	Renal tubular adenoma	Renal cell carcinoma
Skin adnexal glands		
Sweat glands	Sweat gland adenoma	Sweat gland carcinoma
Sebaceous glands	Sebaceous gland adenoma	Sebaceous gland carcinoma
Germ cells (testis and ovary)	—	Seminoma (dysgerminoma), embryonal carcinoma, yolk sac carcinoma
Tumors of Mesenchymal Origin		
Hematopoietic/lymphoid tissue	—	Leukemia, lymphoma, Hodgkin disease, multiple myeloma
Neural and retinal tissue		
Nerve sheath	Neurilemmoma, neurofibroma	Malignant peripheral nerve sheath tumor
Nerve cells	Ganglioneuroma	Neuroblastoma
Retinal cells (cones)	—	Retinoblastoma
Connective tissue		
Fibrous tissue	Fibromatosis (desmoid)	Fibrosarcoma
Fat	Lipoma	Liposarcoma
Bone	Osteoma	Osteogenic sarcoma
Cartilage	Chondroma	Chondrosarcoma
Muscle		
Smooth muscle	Leiomyoma	Leiomyosarcoma
Striated muscle	Rhabdomyoma	Rhabdomyosarcoma
Endothelial and related tissues		
Blood vessels	Hemangioma	Angiosarcoma
		Kaposi sarcoma
Lymph vessels	Lymphangioma	Lymphangiosarcoma
Synovium	—	Synovial sarcoma
Mesothelium	—	Malignant mesothelioma
Meninges	Meningioma	Malignant meningioma

Normal cells also respond to signals instructing them to actively destroy themselves in a process called *apoptosis* (see Chapter 4). Cancer cells, however, do not obey the rules; they have escaped the normal mechanisms of growth control. A number of antisocial properties develop in malignant cells that allow them to proliferate at the expense of other cells and tissues of the body. These abnormal behaviors can be summarized as follows:

- Cancer cells proliferate despite lack of growth-initiating signals from the environment.
- Cancer cells escape apoptotic signals and achieve a kind of immortality in that they are capable of unlimited replication.
- Cancer cells lose their differentiated features and contribute poorly or not at all to the function of their tissue.
- Cancer cells are genetically unstable and evolve by accumulating new mutations at a much faster rate than normal cells.
- Cancer cells invade their local tissue and overrun their neighbors.
- Perhaps worst of all, cancer cells gain the ability to migrate from their site of origin to colonize distant sites where they do not belong.

Most cancers are thought to arise from stem cells that are present in tissues. Tissue stem cells are capable of proliferation, entering the cell cycle to produce two daughter cells—with one cell retaining the original stem cell properties and the other becoming a more differentiated cell, but still capable of proliferation. Normally, the partially differentiated cells can undergo only a limited number of cell divisions before they permanently leave the cell cycle and become senescent. Either a stem cell or a partially differentiated cell has the potential to acquire the genetic mutations necessary to become malignant.

KEY POINTS

- Malignant tumors have the potential to kill the host, whereas benign tumors generally do not. The primary difference between malignant and benign tumors is the propensity of malignant tumors to invade adjacent tissue and spread to distant sites (metastasize).
- The suffix *-oma* is used to indicate a benign tumor (e.g., fibroma). *Carcinoma* and *sarcoma* are used to indicate malignancy (e.g., fibrosarcoma). Exceptions include melanomas, lymphomas, hepatomas, and leukemia, all of which are malignant.
- Malignant cells exhibit antisocial properties that allow them to ignore growth-controlling signals from the environment. Cancer cells proliferate excessively, become immortal, invade locally, and may travel to distant sites where they establish new colonies.

EPIDEMIOLOGY AND CANCER RISK FACTORS

Cancer accounts for approximately 25% of all deaths, which makes it the second-leading cause of death in the United States. Most cancer deaths (78%) occur in persons older than 55 years. The American Cancer Society (ACS) estimates that men have almost a 1 in 2 lifetime risk of developing cancer and women have slightly higher than a 1 in 3 risk. The 5-year relative survival rate for all cancers combined is about 68%. The 5-year survival rate does not distinguish between those who were cured and those who have relapsed or are still in treatment. Fortunately, the current view of cancer causation predicts that many cancers are preventable. Indeed, one-third of cancer-related deaths may be attributed to lifestyle factors. Lifestyle factors of particular importance are tobacco use and nutrition. Sun exposure is a significant risk factor for skin cancer (Chapter 53), and sexual exposure to certain strains of human papillomavirus predisposes to cervical cancer (Chapter 34). The high incidence and relative ease of screening for breast, cervical, colorectal, and prostate cancers have prompted the development of guidelines for early detection of these cancers. The current recommendations for early detection of cancer in average-risk, asymptomatic persons are shown in Table 7.3. Statistics regarding some of the major forms of cancer are shown in Fig. 7.2. Further discussions of particular cancers can be found in chapters relating to corresponding body systems.

Tobacco Use

The impact of tobacco use on cancer-related death can be most vividly seen by looking at cancer death rates in the United States from 1930 to 1998 (Fig. 7.3). Whereas all other cancer-related death rates declined or remained relatively stable, the death rate from lung cancer increased dramatically. The increase is attributable almost entirely to smoking. Lung cancer remains the leading cause of cancer death in both men and women, accounting for nearly 30% of all cancer deaths. Lung cancer has one of the worst survival rates of all cancers—only 15%. In addition to lung cancer, tobacco use has been linked with cancer of the pancreas, bladder, kidney, mouth, esophagus, and cervix (Fig. 7.4). Smoking prevalence among adults in the United States declined from 42% in 1965% to 18% in 2013, with rates being approximately 5% lower in women than in men. An estimated 42 million U.S. adults currently smoke cigarettes. Approximately 16% of high school students reported being cigarette smokers in 2013. Approximately 2% of adults use electronic nicotine devices (e-cigarettes). There is little known about the relative risk of these products compared with traditional tobacco.

Carcinogens can be grouped into two major types: those that cause genetic damage (initiators) and those that promote growth of the tumor (promoters). Tobacco smoke contains hundreds of compounds, many of which have known genotoxicity (e.g., polycyclic aromatic hydrocarbons, nicotine derivatives) and probably serve as initiators. Tobacco smoke also contains promoters, which spur the mutant cells to proliferate.

Nutrition

The scientific study of nutrition and cancer is complex, and it is not clear how single nutrients, combinations of nutrients, overnutrition

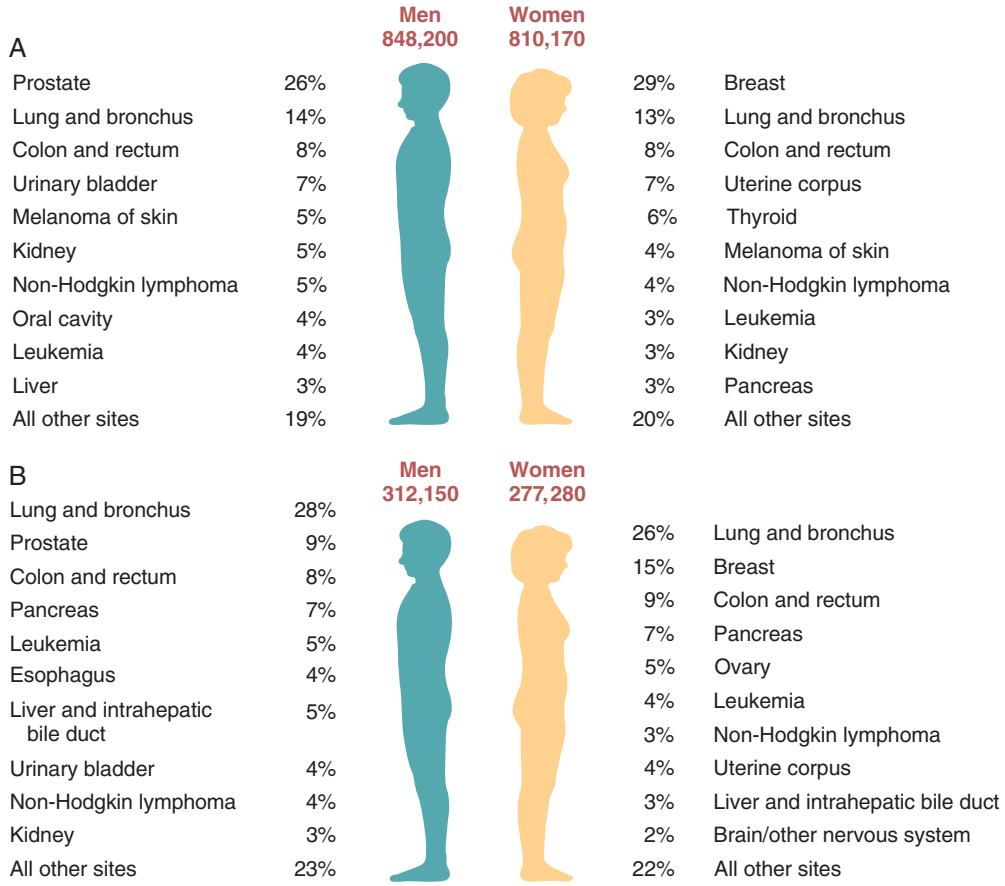
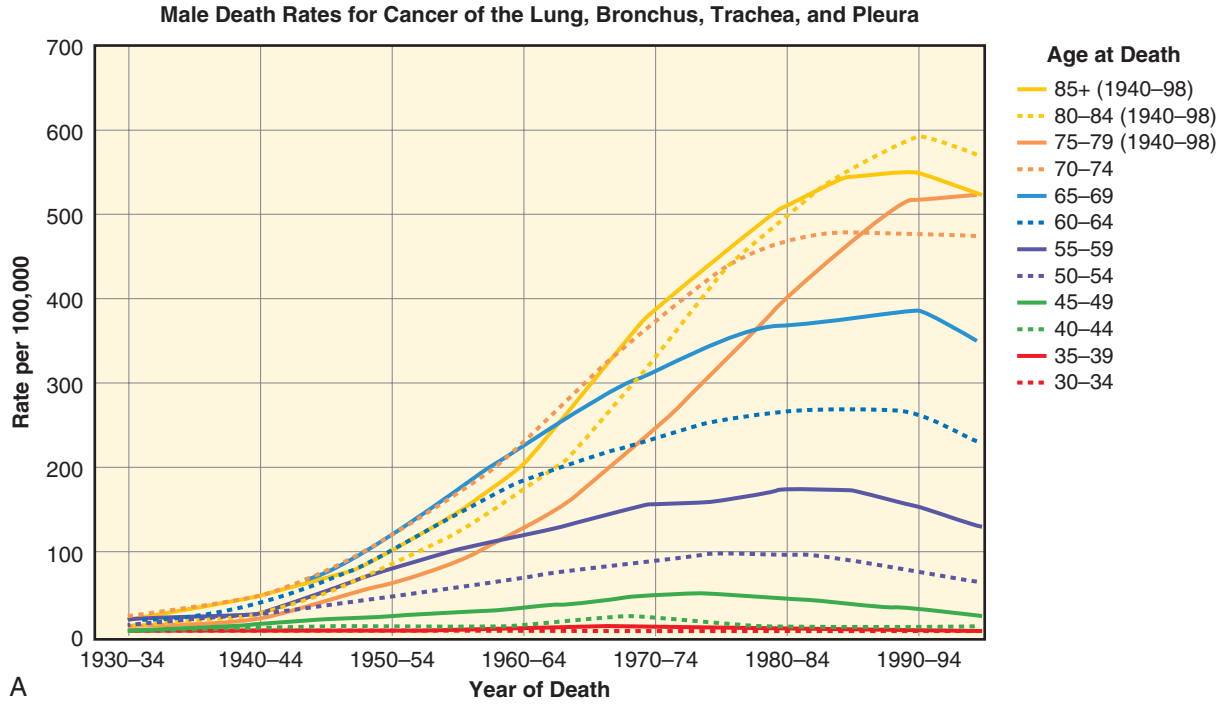


FIG 7.2 United States 2015 estimated new cancer cases **(A)** and estimated cancer deaths **(B)** in 10 leading sites by gender. Excludes basal and squamous cell skin cancers and in situ carcinomas except urinary bladder. (Data from American Cancer Society: *Cancer facts and figures—2015*, Atlanta, GA, 2015, American Cancer Society.)

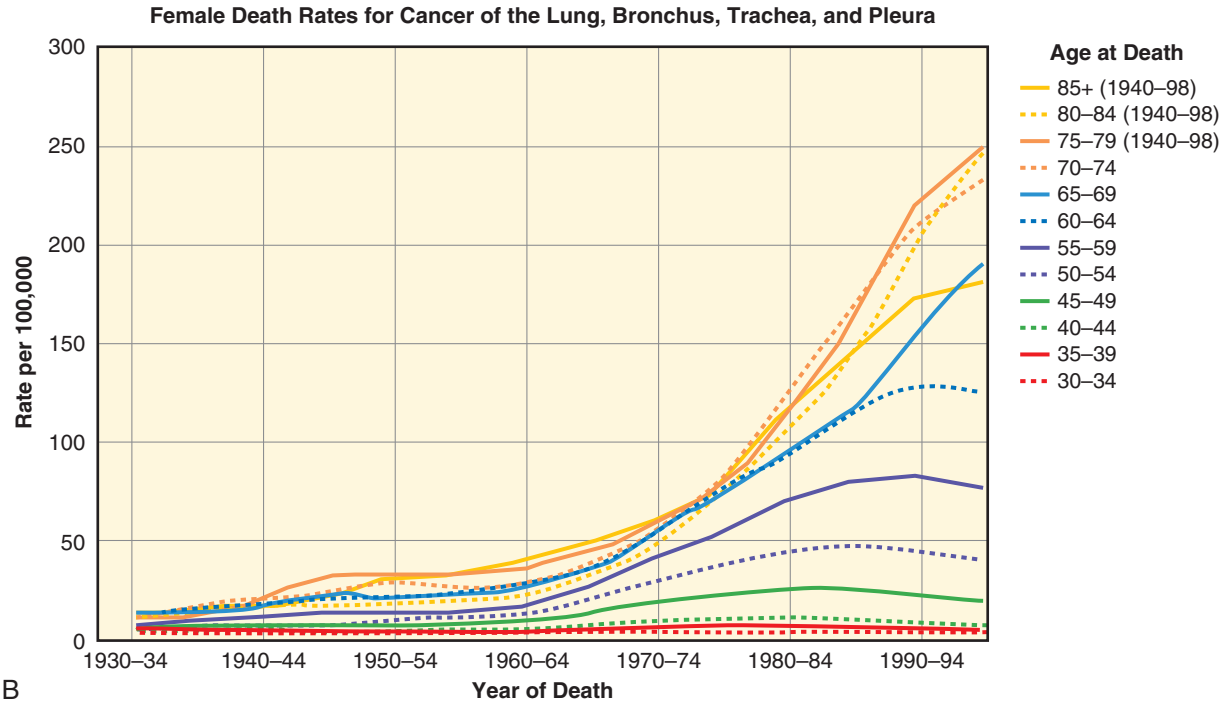
TABLE 7.3 Screening Guidelines for the Early Detection of Cancer in Average-Risk Asymptomatic People

Cancer Site	Population	Test or Procedure	Frequency
Breast	Women, age 20+	Breast self-examination (BSE)	It is acceptable for women to choose not to do a BSE or to do a BSE regularly (monthly) or irregularly. Beginning in their early 20s, women should be told about the benefits and limitations of BSE. Whether a woman ever performs a BSE, the importance of prompt reporting of any new breast symptoms to a health professional should be emphasized. Women who choose to do BSE should receive instruction and have their technique reviewed on the occasion of a periodic health examination.
		Clinical breast examination (CBE)	For women in their 20s and 30s, it is recommended that CBE be part of a periodic health examination, preferably at least every 3 years. Asymptomatic women aged 40+ should continue to receive a CBE as part of a periodic health examination, preferably annually.
		Mammography	Begin annual mammography at age 40*
Cervix	Women, age 21–65	Pap test and HPV DNA test	Cervical cancer screening should begin at age 21. For women ages 21–29, screening should be done every 3 years with conventional or liquid-based Pap tests. For women ages 30–65, screening should be done every 5 years with both HPV test and the Pap test (preferred), or every 3 years with the Pap test alone (acceptable). Women ages 65+ who have had ≥3 consecutive negative Pap tests or ≥2 consecutive negative HPV and Pap tests within the last 10 years, with the most recent test occurring within 5 years, and women who have had a total hysterectomy should stop cervical cancer screening. Women should not be screened annually by any method at any age.
Colorectal	Men and women, age 50+	Fecal occult blood test (FOBT) with at least 50% test sensitivity for cancer, fecal immunochemical test (FIT) with at least 50% test sensitivity for cancer, **or**	Annual, starting at age 50. Testing at home with adherence to manufacturer's recommendation for collection techniques and a number of samples is recommended. FOBT with single stool sample collected on the clinician's fingertip during a DRE in the health care setting is not recommended. Guaiac-based toilet bowl FOBT tests are also not recommended. In comparison with the guaiac-based tests for the detection of occult blood, immunochemical tests are more patient friendly and are likely to be equal or better in sensitivity and specificity. There is no justification for repeating FOBT in response to an initial positive finding.
		Stool DNA test, **or**	Every 3 years, starting at age 50
		Flexible sigmoidoscopy (FSIG), **or**	Every 5 years, starting at age 50. FSIG can be performed alone, or consideration can be given to combining FSIG performed every 5 years with a highly sensitive guaiac-based FOBT or FIT performed annually.
		Double-contrast barium enema (DCBE), **or**	Every 5 years, starting at age 50
		Colonoscopy	Every 10 years, starting at age 50
		CT Colonography	Every 5 years, starting at age 50
Endometrial	Women, at menopause		At the time of menopause, women at average risk should be informed about risks and symptoms of endometrial cancer and strongly encouraged to report any unexpected bleeding or spotting to their physicians.
Lung	Current or former smokers (quit within the past 15 years) ages 55–74 in good health with at least a 30 pack-year history	Low-dose helical CT (LDCT)	Clinicians with access to high-volume, high-quality lung cancer screening and treatment centers should initiate a discussion about lung cancer screening with apparently healthy patients aged 55–74 who have at least a 30 pack-year smoking history and who currently smoke or have quit within the last 15 years. A process of informed and shared decision making with a clinician related to the potential benefits, limitations, and harms associated with screening for lung cancer with LDCT should occur before any decision is made to initiate lung cancer screening. Smoking cessation counseling remains a high priority for clinical attention in discussions with current smokers, who should be informed of their continuing risk of lung cancer. Screening should not be reviewed as an alternative to smoking cessation.
Prostate	Men, age 50+	Digital rectal examination (DRE) and prostate-specific antigen test (PSA)	Men who have at least 10-year life expectancy should have an opportunity to make an informed decision with their health care provider about whether to be screened for prostate cancer, after receiving information about the potential benefits, risks, and uncertainties associated with prostate cancer screening. Prostate cancer screening should not occur without an informed decision-making process.
Cancer-related checkup	Men and women age 20+		On the occasion of a periodic health examination, the cancer-related checkup should include examination for cancers of the thyroid, testicles, ovaries, lymph nodes, oral cavity, and skin, as well as health counseling about tobacco, sun exposure, diet and nutrition, risk factors, sexual practices, and environmental and occupational exposures.

American Cancer Society: *Cancer prevention and early detection facts and figures—2015–2016*, Atlanta, GA, 2015, American Cancer Society.
CT, computed tomography.
*Beginning at age 40, annual clinical breast examination should ideally be performed before mammography.



A



B

FIG 7.3 Death rates for cancer of the lung, bronchus, trachea, and pleura, by age at and year of death, 1930–1998. **A,** Male death rates. **B,** Female death rates. (Data From U.S. Department of Commerce, Bureau of the Census. *Mortality statistics.* Washington: U.S. Government Printing Office, 1930–1936; U.S. Department of Commerce, Bureau of the Census. *Vital statistics of the United States, Part I.* Washington: Government Printing Office, 1937–1944; Federal Security Agency, Public Health Service, National Office of Vital Statistics. *Vital statistics of the United States, Part I.* Washington: Government Printing Office, 1945–1949; U.S. Department of Health, Education, and Welfare, Public Health Service, National Office of Vital Statistics. *Vital statistics of the United States. Volume II.* Washington: Government Printing Office, 1950–1959; Centers for Disease Control and Prevention, National Center for Health Statistics. *Multiple cause of death public use data files.* Hyattsville, MD: National Center for Health Statistics, 1960–1998.)

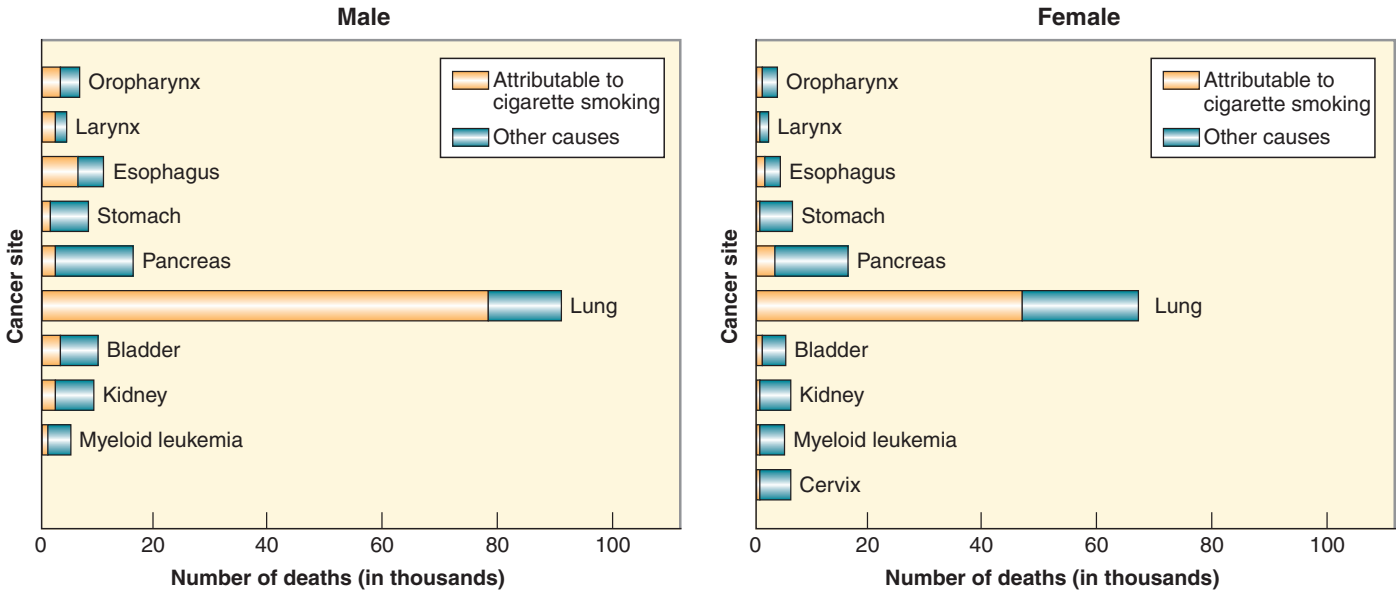


FIG 7.4 Annual cancer deaths attributable to smoking in males and females in the United States. (From Centers for Disease Control and Prevention: Annual smoking-attributable mortality, years of potential life lost, and productivity losses—United States, 2000-2004, *MMWR Morb Mortal Wkly Rep* 2008;57(45):1226–1228.)

and energy imbalance, or the amount and distribution of body fat affect a person's risk for specific cancers. The ACS suggests a mostly plant-based diet emphasizing a variety of vegetables, fruits, and whole grains. The ACS endorses limiting the intake of red and processed meats while controlling total caloric intake to maintain a healthy weight and avoiding excessive alcohol intake. Alcohol intake has been linked to a number of cancers, including breast, esophageal, laryngeal, and liver cancer. Alcohol may exert its cancer-promoting effects through impairment of the liver's ability to metabolize harmful substances and endogenous hormones. Moderate alcohol intake has been shown to increase estrogen levels, which may account for its promoting effects on breast cancer. As a carbohydrate-dense substance, alcohol may contribute to cancer risk through its effects on insulin secretion. Insulin is a general growth factor for a number of tissues. Limiting alcohol intake may provide a modest reduction in cancer risk.

Antioxidants

The emphasis of cancer prevention generally has been on the identification and avoidance of cancer-causing agents. However, increasing interest has been shown in finding substances with cancer-protective properties. The fact that DNA damage is an important step in cancer initiation, coupled with the knowledge that oxygen free radicals can impart this damage, led to the idea that antioxidants may have protective effects for cancer. The specific agents tested in clinical trials included β-carotene, vitamin E, vitamin C, selenium, retinol, zinc, riboflavin, and molybdenum. None of the completed trials produced convincing evidence to justify the use of traditional antioxidant-related vitamins or minerals for cancer prevention.

Vitamin A and the antioxidant trio of vitamin E, β-carotene, and vitamin C have been most widely studied. The use of antioxidants to prevent cancer sounds like a good idea; however, several large-scale studies have failed to reveal a benefit, and some have found that the risk of cancer may be increased. At present, it may be prudent to consume a diet high in natural fruit and vegetable sources of antioxidants.

KEY POINTS

- The risk of developing cancer increases with age. It is estimated that men have almost a 1 in 2 lifetime chance of developing cancer, whereas women have a little more than a 1 in 3 chance.
- The development of many cancers is related to lifestyle, particularly tobacco use and nutrition. Smoking cessation is considered important in reducing cancer risk. Guidelines regarding nutrition are less clear; however, avoiding excess weight gain is recommended. Limiting excessive calorie and alcohol intake while increasing intake of dietary fiber, fruit, and vegetables may be of benefit.

GENETIC MECHANISMS OF CANCER

Despite much progress in our understanding of how mechanisms of growth control and cellular differentiation may go awry, there is still no simple answer to the question, "What causes cancer?" It is increasingly evident, however, that cancer is primarily a disorder of gene expression. Early support for a genetic basis of cancer came from the observation that cancer often resulted from agents known to damage DNA. In the 1970s a number of potential cancer-causing agents (carcinogens) were identified by demonstrating their mutagenic potential. The suggestion that mutant genes were the basis for cancer launched intense research to identify the cancer-causing gene or genes.

Cancer-critical genes are grouped into two broad classes, according to whether overactivity of the gene contributes to cancer (gain-of-function mutations) or whether underactivity is the problem (loss-of-function mutations). Both categories of genes result in similar effects in enhancing cell proliferation and survival. Genes in the first category are called *proto-oncogenes,* which normally code for components of the cellular growth–activating pathways. A proto-oncogene in its mutant, overactive, or overexpressed form is called an *oncogene.* Genes in the second category of cancer-related genes are called *tumor suppressor genes,* which normally inhibit cell proliferation. Cancers may arise when tumor suppressor

gene function is lost or abnormally inhibited. To achieve malignant transformation, a cell must generally suffer mutations in a combination of these growth regulatory genes. A cell thus transformed passes on these mutations to its progeny when it divides and forms a clone of abnormally proliferating cells. Numerous studies have begun to unravel the details of how proto-oncogenes and tumor suppressor genes may dysfunction and contribute to the malignant phenotype. In addition to the genes that regulate the cell cycle, two other categories of genes that monitor and maintain the genome contribute indirectly to the development of cancer. These are the DNA-repair genes and the genes that regulate apoptosis (see Chapter 4).

Proto-Oncogenes

Proto-oncogenes were the first of the tumor-associated genes to be discovered, and hundreds have been described to date. As often happens in the study of genes, a gene associated with a disease process is identified long before its normal cellular function is elucidated. Thus genes associated with cancer are traditionally named for the cancer in which they were first discovered (in mutant form) rather than for their normal cellular function. Many of the first cancer-associated genes, called *oncogenes,* were initially identified in viruses and still retain the name reflecting their viral discovery. The term *proto-oncogene* was created to label the normal cellular gene that can be transformed into an oncogene by activating (gain-of-function) mutations. A representative list of known proto-oncogenes is shown in Table 7.4.

The majority of proto-oncogenes described to date code for components of cell-signaling systems that promote cell proliferation. These components can be grouped into four broad categories: (1) growth factors, (2) receptors, (3) cytoplasmic signaling molecules, and (4) nuclear transcription factors (Fig. 7.5). Excessive activity in any of these components may release the cell from environmental feedback and allow it to proliferate abnormally.

TABLE 7.4 Examples of Gain-of-Function Proto-Oncogenes and Their Mechanisms of Action

Factor	Type of Cancer
Mitogens and Growth Factors	
PDGF	Glioma (brain)
FGF	Melanoma
EGF	Breast
TGF-α	Breast, numerous others
Cell Membrane Receptors	
ERBB1 (EGF receptor)	Breast, brain
HER-2 (ERBB2)	Breast, ovarian
RET	Thyroid
Cytoplasmic Signaling Molecules	
RAS	Lung, ovarian, colon, pancreatic, leukemia
ABL	Leukemia
JAK2	Myeloproliferative, leukemia
Transcription Factors and Cell Cycle Regulators	
C-MYC	Leukemia, breast, lung, neuroblastoma
MYB	Various
JUN	Various
FOS	Various
REL	Various
Cyclin D, E	Breast, lymphoma
Cyclin-dependent kinase 4	Melanoma, sarcoma

EGF, Epidermal growth factor; *FGF,* fibroblast growth factor; *PDGF,* platelet-derived growth factor; *TGF,* transforming growth factor.

Growth Factors (Mitogens)

The first proto-oncogenes to be discovered coded for growth factors. A great deal of intercellular communication is accomplished through the cell-to-cell transmission of growth factors. Growth factors are small peptides that are manufactured by cells and secreted into the extracellular space. They diffuse to nearby cells and interact with receptors on the target cell surface. Binding of growth factors to cell surface receptors activates signaling cascades within the cell that enhance proliferation. As a general principle, cells do not independently produce growth factors sufficient to stimulate their own proliferation. The proliferation signals must be produced by the cell's environment. The cell's environment also conveys growth-inhibiting signals. Overproduction of stimulatory growth factors by a mutant proto-oncogene can shift the balance of signals and produce excessive self-stimulated growth (autocrine signaling). Examples of tumor-secreted growth factors include platelet-derived growth factor (PDGF), transforming growth factor-α (TGF-α), and epidermal growth factor (EGF). Certain cancer types typically secrete particular growth factors. For example, PDGF is commonly oversecreted in glial cell cancers (brain tumors) and connective tissue cancers (sarcomas).

Growth Factor Receptors

Peptide growth factors (mitogens) cannot penetrate the cell membrane directly, so their presence at the cell surface must be transmitted intracellularly by cell surface receptors. Receptors are transmembrane proteins with the mitogen-binding area on the outside of the cell and an enzyme-activating area on the inside of the cell. These receptors are extremely specific; they will bind with only one particular mitogen. Binding activates a series of reactions within the cell that eventually leads to cell proliferation.

A mutational event may allow the expression of receptors that should not be present at all or allow excessive amounts of normally present receptors, or it may produce receptors with abnormally high affinity. All of these changes result in excessive responsiveness to the mitogens normally present in the cell's environment. Some mutant receptors may even be active in the absence of growth factors and spur the cell to divide despite the absence of environmental signals to do so. An important example of a receptor abnormality is the overexpression of human epidermal growth factor receptor type 2 (HER2) receptors in about 25% of breast cancers. The overactive receptors stimulate proliferation of tumor cells even when there is little or no epidermal growth factor bound to them.

Cytoplasmic Signaling Pathways

A third way in which oncogenes may facilitate proliferation is by the manufacture of excessive or abnormal components of the intracellular signaling pathways. These pathways involve numerous enzymes and chemicals that normally function to transmit signals from activated receptors at the cell surface to the cell nucleus. A mutant proto-oncogene that codes for excessive or abnormal cytoplasmic signaling components could cause activation of the pathway even though no signal was received at the cell surface. A well-understood example of this mechanism is mutations of the *ras* gene family. Proteins encoded by *ras* genes are monomeric G-proteins that transmit signals from receptors at the cell surface into the interior of the cell. The *ras* protein is active when it has guanosine triphosphate (GTP) bound to it, but it quickly hydrolyzes the GTP, thus automatically turning itself off after a brief period of

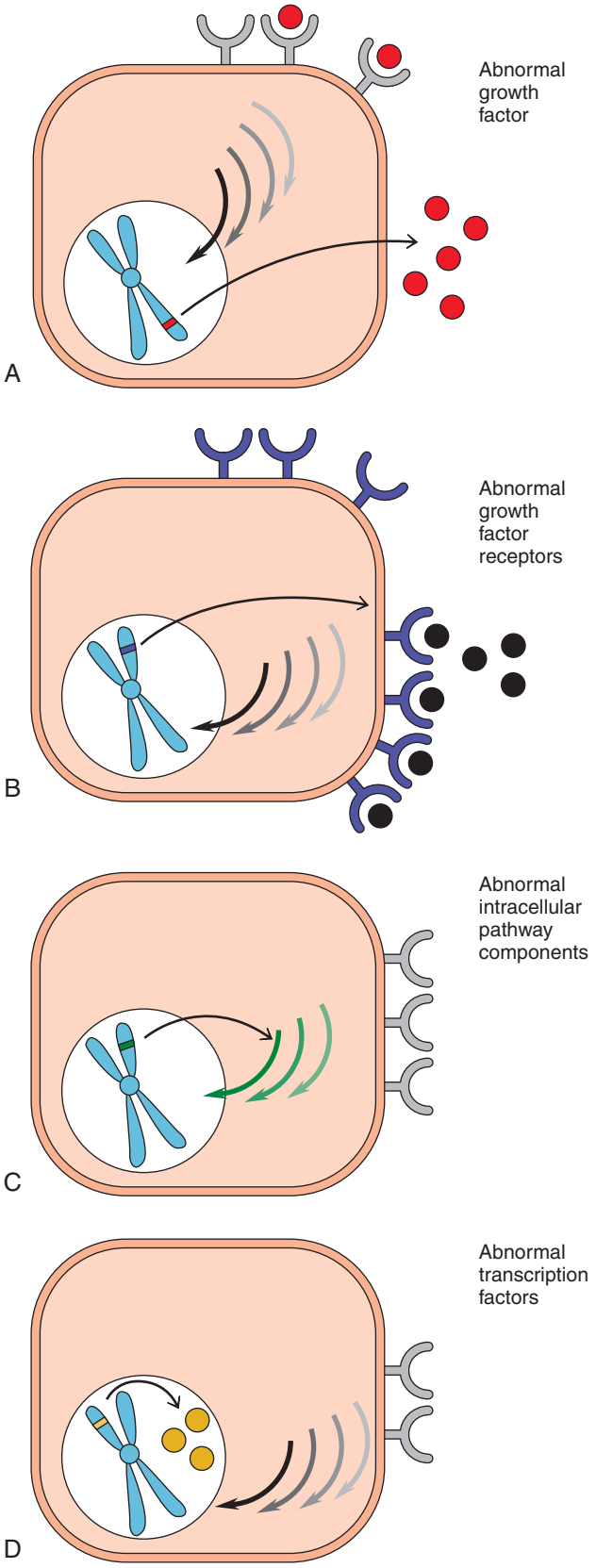


FIG 7.5 Possible effects of proto-oncogene activation on growth signaling pathways. **A,** Production of growth factors (mitogens). **B,** Production of growth factor receptors. **C,** Intracellular pathway disturbances. **D,** Activation of transcription factors for growth.

activity. A mutation in the *ras* gene can code for a protein that is unable to hydrolyze GTP, so it remains persistently active and stimulates cell proliferation inappropriately. Mutations of the *ras* genes occur in about 20% of all human cancers, including leukemias and lung, ovarian, colon, and pancreatic cancer.

Transcription Factors

The entire proliferation pathway, including the growth factor (mitogen), the receptor, and the intracellular cascade, ultimately affects transcription of a set of genes in the nucleus that spur the cell to enter the S phase. A number of proto-oncogenes have been identified that code for transcription factors in the nucleus. Transcription factors are proteins that must be assembled at the promoter area to begin gene transcription (see Chapters 3 and 5). Transcription factors are normally sequestered and prevented from indiscriminate activity until appropriate signals cause their release. Mutations in transcription factor genes may cause overproduction of transcription factors or interfere with the normal mechanisms for keeping them in check. *Myc, jun,* and *fos* are examples of proto-oncogenes that code for nuclear transcription factors. Abnormalities of the *myc* genes are found in numerous cancers, including lung and breast cancer, leukemia, and neuroblastoma.

From Proto-Oncogene to Oncogene

Proto-oncogenes become activated oncogenes when mutations alter their activity so that proliferation-promoting signals are generated inappropriately. At least four general ways in which proto-oncogenes can be activated are known (Fig. 7.6): (1) oncogenes may be introduced into the host cell by a retrovirus; (2) a proto-oncogene within the cell may suffer a mutagenic event that changes its structure and function; (3) a DNA sequence that normally regulates proto-oncogene expression may be damaged or lost and allow the proto-oncogene to become abnormally active; and (4) an error in chromosome replication may cause extra copies of the proto-oncogene to be included in the genome (amplification).

In the early 1960s it was discovered that certain viruses were associated with cancer in various animal models. Researchers speculated that a virus could introduce a mutant, cancer-causing gene (oncogene) into the host's cells. Indeed, malignant cells containing the cancer-causing viruses were shown to have incorporated a small number of viral genes into their cellular DNA. The presence of these oncogenes was required to maintain the malignant state of the cell.

Only a few types of human cancers are thought to be associated with viruses (Table 7.5). For example, human immunodeficiency virus is associated with Kaposi sarcoma, Epstein-Barr virus with Burkitt lymphoma, and human T-lymphocyte virus type I with adult T-cell leukemia-lymphoma. Viruses can interrupt normal cell growth control pathways and interfere with apoptotic pathways. Sometimes viruses incorporate into the host cell's genome and disrupt gene regulation. See Chapter 8 for a review of viral mechanisms.

Proto-oncogene expression is tightly regulated in a normal cell. A number of different mutations can affect proto-oncogene expression and activity. A point mutation in the coding region of the proto-oncogene can alter the structure of its protein product and make it hyperactive. An example of this mechanism is the abnormal *ras* protein described earlier. Even though the protein is synthesized in normal quantity, its activity is enhanced. Other mutations may lead to overproduction of a protein with normal structure. Gene amplification and chromosomal rearrangement during mitosis may release the proto-oncogene from its normal regulation and allow excessive transcription (Fig. 7.7).

In summary, mutational events in the cell's genome may cause overexpression of normal proto-oncogene products or production of altered and hyperactive proteins. Most known oncogenes act by releasing

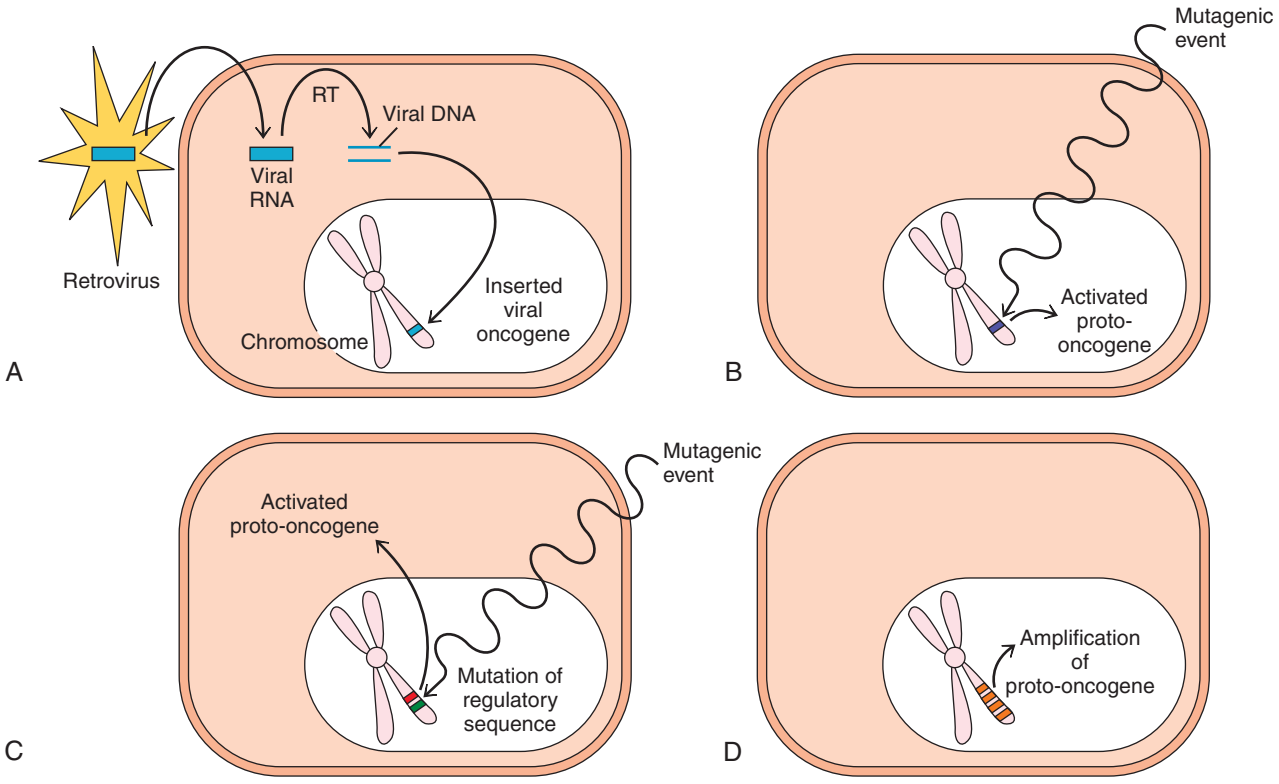


FIG 7.6 Mechanisms of proto-oncogene activation. **A,** Retroviral insertion. **B,** Proto-oncogene mutation. **C,** Regulatory sequence mutation. **D,** Proto-oncogene amplification. *RT,* Reverse transcriptase.

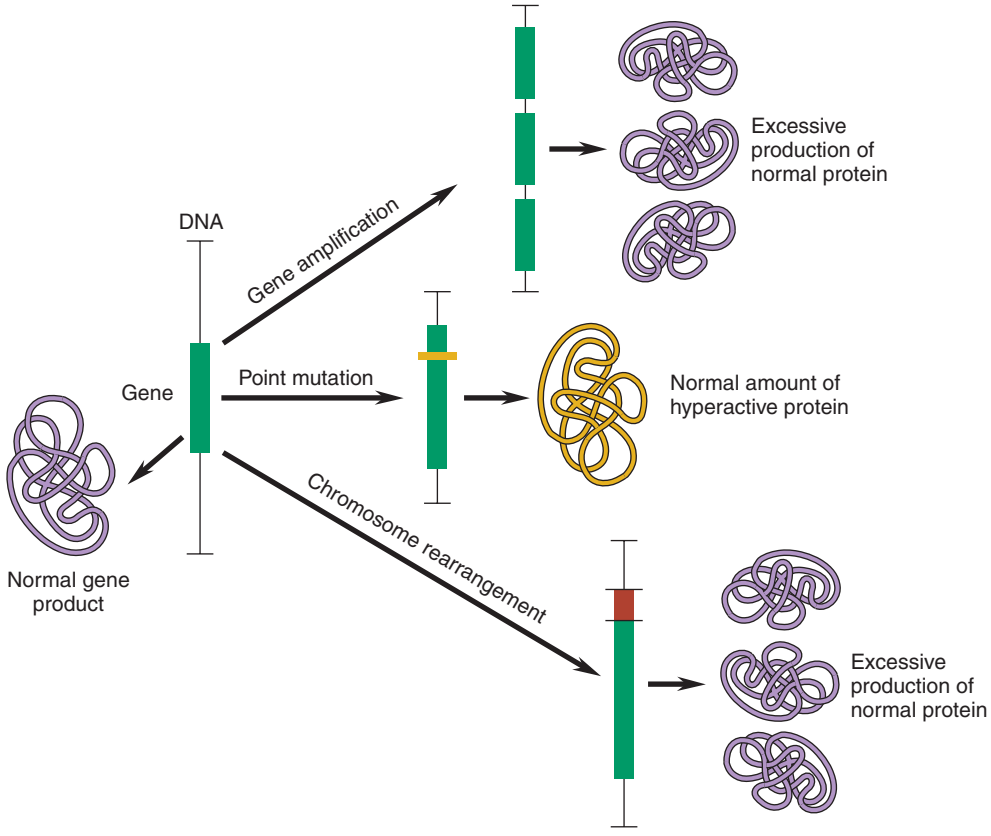


FIG 7.7 Overactivity of proto-oncogenes may be due to normal production of an abnormal protein (mutation in coding sequence) or excessive production of a normal protein (gene amplification or chromosome rearrangement).

TABLE 7.5 Viruses Associated With Human Cancers

Virus	Associated Cancer	Areas of High Incidence
DNA Viruses		
Papovavirus Family		
Papillomavirus (many distinct strains)	Warts (benign)	Worldwide
	Carcinoma of the uterine cervix	Worldwide
Hepadnavirus Family		
Hepatitis-B virus	Liver cancer (hepatocellular carcinoma)	Southeast Asia, tropical Africa
Herpesvirus Family		
Epstein-Barr virus	Burkitt lymphoma (cancer of B lymphocytes)	West Africa, Papua New Guinea
	Nasopharyngeal carcinoma	Southern China, Greenland
Human herpesvirus 8	Kaposi sarcoma	Central and Southern Africa
RNA Viruses		
Retrovirus Family		
Human T-cell leukemia virus type I (HTLV-1)	Adult T-cell leukemia/lymphoma	Japan, West Indies
Human immunodeficiency virus (HIV, the AIDS virus)	Kaposi's sarcoma (via human herpesvirus 8)	Central and Southern Africa
Flavivirus Family		
Hepatitis-C virus	Liver cancer (hepatocellular carcinoma)	Worldwide

For all these viruses, the number of people infected is much larger than the number who develop cancer: the viruses must act in conjunction with other factors.

From Alberts B et al, editors: *Molecular biology of the cell,* ed 6, New York, 2015, Garland Science, p 1130.

TABLE 7.6 Examples of Tumor Suppressor Genes

Gene	Cancer
RB	Retinoblastoma, sarcoma
P53	Li-Fraumeni syndrome, 50% of all tumors
DCC	Colorectal carcinoma
APC	Colorectal, stomach, pancreatic
BRCA1	Breast, ovarian
BRCA2	Breast, ovarian, prostate
WT1	Wilms tumor
WT2	Rhabdomyosarcoma
NF1	Neurofibromatosis type 1, astrocytoma
NF2	Neurofibromatosis type 2, meningioma
VHL	Renal cell carcinoma
MEN1	Multiple endocrine neoplasia
MTS1	Melanoma, leukemia, sarcomas, several carcinomas
SMAD2, SMAD4	Colon, pancreatic
PTCH	Basal cell carcinoma

the cell from its dependence on growth and survival signals in its environment. This effect usually is accomplished by gain-of-function abnormalities in the proliferation pathways that involve growth factors (mitogens), receptors, cytoplasmic signaling molecules, or nuclear transcription factors.

Tumor Suppressor Genes

To become malignant, cells must devise ways to evade the normal inhibitory mechanisms that keep the brakes applied to cell division. It is not enough to simply overstimulate growth-promoting signals. Critical elements of the proliferation-inhibiting pathways are defective in most cancers. The components of the inhibitory machinery are specified by the so-called *tumor suppressor genes.* Tumor suppressor genes were more difficult to study because they contributed to cancer only when they were not there. The first tumor suppressor gene to be discovered was the *Rb* gene, so named because of its role in retinoblastoma, a cancer of the eye. A familial form of retinoblastoma is associated with the transmission of a genetic defect; a portion of chromosome 13 is missing, which is where the *Rb* gene is normally located. An absent *Rb* gene predisposes an individual to cancer, but cancer will not develop unless the other copy of the *Rb* gene (from the other parent) is also damaged (Fig. 7.8).

Since the initial discovery of the *Rb* tumor suppressor gene, researchers have compiled an impressive list of other genes that appear to function as inhibitors of cellular proliferation (Table 7.6). As with the *Rb* gene, both copies of the tumor suppressor genes usually are inactivated when cancer develops. A person who inherits a defective copy of a tumor suppressor gene from one parent has a much higher risk of cancer than a person who inherits two healthy copies. Knowledge about the sequence of many of these genes provides the opportunity to screen individuals with familial cancers to determine whether they carry a defective gene. Detection of defective tumor suppressor genes is easier than determining their normal cellular functions, but steady progress is being made.

Why do tumor suppressor genes stop functioning? As with proto-oncogene activation, genetic mutations are the usual culprits. Chromosome deletions, point mutations, or chromosome loss through nondisjunction may knock out tumor suppressor gene function. Tumor suppressor gene function may also be lost through an "epigenetic" process that "silences" the gene. Epigenetic influences do not change the DNA sequence (no mutation is required), but change the packaging of DNA and chemically modify it (e.g., methylation) so that the gene is inactivated. This inactivation can be passed on to daughter cells during mitosis.

The Rb Gene

The *Rb* gene codes for a large protein in the cell nucleus (pRb) that has been labeled the "master brake" of the cell cycle (see Chapter 3). It blocks cell division by binding transcription factors (E2F), thereby inhibiting them from transcribing the genes that initiate the cell cycle (Fig. 7.9). The Rb protein can be induced to release the transcription factors when it is sufficiently phosphorylated. Proliferation-promoting signals in the cell increase cyclin-dependent kinase (cdk) enzymes and promote pRb phosphorylation, whereas growth-inhibiting signals prevent phosphorylation. Thus an inactivating mutation of the *Rb* genes removes one of the major restraints on cell division. Defective pRb is common to a number of different cancers.

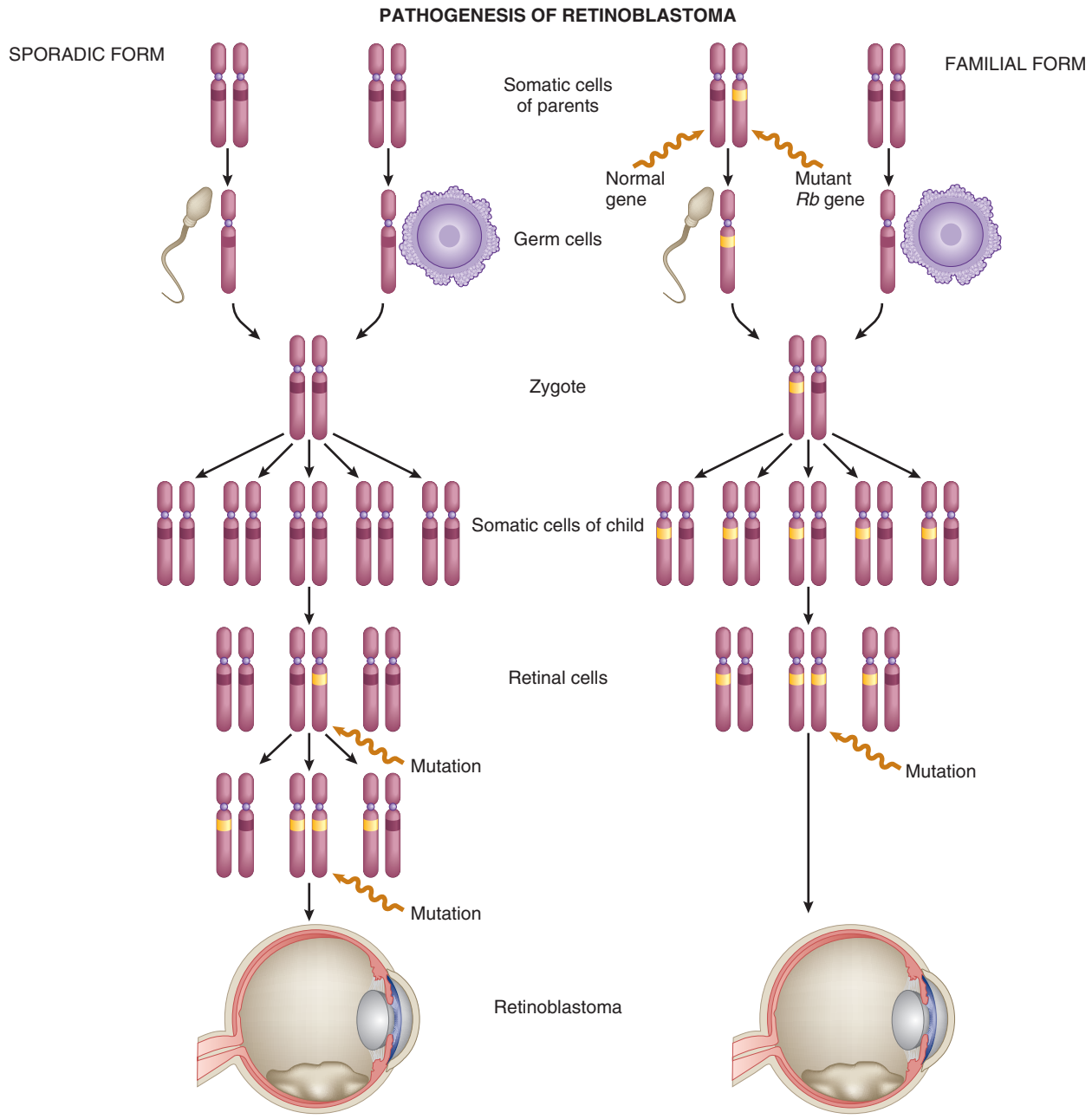


FIG 7.8 Both DNA copies (alleles) of the *Rb* tumor suppression gene must be dysfunctional for occurrence of retinoblastoma. Inheriting a defective *Rb* gene predisposes an individual to the development of cancer because only a single mutational event is required to inactivate pRb function.

The P53 Gene

The most common tumor suppressor gene defect identified in cancer cells involves *P53,* so named because of the protein's molecular mass of 53 kilodaltons (also called *TP53*). More than half of all types of human tumors lack functional *P53.* The p53 protein, like pRb, inhibits cell cycling. Unlike pRb, however, normally very little p53 is found in cells, and it accumulates in response to cellular, particularly DNA, damage. P53 is a transcription factor that binds to damaged DNA and regulates hundreds of genes. It stalls cell division, presumably to allow time for DNA repair before DNA replication in the S phase (Fig. 7.10). In the face of excessive damage (or other distress signals), p53 may direct the cell to initiate apoptosis. A defect in p53 function disrupts this important quality control system, allowing genetically damaged and unstable cells to survive and continue to replicate (see Fig. 7.10). Genetically unstable cells have a propensity to accumulate more cancer-promoting mutations while they proliferate. The *P53* gene is important for therapeutic reasons as well. Chemotherapy- and radiation-induced cell death is mediated in large part by p53. These agents usually do not kill cancer cells directly; rather, they cause enough cellular damage in the target cell to trigger p53-mediated apoptosis. Cancer cells that lack functional p53 may therefore be resistant to some radiation and chemotherapeutic protocols.

BRCA1 and BRCA2 Genes

Many tumor suppressor genes have been identified through studies of inherited predisposition to certain types of cancer. The breast cancer genes *BRCA1* and *BRCA2* are important examples. Women with a family

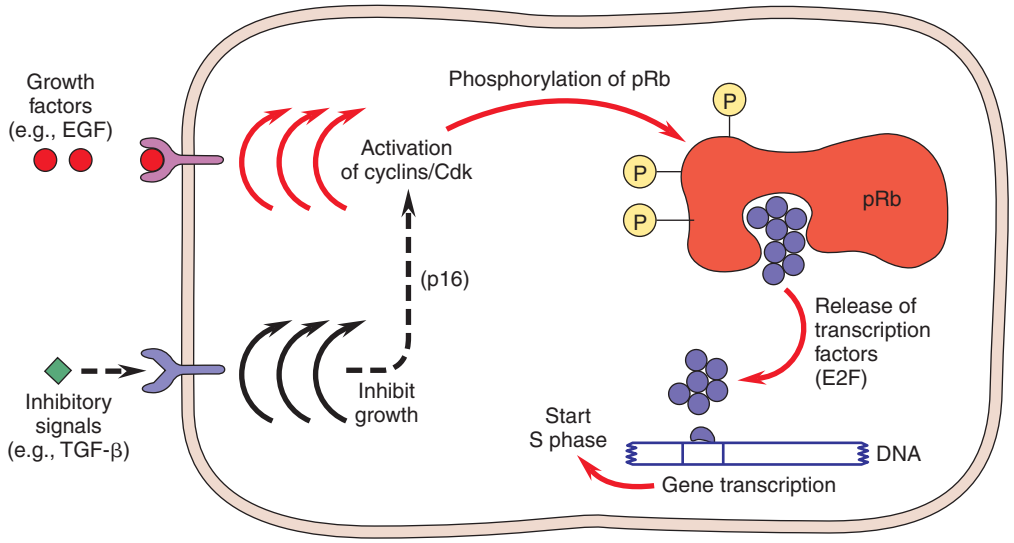


FIG 7.9 The Rb protein functions to bind transcription factors in the nucleus and keep them from participating in the transcription of cell cycle–related genes. pRb is induced to release its hold on the E2F transcription factors when it is sufficiently phosphorylated by cyclin-dependent kinases (cdk). Cyclin-dependent kinases are activated by cyclin proteins that accumulate when growth factors bind to receptors and stimulate growth pathways. Other signals, such as transforming growth factor-β (TGF-β), inhibit the activity of cyclin/cdk through activation of inhibitory proteins such as p16. A loss of pRb function removes the "major brake" on cell division. *P,* Phosphate group; *EGF,* epidermal growth factor.

history of breast cancer and an inherited defect in the *BRCA1* gene have about a 50% risk of developing breast cancer. The age of onset of inherited breast cancer is earlier than the onset of noninherited (sporadic) forms, and the prevalence of bilateral breast cancer is higher. Inherited forms of breast cancer account for only about 5% to 10% of all cases of breast cancer, but study of the genes involved can provide important insights into breast cancer biology in general.

Defects in numerous other tumor suppressor genes have been identified in certain types of cancers (see Table 7.6), including *APC* and *DCC* in colorectal cancer, *NF1* and *NF2* in neurofibromatosis, and *VHL* in renal cell cancers. The functions of tumor suppressor genes are varied, but most appear to inhibit proliferation, repair DNA, or induce apoptosis in defective cells. In general, any cellular alteration that promotes proliferation or inhibits cell death can contribute to an increased risk of tumor development.

KEY POINTS

- Cancer is thought to develop when proto-oncogenes become inappropriately activated in a cell or when tumor suppressor genes become inactivated. This change in gene function is usually due to mutations in the cell's DNA.
- Mutant proto-oncogenes disrupt the intercellular communication pathway that normally regulates cell proliferation. This disruption may occur through abnormal production of growth factors, receptors, cytoplasmic signaling molecules, or nuclear transcription factors.
- Both copies of a tumor suppressor gene usually must be inactivated to eliminate its function. Tumor suppressor genes inhibit cellular proliferation in various ways. The Rb protein serves as a "master brake" on cell proliferation by inhibiting transcription factors. *P53* inhibits cell cycling when the cell is damaged to allow time for DNA repair. *P53* is also important in initiating apoptosis of damaged or unwanted cells.

MULTISTEP NATURE OF CARCINOGENESIS

From the preceding discussion, it might seem that simply activating an oncogene in a normal cell or knocking out a tumor suppressor gene would be sufficient to transform it into a malignant cell. Such has not proved to be the case. Growth regulation of mammalian cells appears to be organized in such a manner that a single aberrant gene is unlikely to induce conversion to full malignancy. Different genes function in distinct ways and may affect only a subset of the changes necessary to achieve full malignancy. For example, introduction of the *ras* oncogene into normal cells in culture causes them to show anchorage independence, but they are unable to form tumors when inoculated into an animal. Anchorage independence is a typical feature of most transformed cells and means that they are capable of proliferating even if they are not attached to a matrix. Normal cells will not divide and will initiate apoptosis if they do not have a space on the matrix on which to anchor themselves. Similarly, the *myc* oncogene allows cells to grow indefinitely in culture, but these immortal cells are still unable to induce tumor formation. However, when both the *ras* and the *myc* oncogenes are introduced into normal cells, they become fully malignant.

These culture experiments support the clinical observation that carcinogenesis is a multistep phenomenon. The steps of carcinogenesis have been labeled *initiation, promotion,* and *progression* (Fig. 7.11).

Initiation

Initiating events are thought to be the genetic mutations that inappropriately activate proto-oncogenes and inactivate tumor suppressor genes. However, the genetic mutations are not evident until the mutant cell proliferates. *Proliferation* is a requirement for cancer development, and nonproliferating cells are unlikely to cause cancer. It has been suggested that several mutations may be necessary to achieve full malignancy. The development of colorectal cancer is a well-documented example of these sequential changes (Fig. 7.12). Each individual cancer is likely to have its own unique combination of mutations that eventually

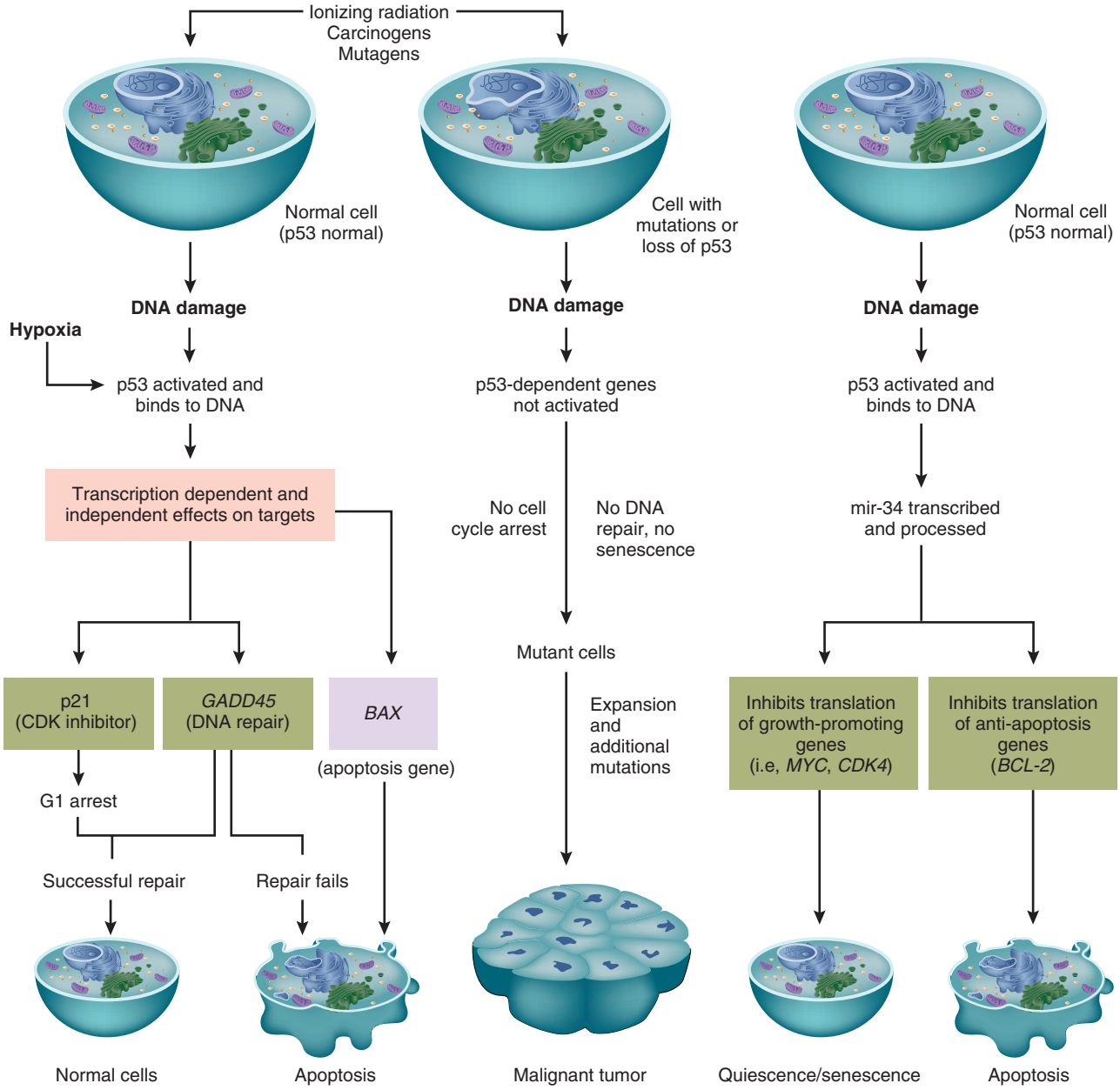


FIG 7.10 Role of *P53* (TP53) in maintaining the integrity of the genome. Damage to DNA in cells with functional *P53* stalls the cell cycle so that DNA can be repaired. If repair fails, then the cell undergoes apoptosis to prevent the proliferation of DNA-damaged cells. If the *P53* is not functional, genetically unstable cells may be allowed to survive and proliferate.

lead to malignant behavior. Genotyping of tumor cells is an important development in understanding tumor cells and individualizing therapy.

A number of etiologic agents are considered important initiators of cancer. The term *carcinogen* is applied to agents and substances capable of inducing cancer. Some carcinogens are complete carcinogens in that they are capable of initiating genetic damage as well as promoting cellular proliferation, whereas many others are only partial carcinogens. Partial carcinogens are often promoters that stimulate growth but are incapable of causing genetic mutations sufficient to initiate cancer by themselves. Examples of known carcinogens are ultraviolet and ionizing radiation, certain viruses, asbestos, and numerous chemicals. Most known chemical carcinogens are encountered through repeated occupational exposure (Box 7.1).

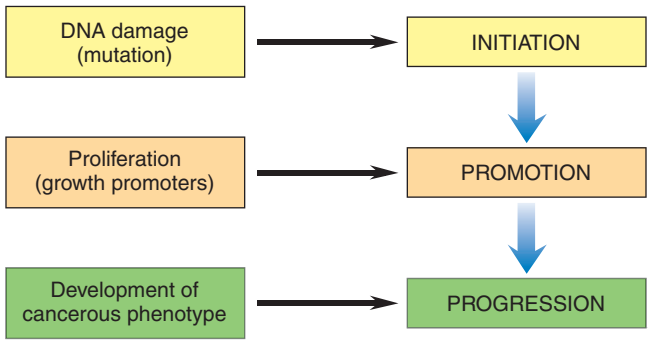


FIG 7.11 Theoretical steps in the development of cancer include initiation, promotion, and progression.

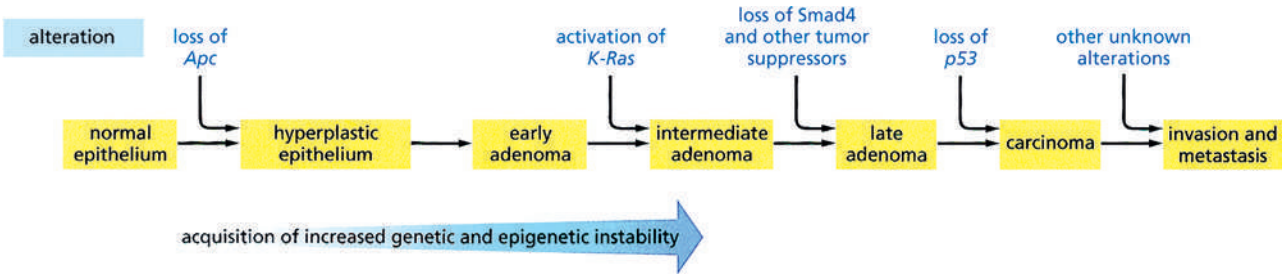


FIG 7.12 The development of colorectal cancer illustrates the concept of multistep carcinogenesis. Derangement of several genes is likely to occur in most types of cancer. (Redrawn from Alberts B et al, editors: *Molecular biology of the cell,* ed 6, New York, 2015, Garland Science, p 1126.)

BOX 7.1 Major Chemical Carcinogens

Direct-Acting Carcinogens

Alkylating Agents

β-Propiolactone
Dimethyl sulfate
Diepoxybutane
Anticancer drugs (cyclophosphamide, chlorambucil, nitrosoureas, and others)

Acylating Agents

1-Acetylimidazole
Dimethylcarbamoyl chloride

Procarcinogens That Require Metabolic Activation

Polycyclic and Heterocyclic Aromatic Hydrocarbons

Benz[a]anthracene
Benzo[a]pyrene
Dibenz[a,h]anthracene
3-Methylcholanthrene
7,12-Dimethylbenz[a]anthracene

Aromatic Amines, Amides, Azo Dyes

- 2-Naphthylamine (β-naphthylamine)
- Benzidine
- 2-Acetylaminofluorene
- Dimethylaminoazobenzene (butter yellow)

Natural Plant and Microbial Products

Aflatoxin B_1
Griseofulvin
Cycasin
Safrole
Betel nuts

Others

Nitrosamine and amides
Vinyl chloride, nickel, chromium
Insecticides, fungicides
Polychlorinated biphenyls

Promotion

Promotion is the stage during which the mutant cell proliferates. The transition from initiation to promotion may involve the activation of another oncogene or the inactivation of a tumor suppressor gene that has kept proliferation in check. Nonmutating factors may also be important in promoting cellular proliferation. Nutritional factors and infection may provide a stimulus for cellular proliferation. As previously described, proliferation is regulated by numerous hormonal growth factors. It is not surprising, then, that hormones may act as promoters of certain types of cancer. The relationship between estrogen hormones and breast, ovarian, and uterine cancer is an important example. Epidemiologic studies indicate that the greater the number of menstrual cycles experienced, the higher the risk of these types of cancer developing. Women with early menarche, late first pregnancy, lack of breast-feeding, and late menopause have a greater risk of developing breast, uterine, and ovarian cancer. This enhanced susceptibility is thought to occur in part because of the greater lifetime estrogen exposure. Estrogen is a trophic hormone for these tissues and may therefore be viewed as having promoter effects. Treatment protocols using antiestrogen agents (tamoxifen) indicate that breast cancer risk may be reduced by blocking the effects of estrogen. However, estrogen is not considered to be carcinogenic and does not cause genetic mutations.

A similar relationship has been identified for prostate cancer and testosterone hormones. In males, testosterone is secreted primarily from the testes under the influence of pituitary gonadotropins. Testosterone is a growth factor for the prostate gland and can act as a promoter of tumor formation in this tissue. This relationship is supported by the fact that therapeutic blocking of testosterone activity in persons with prostate cancer can help shrink the tumor.

Tumor cells frequently produce more than the normal amount of telomerase, an enzyme that allows the cell to repair the ends of the chromosomes (telomeres). Telomere shortening with each cell division normally limits the number of cell cycles in normal cells. Tumor cells must overcome this limit on cell division in order to achieve the immortality required to continue to divide. In the early 1970s it was discovered that normal somatic cells replicate only a finite number of times in culture: fetal cells may replicate approximately 80 times, whereas cells from older individuals divide only 20 or 30 times. Each time a cell divides, it must replicate its DNA, but DNA polymerase is unable to copy the DNA strands all the way to the very tips of the chromosomes (called *telomeres*). The telomere thus shortens slightly with each cell division until some critical length is reached and cell division stops (see Chapter 4). Stem cells and germ cells produce an enzyme called *telomerase* that promotes synthesis of the telomere ends and permits these cell types to replicate indefinitely, but normal somatic cells produce little of this enzyme. Most cancer cells begin to synthesize telomerase while they acquire the malignant phenotype, thus rescuing themselves from critical telomere shortening and gaining a mechanism for achieving immortality. In addition, the majority of cancers are deficient in *p53* activity, which allows them to escape apoptosis despite gross derangements in DNA structure.

Progression

Progression is the stage during which the mutant proliferating cells begin to exhibit malignant behavior. The mutations suffered during initiation are not sufficient to cause all the biochemical changes necessary

for malignant behavior. The proliferating cells are genetically unstable and undergo chance mutations that give them a growth advantage. Clones of mutant cells exhibit a wide variation in phenotype. *Phenotype* refers to the cell's traits, such as morphology, metabolism, and biochemical composition. Cells whose phenotype gives them a growth advantage proliferate more readily. With each cycle of proliferation, an opportunity for chance variation arises. In the end, highly evolved tumor cells are generated that differ significantly from their normal ancestors. These cells have developed characteristics such as the presence of laminin receptors, lytic enzymes, and anchorage independence that enable them to behave malignantly.

Cancer cells often have numerous abnormalities of chromosome structure, and the karyotype can be quite bizarre with bits and pieces of chromosomes attached in the wrong places and extra or missing chromosomes. An example of the chromosomes obtained from a typical cancer cell is shown in Fig. 7.13. The color stains are specific for a particular chromosome, and each chromosome pair should be one color.

The fact that conversion from a normal cell type to a malignant cell type requires multiple steps implies many opportunities to intervene in the process. Prevention of the initiating mutation may be difficult inasmuch as carcinogens are ubiquitous; however, therapies to prevent promotion and progression could render the initial mutation harmless. As the biochemical processes governing promotion and progression become clearer, strategies for blocking these stages continue to be developed.

KEY POINTS

- Full expression of cancer in a host is a multistep process. These steps have been described as initiation, promotion, and progression. The initiating event is usually from genetic mutations. Promotion refers to the stage in which the mutant cell is induced to proliferate. Progression is the stage during which the mutant proliferating cells acquire properties that allow malignant behavior.
- Malignant cells commonly produce telomerase, an enzyme that repairs the telomeres and may be a key for attaining immortality. The majority also have insufficient p53, which allows the tumor cells to escape apoptosis despite DNA damage.

METASTASIS

Metastasis is the process whereby cancer cells escape their tissue of origin and initiate new colonies of cancer in distant sites. For tumor cells to gain access to the blood or lymphatic circulation, they must first escape the basement membrane of the tissue of origin, move through the extracellular space, and penetrate the basement membrane of the vessel (Fig. 7.14). This process is thought to involve loss of cell-to-cell adhesion and binding to matrix components such as laminin via specific laminin receptors on the tumor cell, followed by release of enzymes such as proteases and collagenases that digest the basement membrane. The cancer cell then squeezes through the rift by ameboid movement. The process is repeated at the vessel basement membrane to access the blood or lymphatic vessel. When the cell reaches the tissue to be colonized, it must again traverse the basement membranes by using similar mechanisms (Fig. 7.15). Once in a new tissue setting, the cancer cell colony must acquire nutrients and a blood supply and cope with an environment that may differ considerably from its origin. In general, less differentiated cancer cells are better able to adapt to foreign tissues and survive.

Patterns of Spread

The survival of tumor cells in the circulation is not guaranteed. They may be detected by immune cells and destroyed, or they may undergo apoptosis unless they quickly find a matrix on which to adhere. Fewer than 1 in 10,000 of the cancer cells that enter the circulation survives to form a new tumor at a distant site. Some tumor cell types appear to prefer specific target organs. Sometimes the pattern of metastasis is related to the circulatory flow. For example, metastatic tumors from the colon often seed the liver because they travel within the portal vein. The localization of most metastatic tumors is not so easily explained by blood flow patterns, and some tumor cells appear to "home" to specific targets. This homing tendency is poorly understood but may involve chemotactic signals from the organ to which the tumor cells respond. Cell surface receptors of the integrin and cell adhesion molecule families, which mediate cell-to-matrix and cell-to-cell adhesion, are likely to influence the choice of tissues that cancer cells invade. Dissemination via lymphatics is somewhat more predictable than distribution by blood flow. Generally, the lymph nodes that immediately drain the tissue of cancer origin are colonized first, and then the

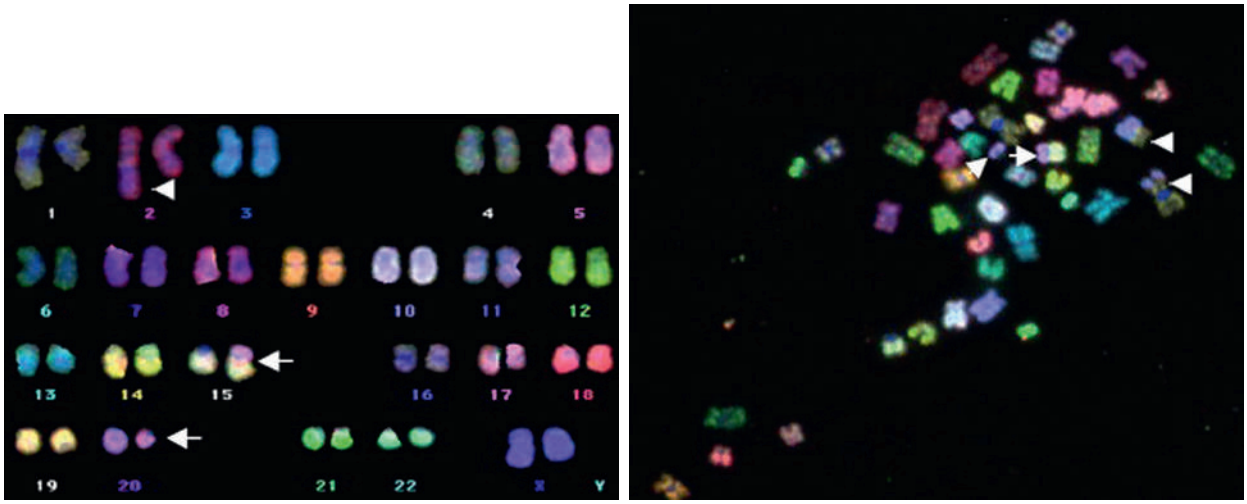


FIG 7.13 Fluorescent images from human colon cancers. (Redrawn from Alberts B et al, editors: *Molecular biology of the cell,* ed 6, New York, 2015, Garland Science, p 1125. Courtesy of Wael Abdel-Rahman and Paul Edwards.)

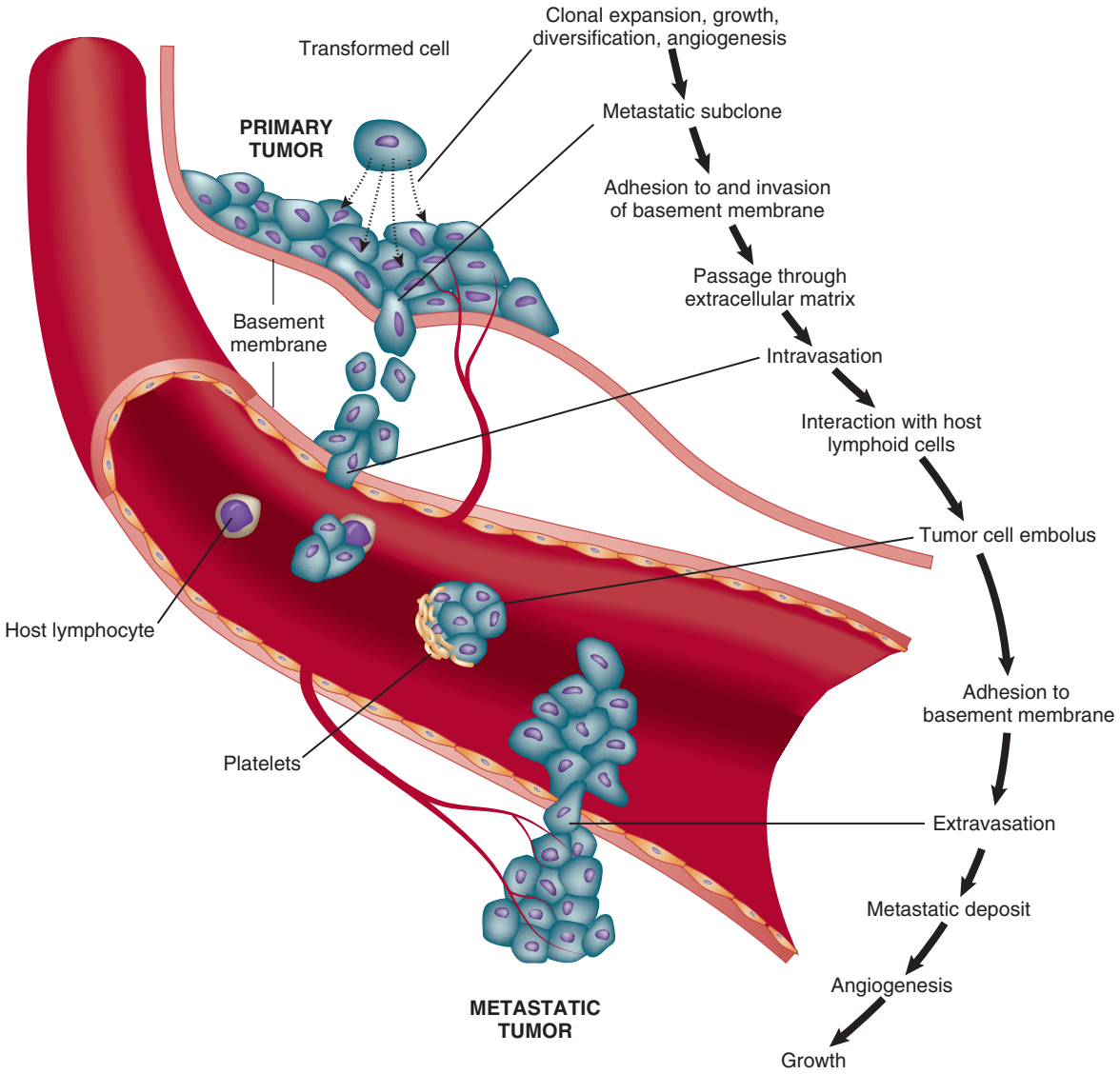


FIG 7.14 The metastatic cascade of events.

tumor cells tend to spread contiguously from node to node. Hodgkin disease, a lymphoma, is particularly noted for its orderly spread via the lymphatics.

Because tumor cells exhibit various degrees of differentiation or resemblance to the parent tissue of origin, it may be difficult to determine the metastatic cancer's tissue of origin. Tumor markers are substances associated with tumor cells that may be helpful in identifying their tissue type. Identification of the tissue of origin has important implications for prognosis and selection of treatment options. Tumor markers rely on the retention of at least some characteristics of the parent tissue type. Some tumor markers are released into the circulation, whereas others must be identified through biopsy of the metastatic tissue. Enzymes and other proteins that are specific to a particular cell type are commonly used as tumor markers. For example, production of thyroglobulin protein is specific for thyroid tumor cells. Melanoma cells express the antigens HMB-45 and S-100, which is helpful in identification as melanocytes. Unfortunately, most tumor markers are not specific for cancer because the normal cells in the tissue of origin also produce them. Tumor markers are most useful as indicators for further diagnostic evaluation and to track the tumor activity. An increasing blood concentration of a specific tumor marker may indicate progression and proliferation of the cancer cells (increased tumor burden). See Table 7.7 for other examples of antigen, hormone, isoenzyme, and immunoglobulin markers used to identify tumor cell types.

Angiogenesis

Tumors cannot enlarge more than about 2 mm in diameter unless they grow blood vessels into the tumor mass to provide oxygen and nutrients. Angiogenesis is the process of forming new blood vessels. Most tumors do not induce angiogenesis until late in the stage of cancer development and so remain small and nonvascularized for years. The triggers that spur the cancer to begin angiogenesis are not completely understood. Tumor cells may begin to produce angiogenic factors such as vascular endothelial growth factor (VEGF) in response to hypoxia or other signals. VEGF stimulates proliferation of vascular endothelial cells, which then migrate to the tumor and orchestrate blood vessel development. Metastatic tumors must also initiate angiogenesis in their new locations or they will not survive. Therefore inhibition of angiogenesis is an important therapeutic goal to limit tumor growth and metastasis and continues to be an area of active research.

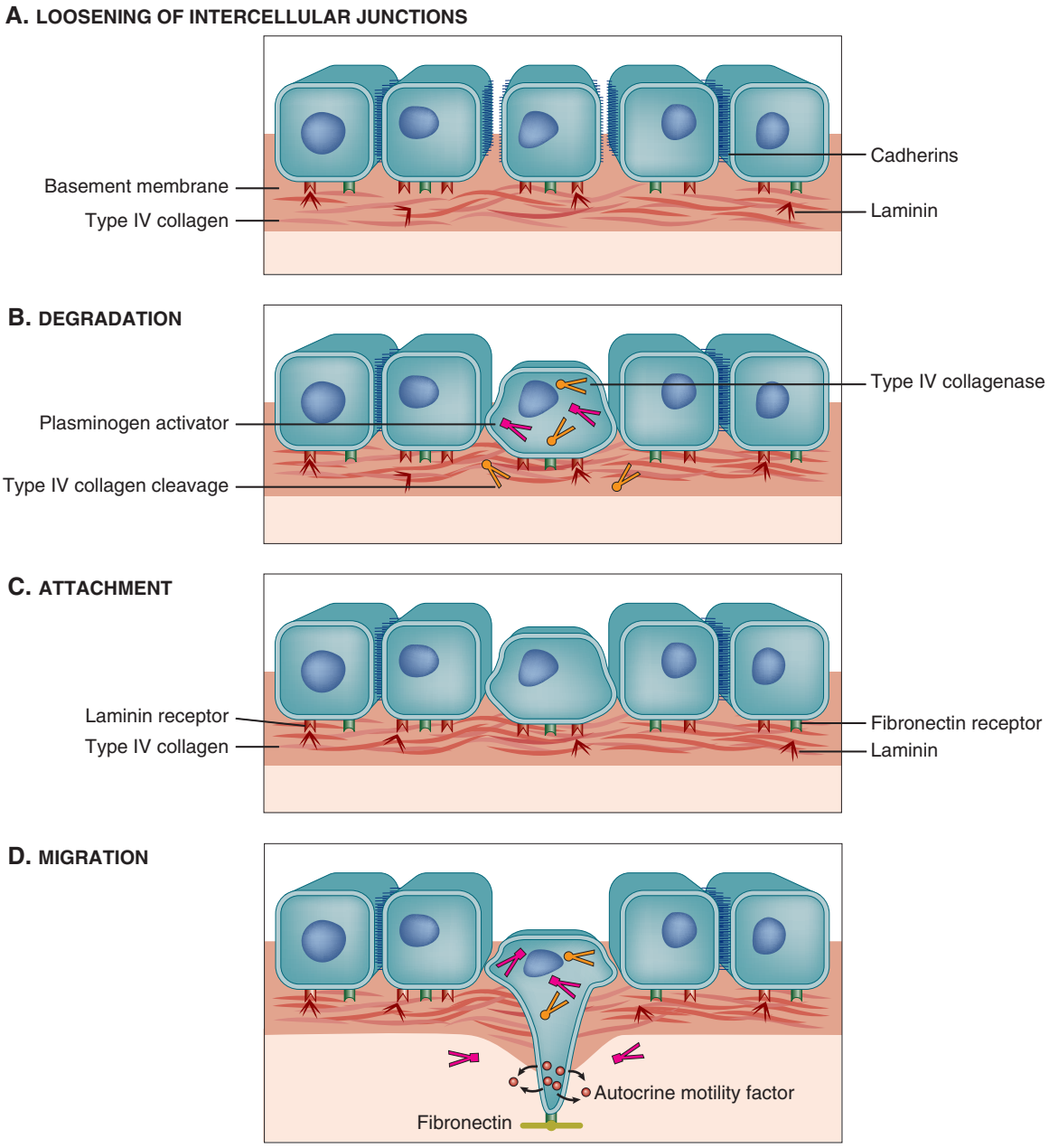


FIG 7.15 Mechanisms of tumor invasion allow tumor cells to escape the site of origin, penetrate the basement membrane, and travel to distant sites. **A,** Tumor cells decrease cell-to-cell attachments via cadherins that allow detachment and migration toward the basement membrane. **B,** Enzymes that degrade proteins are released into the area to form a rift. **C,** The tumor cell migrates away from the site of origin using laminin and fibronectin receptors to pull through the tissue. **D,** Finally the cell moves through a rift in the matrix.

Grading and Staging of Tumors

Grading and staging of tumors are done to predict the clinical behavior of a malignant tumor and to guide therapy. Grading refers to the histologic characterization of tumor cells and is basically a determination of the degree of anaplasia. Most grading systems classify tumors into three or four classes of increasing degrees of malignancy. A greater degree of anaplasia indicates a greater malignant potential. The correlation between the grade of the tumor and its biological behavior is not perfect. Some low-grade tumors have proved to be quite malignant.

The choice of treatment modality is usually influenced more by the stage of the tumor than by its histologic grade. Staging describes the location and pattern of spread of a tumor within the host. Factors such as tumor size, extent of local growth, lymph node and organ involvement, and presence of distant metastases are considered. Several staging systems exist; however, the international TNM (*t*umor, *n*ode, *m*etastasis) system is used extensively as a general framework for staging tumors. Particular staging criteria vary with tumors in different organ systems. An example of staging criteria for breast cancer is shown in Table 7.8.

In the past, tumor staging was based primarily on results of radiography and exploratory surgery. The availability of computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), and other highly sophisticated imaging techniques has revolutionized cancer detection. These imaging modalities allow noninvasive

TABLE 7.7 Selected Tumor Markers

Hormones	
Human chorionic gonadotropin	Trophoblastic tumors, nonseminomatous testicular tumors
Calcitonin	Medullary carcinoma of thyroid
Catecholamine and metabolites	Pheochromocytoma and related tumors
Ectopic hormones	Paraneoplastic syndromes
Oncofetal Antigens	
α-Fetoprotein	Liver cell cancer, nonseminomatous germ cell tumors of testis
Carcinoembryonic antigen	Carcinomas of colon, pancreas, lung, stomach, and heart
Isoenzymes	
Prostatic acid phosphatase	Prostate cancer
Neuron-specific enolase	Small cell cancer of lung, neuroblastoma
Specific Proteins	
Immunoglobulins	Multiple myeloma and other gammopathies
Prostate-specific antigen and prostate-specific membrane antigen	Prostate cancer
Mucins and Other Glycoproteins	
CA-125	Ovarian cancer
CA-19–9	Colon cancer, pancreatic cancer
CA-15–3	Breast cancer
Cell-Free DNA Markers	
p53, APC, and RAS mutants in stool and serum	Colon cancer
p53 and RAS mutants in stool and serum	Pancreatic cancer
p53 and RAS mutants in sputum and serum	Lung cancer
p53 mutants in urine	Bladder cancer

exploration of the tissues of the entire body. The computer-generated images can then be scrutinized for any signs of abnormality that might signal the presence of hidden tumors. CT and MRI rely primarily on detection of differences in tissue density and are therefore not totally specific for tumors. They can, however, guide the selection of sites for exploration and biopsy and potentially reduce unnecessary surgery. PET is a promising staging technology because it facilitates cancer detection based on molecular and biochemical processes within the tumor tissues (Fig. 7.16). PET may be used in certain clinical situations in which CT has known limitations, such as differentiation of benign from malignant lymph nodes or other lesions, differentiation of residual tumor from scar tissue, or detection of unsuspected distant metastases.

Antibodies can also be used to locate cancer cells in the body. Antibodies can be raised against specific antigens present on the surface of tumor cells. The antibodies are also bound to a tracer (e.g., a radioactive isotope such as iodine-125), which can be detected by imaging. Because methods for identifying tumor antigens and raising specific antibodies have improved, this technology provides the potential for finding very small numbers of tumor cells hidden in the body.

The results of the staging procedure will determine which of the mainstays of cancer treatment—surgery, radiation therapy, or chemotherapy—may be used, singly or in combination, to destroy the cancer cells. Localized tumors may be managed with surgery and radiation therapy, whereas evidence of metastasis generally necessitates the addition of chemotherapy.

KEY POINTS

- Malignant cells produce specialized enzymes and receptors to enable them to escape their tissue of origin and metastasize.
- The spread of tumors generally occurs by way of the bloodstream or lymphatics. Tumor cells often lodge in the capillary beds of the organs that drain them, such as liver and lung. Some tumors appear to "home" to certain tissues.
- Grading and staging are done to predict tumor behavior and guide therapy. Grading is the histologic characterization of tumor cells, whereas staging describes the location and pattern of tumor spread within the host.
- The TNM (*t*umor, *n*ode, *m*etastasis) staging system is used to describe the tumor size, lymph nodes affected, and degree of metastasis.

TABLE 7.8 American Joint Committee on Cancer and Union Contre Le Cancer Staging*

Stage	T: Primary Cancer	N: Lymph Nodes	M: Distant Metastasis	10-Year Survival (%)
0	DCIS or LCIS	No metastases	Absent	92
I	Invasive carcinoma ≥2 cm	No metastases or only micrometastases	Absent	87
II	Invasive carcinoma >2 cm	1–3 positive nodes	Absent	65
	Invasive carcinoma >5 cm but ≤5 cm	0–3 positive nodes	Absent	
III	Invasive carcinoma >5 cm	Negative or positive nodes	Absent	40
	Any size invasive carcinoma	≥4 positive nodes	Absent	
	Invasive carcinoma with skin or chest wall involvement or inflammatory carcinoma	Negative or positive nodes	Absent	
IV	Any size invasive carcinoma	Negative or positive lymph nodes	Present	5

DCIS, ductal carcinoma in situ; *LCIS,* lobular carcinoma in situ; *M,* metastasis; *N,* nodes; *T,* tumor.
*The groups listed in the table are based on the characteristics of the primary carcinoma and the axillary lymph nodes. For rare women with involved internal mammary lymph nodes or supraclavicular lymph nodes, there are additional staging criteria.
From Kumar V, Abbas A, Aster JC: Neoplasia. In Kumar V, Abbas A, Aster JC, editors: *Robbins and Cotran pathologic basis of disease,* ed 9, Philadelphia, 2015, Saunders, p 1068.

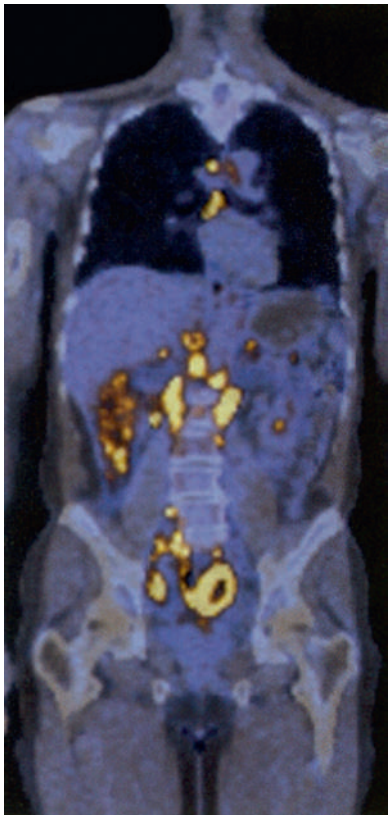


FIG 7.16 PET scan that detects uptake of radioactively labeled glucose is overlaid onto a CT scan background image. The yellow spots in the abdomen and mediastinum are indicative of multiple metastases of non-Hodgkin lymphoma. (From Alberts B et al, editors: *Molecular biology of the cell,* ed 6, New York, 2015, Garland Science, p 1092. Courtesy Dr. Sanjiv Sam Gambhir.)

FIG 7.17 General emaciated appearance in cancer cachexia. (Courtesy Dr. P. Marazzi/Science Photo Library.)

BOX 7.2 Cancer's Seven Warning Signs

Change in bowel or bladder habits
A sore that does not heal
Unusual bleeding or discharge
Thickening or lump in breast or elsewhere
Indigestion or difficulty swallowing
Obvious change in wart or mole
Nagging cough or hoarseness

BOX 7.3 Cancer's Warning Signs in Children

Continued, unexplained weight loss
Headaches with vomiting in the morning
Increased swelling or persistent pain in bones or joints
Lump or mass in abdomen, neck, or elsewhere
Development of whitish appearance in pupil of the eye
Recurrent fevers not caused by infections
Excessive bleeding or bruising
Noticeable paleness or prolonged tiredness

EFFECTS OF CANCER ON THE BODY

The effects of cancer on the host vary widely, depending on the location of the tumor and the extent of metastasis. Early-stage cancer may be asymptomatic. As the tumor increases in size and spreads through the body, a number of symptoms typically become apparent, including pain, cachexia, immune suppression, and infection. Once treatment has begun, patients may also suffer hair loss and sloughing of mucosal membranes. The ACS has published the seven warning signs of cancer as a way of encouraging the public to seek early evaluation of potential cancers (Box 7.2). The presentation of cancer in children differs from that in adults, and special warning signs have been identified for the pediatric population (Box 7.3).

Pain is a common and feared complication of the disease process. Pain may be due to invasion of metastatic cells into organs or bone and subsequent activation of pain and pressure receptors in these tissues. Tissue destruction and inflammation may contribute to cancer pain. Cancer treatment may contribute to overall pain because of procedures requiring biopsy and intravenous drug administration. Pain can usually be controlled through the use of analgesics. The use of patient-controlled analgesia has been effective in reducing patient fears of inadequate therapy for pain (see Chapter 47).

Cachexia refers to an overall weight loss and generalized weakness (Fig. 7.17). Many factors contribute to cancer cachexia, including loss of appetite (anorexia) and increased metabolic rate. Anorexia accompanies many disease processes and may result from toxins released by the cancer cells or immune cells. Cancer patients may have aversions to specific foods and may feel full after only a few bites. Nausea and vomiting are common complications of cancer therapy and contribute to decreased nutrient intake. Despite the minimal nutrient intake, body metabolism remains high. Production of tumor necrosis factor and other immune cytokines is thought to be important in producing the hypermetabolic state. Nutrients are mobilized from fat and protein stores in the body and consumed by the hypermetabolic cells (see Chapter 42). Some patients may require nutritional supplementation by enteral or parenteral routes.

Individuals with cancer often demonstrate deficits in immune system competence. Cancer cells secrete substances that suppress the immune system. Individuals with cancer may have reduced populations of T and B cells and may respond poorly to injected antigens. The mechanisms by which cancer cells depress immune responses are varied, but the prognosis for cancer recovery is poorer when the immune system is depressed. Immune cells, including cytotoxic T cells and natural killer

(NK) cells, are needed to actively detect and destroy cancer cells (see Chapter 9).

Bone marrow suppression contributes to the anemia, leukopenia, and thrombocytopenia that often accompany cancer. Bone marrow suppression may be due to invasion and destruction of blood-forming cells in the bone marrow, poor nutrition, and chemotherapeutic drugs. *Anemia* refers to a deficiency in circulating red blood cells. In addition to decreased production of blood cell precursors in the bone marrow, anemia may result from chronic or acute bleeding. The signs and symptoms of anemia, such as fatigue, increased heart rate, and increased respiratory rate, are related to a decrease in oxygen-carrying capacity.

Leukopenia refers to a decrease in circulating white blood cells (leukocytes). Malignant invasion of the bone marrow is a primary cause of leukopenia, with malnutrition and chemotherapy being contributing factors. A deficiency in white blood cells reduces the patient's ability to fight infection, which is a major cause of morbidity and mortality in cancer patients. Often the offending organism is *opportunistic;* it is unable to infect an immunocompetent host and becomes virulent only when a person is immunocompromised. Infections are quite difficult to manage because the host is unable to mount an effective immune response. Infections are also difficult to prevent because the majority of the infecting organisms are from the patient's own endogenous flora (e.g., skin, gastrointestinal tract). The development of severe leukopenia or infection during treatment may necessitate changes in the chemotherapeutic regimen to allow bone marrow recovery.

Thrombocytopenia is a deficiency in the number of circulating platelets, which are important mediators of blood clotting. Platelet deficiencies predispose to life-threatening hemorrhage. A platelet count of less than 20,000/mm^3 has been associated with spontaneous hemorrhage.

Anemia, leukopenia, and thrombocytopenia may be managed by administration of blood products containing red blood cells, white blood cells, and platelets, respectively. In fact, blood replacement therapy is used more often in cancer patients than in patients with any other medical condition. When chemotherapy is terminated, stem cells in the bone marrow generally recover and the production of blood cells resumes. In some cases, the production of red and white blood cells can be enhanced by treating the patient with specific growth factors, such as erythropoietin (Epogen) or granulocyte-stimulating factors (Neupogen).

Hair loss and the sloughing of mucosal membranes are complications of radiation therapy and chemotherapy. Treatment is designed to kill the rapidly proliferating cancer cells, but normal cells with high growth rates such as mucosal epithelia and hair follicle cells are also damaged. Damaged mucosa is a primary source of cancer pain and anorexia, and may provide a portal for the invasion of organisms from the skin or gastrointestinal tract.

Paraneoplastic syndromes are symptom complexes that cannot be explained by obvious tumor properties and occur in 10% to 15% of patients with cancer. Many of the syndromes are associated with excessive production of hormones or cytokines by the tumor. Common paraneoplastic syndromes include (1) hypercalcemia, (2) Cushing syndrome secondary to excess adrenocorticotropic hormone (ACTH) secretion, and (3) hyponatremia and water overload secondary to excess antidiuretic hormone (syndrome of inappropriate ADH) secretion. Small cell carcinoma of the lung is commonly the culprit for excess ACTH and ADH syndromes. Hypercalcemia (elevated concentration of serum calcium) is a paraneoplastic syndrome associated with abnormal production of parathyroid hormone–related protein by the tumor cells. Unexplained hypercalcemia is regarded as evidence of cancer until proven otherwise. Hypercalcemia may be a consequence of metastatic bone cancer, and in this case it would be an expected finding rather than a paraneoplastic syndrome.

If left untreated, cancer has the potential to kill the host. The cause of death is multifactorial. Infection, hemorrhage, and organ failure are the primary causes of cancer death. The failure of cancer-ridden organs such as the liver, kidney, brain, and lung results in the loss of life-sustaining functions. Treatment for cancer can also be detrimental to the host by contributing to immunosuppression and platelet deficiencies. The cumulative effects of one or more of these factors may lead to death.

KEY POINTS

- Regardless of the type of malignancy, affected individuals exhibit characteristic signs and symptoms, including pain, cachexia, bone marrow suppression, and infection.
- Bone marrow suppression is manifested as anemia, leukopenia, and thrombocytopenia.
- Immunosuppression with consequent infection is a primary cause of cancer-associated death.

CANCER THERAPY

The overall 5-year survival rate for patients with cancer is approximately 68%, with some types of cancer having much higher or lower rates. Early detection of cancer, while it remains localized in the tissue of origin, is associated with the best prognosis for cure. Cure implies eradication of all cancer cells in the body and is different from the 5-year survival rate. Patients with metastatic invasion of regional lymph nodes still have a good opportunity for cure with appropriate therapy. Widespread invasion of multiple tissues and organs is associated with a poor prognosis, and therapy may be aimed at remission or palliation of symptoms rather than cure. The mainstays of cancer therapy are surgery, radiation therapy, and drug therapy. In some hormone-sensitive tumors (breast, prostate), hormonal-blocking drugs may be used.

Immunotherapy and targeted molecular therapies have begun to emerge as important treatments for specific cancers. Traditional forms of treatment are not selective for cancer cells and result in unavoidable damage to normal tissue. The immune system, on the other hand, is noted for its ability to make subtle distinctions between normal and abnormal or foreign cells. Recognition of tumor cells as different from their normal counterparts is the basis of tumor immunology. Recognition depends on the expression of abnormal molecules or antigens on the cancer cell surface. Unfortunately, most tumor-associated antigens are also expressed to some degree on normal cells, which makes it difficult to develop strategies to target cancer cells selectively. However, genetic analysis of a tumor sometimes can give clues to potential and unique therapeutic targets. Ultimately, the choice of treatment depends largely on the results of the staging procedure. A greater degree of metastasis generally requires a more aggressive therapeutic approach.

Surgery

The majority of patients with solid tumors are treated surgically, which can be curative in some localized cancers. The main benefit of surgery is removal of a tumor with minimal damage to other body cells. The surgeon generally removes a margin of normal-appearing tissue around the resected tumor to ensure complete tumor removal. Lymph nodes are subjected to biopsy and also removed if evidence of metastasis is present. Surgical resection of some tumors can be tricky if vital structures such as neurons or blood vessels are involved.

Surgery involves risks related to the effects of anesthesia, infection, and blood loss. The surgical procedure may be disfiguring or may result in loss of function. Surgical resection as the sole treatment for solid

tumors is curative in a minority of patients because most patients already have undetectable metastases at the time of diagnosis. Therefore surgical resection is commonly accompanied by radiation therapy or chemotherapy. Even one remaining cancer cell could be sufficient to reinitiate tumor formation.

Radiation Therapy

Ionizing radiation is used for two principal reasons: to kill tumor cells that are not resectable because of location in a vital or inaccessible area and to kill tumor cells that may have escaped the surgeon's scalpel and remain undetected in the local area. Radiation kills cells by damaging their nuclear DNA. Cells that are rapidly cycling are more susceptible to radiation death because there is little time for DNA repair. Radiation doses generally do not kill cells directly, but cause enough cell damage to initiate apoptosis. The *P53* tumor suppressor gene is an important mediator of this response. Many tumors have mutant *P53* and may be less susceptible to radiation-induced cell death.

It is difficult to kill all the cells of a large tumor by irradiation because they are heterogeneous—they are in different phases of mitosis and are cycling at different rates. A single radiation dose large enough to kill all the tumor cells would be sufficient to kill the normal cells as well. Radiation is often administered in smaller doses over several treatments and is most effective at eradicating small groups of tumor cells. It is often used in combination with surgery. Radiation is also useful for palliative reductions in tumor size. Pain from bone and brain tumors may be effectively managed with radiation therapy that shrinks the tumor. Tumors with bleeding surfaces may be coagulated with radiation to decrease blood loss.

A certain degree of destruction of normal cells in the irradiated field is expected with radiation therapy. Radiation is best used when tumor cells are regionally located. Total-body irradiation to kill tumor cells in disseminated locations is not recommended because of the likelihood of life-threatening tissue damage, although it may be used in preparation for bone marrow or peripheral stem cell transplantation.

Drug Therapy

Chemotherapy generally refers to the systemic administration of anticancer chemicals as treatment for cancers that are known or suspected to be disseminated in the body. Unlike surgery or radiation therapy, which is locally or regionally applied, parenterally administered chemotherapeutic drugs can find their cancer cell targets in areas throughout the entire body.

Most chemotherapeutic agents are cytotoxic because they interfere with some aspect of cell division. The more rapidly dividing cells are more susceptible to the killing effects of chemotherapeutic agents. In a large tumor mass, the rates of cell division are highly diverse, with many slowly dividing cells. At any one time, only a portion of the tumor cells are in a cell cycle stage that is susceptible to chemotherapy. Several courses of chemotherapy are generally necessary to ensure that all tumor cells have been killed. It is difficult to kill slowly cycling tumor cells without also killing normal cells that are cycling at approximately the same rate. Small tumors are easier to eradicate because rates of cell division are generally faster. To prevent relapse, the "stem" cells that develop into clones of malignant cells must be destroyed. Unfortunately, stem cells may not divide as rapidly as other cells. Resection or irradiation to reduce tumor size may prompt the stem cells to divide, thus making them more susceptible to chemotherapy. Tumor cells with mutations of the *P53* gene may be resistant to chemotherapeutic agents that work by damaging DNA, so drugs that act by interfering with the cancer cell cycle in other ways may be more effective.

Chemotherapeutic agents are not selective for tumor cells, and a certain amount of normal cell death also occurs. Rapidly dividing cells, particularly those of the bone marrow, intestinal epithelia, and hair follicles, are most affected. Bone marrow depression is a most serious side effect inasmuch as it predisposes the patient to anemia, bleeding, and infection.

New approaches to cancer drug therapy have emerged that indirectly inhibit tumors rather than seeking to eradicate tumor cells directly. One approach is to interrupt the tumor's blood supply. To proliferate, solid tumors must be supplied by a progressively expanding network of capillaries. The development of new capillaries, called *angiogenesis,* is accomplished by migration and growth of endothelial cells. Antiangiogenic drugs block the development of new capillaries.

Immunotherapy

Harnessing the power of the immune system to fight cancer is a particularly appealing idea because of the potential for *specificity.* Current modes of immunomodulation primarily involve the use of interferons, interleukins, and monoclonal antibodies. These therapies are generally used as adjuncts to surgery, irradiation, and chemotherapy.

Interferons are glycoproteins produced by immune cells in response to viral infection. Interferons inhibit cell proliferation and are stimulatory to NK cells, T cells, and macrophages. Interferon-α has been used successfully to treat hairy cell leukemia (a rare B-cell malignancy), chronic myelogenous leukemia, and multiple myeloma. Interferon therapy produces symptoms similar to those of a viral infection: fever, chills, and muscle aches.

Interleukins are peptides produced and secreted by white blood cells. They are also called *lymphokines* or *cytokines.* Interleukin-2 (IL-2) is an important cytokine secreted by activated T helper cells. It stimulates the proliferation of T cells, NK cells, and macrophages. IL-2 can be used to stimulate the growth of these immune cells in culture. Immune cells taken from a patient's blood can be grown in culture in the presence of IL-2. Then the greatly expanded number of immune cells can be given back to the patient, along with intravenous infusions of IL-2. Such treatment has been associated with regression of some tumors (melanoma, renal cell carcinoma). However, because IL-2 toxicity is high and many individuals have severe allergic reactions, the benefit of therapy must be weighed against the risks for each individual situation.

The use of monoclonal antibodies (antibodies having identical structure) in cancer therapy is currently the subject of intense investigation. Monoclonal antibodies specifically bind with target antigens and can therefore be used in several ways as treatment for cancer. Antibodies can be used to deliver a cytotoxic drug preferentially to the cancer cell and thus minimize drug interactions with normal cells. Similarly, antibodies can be used to direct other cytotoxic cells, such as NK and T cells, to tumor cells lurking in the body. Antibodies can be attached to a radioactive label and injected into a patient to screen for recurrence of tumor growth. Antibodies can also be directed against cells that support tumor growth.

Monoclonal antibodies have been developed for management of several cancers. For example, nearly 25% of breast cancers have overexpression of the HER2 receptor on the surface of malignant cells. The monoclonal antibody trastuzumab specifically binds to this HER2 protein and helps immune cells find and kill the tumor cells. A summary of selected monoclonal antibody agents and their main tumor protein targets is shown in Fig. 7.18.

Gene and Molecular Therapy

Because cancer is fundamentally a disorder of gene function, the use of gene therapy to alter the malignant behavior of cells may have high therapeutic potential. As specific gene derangements are identified for particular tumors, gene therapy may be used to suppress overactive

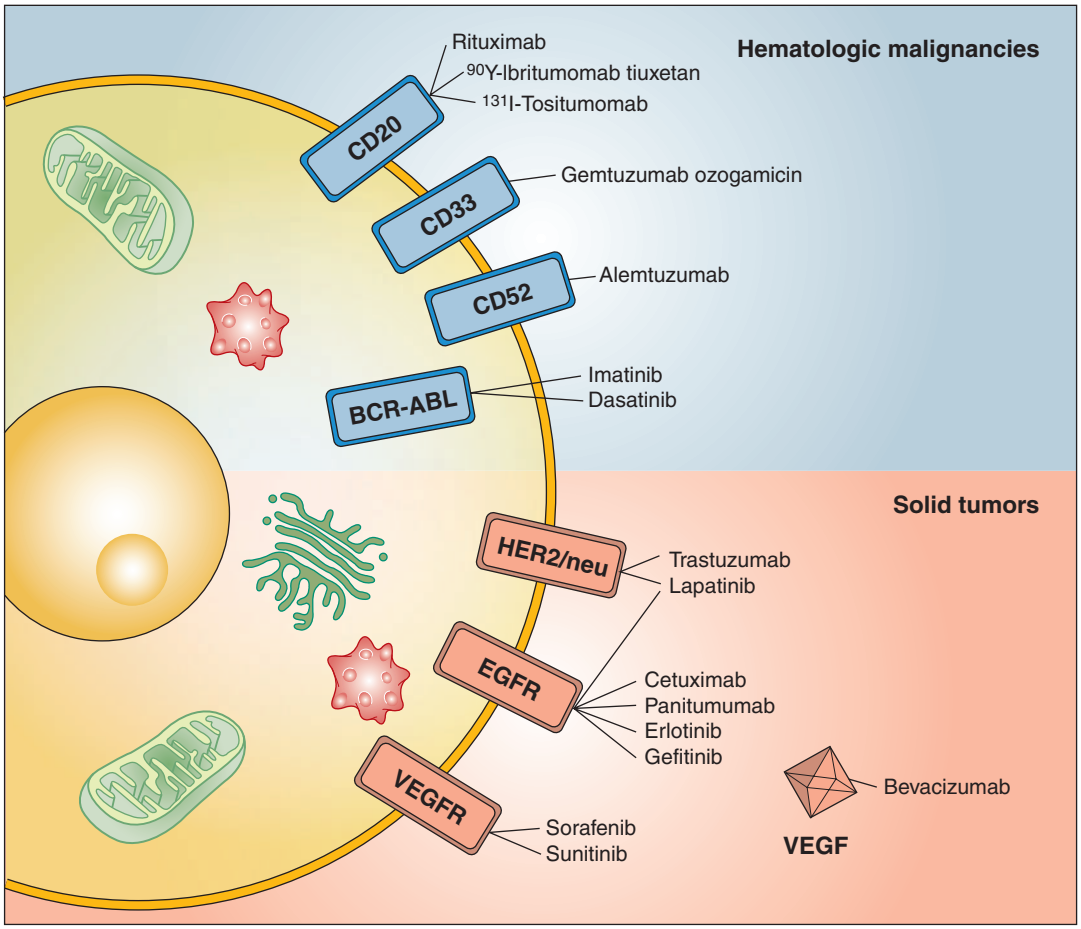


FIG 7.18 Cancer cells express abnormal antigens (tumor-associated antigens) on their cell surface that can activate immune cells or be used as targets for monoclonal antibodies. Numerous medications are now available that use monoclonal antibodies to target cellular proteins relevant to several different types of cancer.

oncogenes or replenish missing tumor suppressor function. Current uses of gene therapy for cancer include genetic alteration of tumor cells to make them more susceptible to cytotoxic agents or immune recognition, and genetic alteration of immune cells to make them more efficient killers of tumor cells.

Tumor cells can also be made more recognizable to immune cells by insertion of genes that cause the tumor cells to express "foreign" proteins on their cell surface. This type of gene therapy has shown some benefit in melanoma and renal carcinoma. Replacement of genes for *P53* is an attractive therapy because tumor cells would be more susceptible to apoptosis. Gene replacement of other tumor suppressors such as *pRb* or *APC* in those tumors that are deficient could help inhibit tumor proliferation.

Gene therapy can be directed at cells other than tumor cells to enhance the body's cancer defenses. One such approach involves harvesting immune cells from the cancer patient, inserting IL-2 genes, and then returning the genetically enhanced immune cells to the patient. The enhanced immune cells attack the tumor cells more vigorously than normal immune cells do and have been shown to persist in the body for 6 months or longer.

At present, gene therapy is limited by difficulty in delivering the new genes to the target cells. As methods improve, gene therapy will become an increasingly important part of cancer prevention and management.

Molecular therapies that target cytoplasmic signaling pathways have also been developed. For example, in chronic myelogenous leukemia a chromosomal rearrangement results in the abnormal production of an enzyme, BCR/ABL. This enzyme stimulates cell proliferation and contributes to the overproduction of leukemic cells. An agent that specifically inhibits this enzyme (imatinib) has dramatically improved the management of this disease. Other drugs that specifically target abnormal tumor characteristics continue to be developed.

Stem Cell Transplantation

Transplantation of hematologic stem cells is used to manage life-threatening disorders in which the patient's bone marrow is incapable of manufacturing white blood cells, red blood cells, or platelets. Most often, nonfunctional marrow is a consequence of the high-dose chemotherapy and radiation used to manage hematologic malignancies such as leukemia and lymphoma. Stem cell transplantation also has been applied to other malignancies (e.g., breast cancer) and to non-malignant disorders (e.g., aplastic anemia, sickle cell anemia, and thalassemia). Stem cells can be harvested from aspirates of bone marrow or from the donor's peripheral bloodstream. Bone marrow is rich in stem cells, but the peripheral blood is poor. The stem cell donor can be a tissue-matched individual (allogeneic), an identical twin (syngeneic), or the patient in question (autologous). A closer match between donor and recipient is associated with a better outcome.

Before infusion of donor stem cells, the patient's own immune cells must be suppressed to prevent transplant rejection. It is also necessary to eliminate any residual malignant cells from the body to avoid relapse of the cancer. Both of these objectives can be accomplished through high-dose chemotherapy and total-body irradiation regimens, which leave the patient susceptible to severe anemia, infection, and bleeding. The therapeutic goal of stem cell transplantation is to restore immune and hematopoietic function. It may take weeks to months for the infused stem cells to reestablish themselves and begin to proliferate in their new host. During this time, the transplant recipient requires intensive monitoring and management of complications.

The success of stem cell transplantation depends on a number of factors, including the age of the patient, closeness of tissue matching, stage of cancer, and general health status of the patient before transplantation. Transplantation is an expensive undertaking, but may significantly improve disease survival rates in some malignancies.

KEY POINTS

- Early detection of cancer while it remains localized is associated with the best prognosis for cure. The overall 5-year survival rate for patients with cancer is about 68%.
- The mainstays of cancer therapy are surgery, radiation therapy, and chemotherapy. Surgery and radiation therapy are effective for cancers that are localized. Chemotherapy is usually the treatment of choice for cancers known or suspected to be disseminated in the body.
- Cells that divide rapidly are the most susceptible to damage from radiation therapy or chemotherapy. However, in addition to cancer cells, rapidly dividing normal cells may be killed. Cells of the bone marrow, hair follicles, and gastrointestinal mucosa are particularly susceptible.
- Immunotherapy has the potential to specifically target cancer cells. Numerous monoclonal antibodies have been developed to boost the immune system's ability to locate and destroy cancer cells.
- Gene and molecular therapy may be used to alter cancer cells to suppress oncogenes, enhance tumor suppressor genes, make tumor cells more susceptible to cytotoxic agents, or interfere with the function of cancer gene products.
- Transplantation of hematopoietic stem cells is an important adjunct to cancer therapy that provides a method to restore bone marrow function after high-dose irradiation or chemotherapy.

SUMMARY

Neoplasia is abnormal cell proliferation of a benign or malignant nature. Benign tumors resemble their parent cells and are strictly local, whereas malignant tumors are anaplastic, invade local tissues, and may spread to distant sites (metastasize). The most important consideration for cancer management is the degree of cancer spread in the body, which can be determined by staging procedures. Cancer is managed by surgical removal, radiation therapy, chemotherapy, and immunotherapy.

Cancer cells have complex relationships with the host. The host immune system is capable of—but not always successful in—recognizing and killing cancer cells. Cancer cells exert immunosuppressive effects on the host and eventually cause pain, cachexia, and bone marrow suppression. If untreated, cancer has the potential to kill the host by multifactorial processes, including infection, hemorrhage, and organ failure. If treated, cancer has an overall 5-year survival rate of approximately 68%.

Cancer is an outcome of genetic predispositions and environmental carcinogens. Tobacco use and improper nutrition are the two most studied carcinogenic lifestyle factors. Tobacco is clearly carcinogenic through its ability to cause genetic damage and to promote the growth of mutant cells.

Cancer is thought to develop when proto-oncogenes become inappropriately overactivated in the cell or tumor suppressor genes become inactivated. This change in activation is usually due to a mutational event in the cell's DNA. Oncogenes are believed to disrupt intercellular communication, which normally exerts growth-controlling effects on the cell. This disruption is accomplished primarily through the production of abnormal growth factors, growth factor receptors, cytoplasmic signaling molecules, or nuclear transcription factors that allow the cancer cell to manufacture its own growth-promoting signals. The tumor suppressor genes *Rb* and *P53* are important inhibitors of cell replication. The Rb protein binds and sequesters transcription factors, whereas p53 monitors the integrity of cellular DNA and may initiate apoptosis (cell suicide) when significant cell damage occurs. Loss of tumor suppressor gene function occurs in most cancers.

RESOURCES

Cancer Epidemiology

Albanes D, et al: Effects of alpha-tocopherol beta-carotene cancer prevention study. *Am J Clin Nutr* 61:S1427–S1430, 1995.

Alberts B, et al: Cancer. In Alberts B, et al, editors: *Molecular biology of the cell*, ed 6, New York, 2015, Garland Science, pp 1091–1144.

American Cancer Society: *Cancer facts and figures—2015*, Atlanta, GA, 2015, Author.

American Cancer Society: *Cancer prevention and early detection facts & figures 2015-2016*, Atlanta, GA, 2015, Author.

Goodman M, Bostick RM, Kucuk O, Jones DP: Clinical trials of antioxidants as cancer prevention agents: past, present, and future. *Free Radic Biol Med* 51(5):1068–1084, 2011.

Kumar V, Abbas A, Aster JC: Neoplasia. In Kumar V, Abbas A, Aster JC, editors: *Robbins and Cotran pathologic basis of disease*, ed 9, Philadelphia, 2015, Saunders, pp 265–340.

Omenn GS, et al: Risk factors for lung cancer and for intervention effects in CARET, the beta-carotene and retinol efficacy trial. *J Natl Cancer Inst* 88(21):1550–1559, 1996.

Cancer Biology

Cuzick J, DeCensi A, Arun B, et al: Preventive therapy for breast cancer: a consensus statement. *Lancet Oncol* 12(5):496–503, 2011.

Dulbecco R: Cell transformation by viruses. *Science* 166:962–968, 1969.

Hayflick L: The biology of human aging. *Adv Pathobiol* 7(2):80–99, 1980.

McCann J, Ames BN: Detection of carcinogens as mutagens in the *Salmonella*/microsome test: assay for 300 chemicals: discussion. *Proc Natl Acad Sci USA* 73:950–955, 1976.

Paradiso A, Formenti S: Hereditary breast cancer: clinical features and risk reduction strategies. *Ann Oncol* 22(Suppl 1):I31–I36, 2011.

Weinberg RA: Tumor suppressor genes. *Science* 254:1138–1146, 1991.

8

Infectious Processes

Brent A. Banasik

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- How does the human microbiome regulate the environment of mucosal surfaces and protect from pathogens?
- What are opportunistic infections, and when do they develop?
- What conditions compromise host defenses against microorganisms?
- What factors influence the virulence of infectious agents?
- How do environmental factors such as climate, poverty, global travel, and bioterrorism influence the risk of infectious diseases?
- What are the important microorganisms that affect humans, including bacteria, viruses, fungi, and parasites, and how do differences in structure and life cycle affect infectious processes?

CHAPTER OUTLINE

Host–Microbe Relationship, 142
The Human Microbiome, 142
Host Characteristics, 142
Physical and Mechanical Barriers, 142
Impaired Immune Function, 143
Pathogen Characteristics, 144
Adherence and Invasion, 144
Bacterial Endotoxin, 144
Bacterial Exotoxins and Enzymes, 145
Evasion of Immune Cells, 145
Endospore Formation, 145
Antimicrobial Resistance, 145
Transmission of Infection, 147
Routes of Transmission, 148
Emerging Infectious Diseases, 148
Weapons of Bioterrorism, 149
Types of Pathogenic Organisms, 149
Bacteria, 149
Viruses, 150
Fungi, 151
Parasites, 152

Infectious diseases can be caused by a variety of pathogenic microorganisms when they gain access to the body through skin and mucous membranes. The importance of maintaining the integrity of these surface barriers and protecting them from contamination has been an important principle of infection control since the days of Florence Nightingale. Only recently have researchers come to realize that the human body is teeming with a staggering number of diverse microorganisms that outnumber the cells in the human body by more than tenfold! This so-called *microbiome* is important for protection against transient pathogens, but is also a ready source of *opportunistic* organisms should the immune system be altered.

A variety of pathogens inhabit different environments such as hospitals, the food supply, water, animals, and humans. Globalization of the world's population and extensive travel by air have had major implications for the worldwide spread of infectious agents, sometimes even before the infected individual becomes symptomatic or identifiable.

Medications that inhibit the immune response in populations such as transplant or cancer patients and the increasing use of immunomodulators for treatment of diseases like rheumatoid arthritis increase the likelihood of infections. Excessive use of antibiotics in humans and domesticated animals has contributed to the emergence

of treatment-resistant infections. Health care professionals have a vital role in the prevention, early detection, and management of infections.

HOST–MICROBE RELATIONSHIP

Microorganisms interact with human hosts in different ways depending on host and microorganism characteristics. Some microorganisms may be encountered in food or water and are of little consequence as they pass through the host without causing harm and without establishing residence in the host. These are called *transient microorganisms* (Table 8.1). The normal flora that have established long-term residency with the host but do not cause harm, and may even benefit the host, are called *commensal.* Sometimes commensal organisms may cause opportunistic infections when they gain access to parts of the body where they do not belong or when the host's immune system is depressed. Other *opportunists* may be low-virulence, ubiquitous types of organisms that do not cause disease in humans with normal immune systems and only occur in the immunosuppressed. Microorganisms that frequently cause diseases in humans with normal immune defenses are called *pathogens.* Human pathogens are passed from one human reservoir to another human, either directly or through a fomite (e.g., contaminated instruments, contaminated food). Sometimes humans come into contact with pathogens that do not usually affect humans, through insect or animal bites, for example. These *accidental* pathogens (see Table 8.1) may cause severe disease in humans.

The Human Microbiome

The relationship between the host and the commensal organisms that live in residence is a complex and diverse one that could affect and be affected by diet, hormones, medications, and metabolism. Each individual is therefore likely to have a different microbiome depending on where they live, what they eat and drink, what medications they take, their genetics, and with whom they come into contact. Understanding the microbiome ecosystem and finding ways to modulate it to advantage is an active area of research. Not all commensal organisms are bacteria, although the most research has been on the microbiome of the gastrointestinal tract, which contains numerous bacterial species. The majority of gut microbiota are known phyla, such as Actinobacteria, Firmicutes, and Bacteroidetes; however, about 25% of an individual's diverse niches throughout the body are unknown. Recent research suggests that the average individual houses 1000 to 1150 different bacterial species and contains more than 100-fold as many genes than the human host's genome. Bacteria have been detected in fetal umbilical cord, meconium, and amniotic fluid, suggesting that the complex colonization of commensal bacteria develops prebirth and not just during the birthing process. Commensal microbiota development in infants after delivery is quite rapid and dependent on many sources, including introduction of maternal microbiota via feces from the gastrointestinal tract through vaginal delivery; microflora from cesarean section and breast milk; and environmental inoculum from the air, hospital staff, and eventually from consuming solid foods.

Often recognized as a gram-negative pathogen, *Escherichia coli* was determined to be among the first commensal microbes that inhabit the infant gut and contribute to broad-spectrum immunity and reduced infections. The harmony of human microbiota can be drastically affected by probiotic or antibiotic usage. Traditionally, probiotics have been administered as gastrointestinal remedies, but recent research has linked them to prevention and treatment of many diseases that include immune and atopic disease, obesity, and even diabetes. It is commonly accepted that antibiotics disrupt gut microflora and provide a niche for enteric pathogens to take root in their absence. In addition, recent research suggests that indigenous commensal bacteria unexpectedly translocated across the gut epithelium in the presence of antibiotics induced an inflammatory response and epithelial injury that led to pathogen growth.

The development of infection in a host depends on complex relationships among environmental exposure, host susceptibility, and microbial virulence (Fig. 8.1).

TABLE 8.1 Definitions of Microorganism–Host Interactions

Transient	A microorganism that is encountered in food or elsewhere in our environment. In general, it is just "passing through" and of little consequence; however, regular encounters over extended periods might lead to host adaptation or even dependence.
Commensal (literally, those that "eat at the same table")	A microorganism that is a normal inhabitant of the human body. In commensal relationships, either the microbe or host derives benefit; in mutualistic relationships, both derive benefit.
Pathogen (derived from the Greek, *pathos,* meaning the "birth of suffering")	A microbe that may or may not be a member of the indigenous microbiota, but it regularly causes disease in apparently normal individuals.
Opportunistic pathogen	A microbe that causes disease *only* in humans who are in some way compromised in their normal defense mechanisms.
Accidental pathogen	A microorganism that is encountered by accidental contact with animals, insects, or the environment. These microorganisms are often deadly in humans and sometimes the causative agent of disease in other animals. These microbes are often distinguished from human-specific pathogens because they are not directly or readily transmissible from human to human.

Host Characteristics

The host's defense system is responsible for identifying, repelling, and eliminating pathogens while minimizing injury to its own tissues. Innate cells that recognize a wide variety of pathogens on first exposure accomplish initial responses to pathogens. Innate responses are generally most efficient against bacterial and parasitic infections, but not effective against viral antigens on first exposure. The specific cells of the immune system—B cells and T cells—are able to respond to viral antigens as well as other foreign invaders, but generally take days to weeks to become effective. With subsequent exposure these cells are able to mount such a robust response that the host may not become ill. Innate and specific immune responses can be modified by the characteristics of the host, including genetics; age; metabolism; and anatomic, physiologic, and environmental factors (Table 8.2). Host immune systems are discussed in greater detail in Chapter 9.

Physical and Mechanical Barriers

Intact physical barriers act as a blockade to foreign material entering the body (Fig. 8.2). Epithelial cells of the skin and those that line the gastrointestinal, genitourinary, and respiratory tracts form sheets that provide a physical barrier to invasion. The constant shedding of the epidermis and mucosal membranes aids in the removal of any microorganisms that attach to their surfaces. In addition, the high fat content

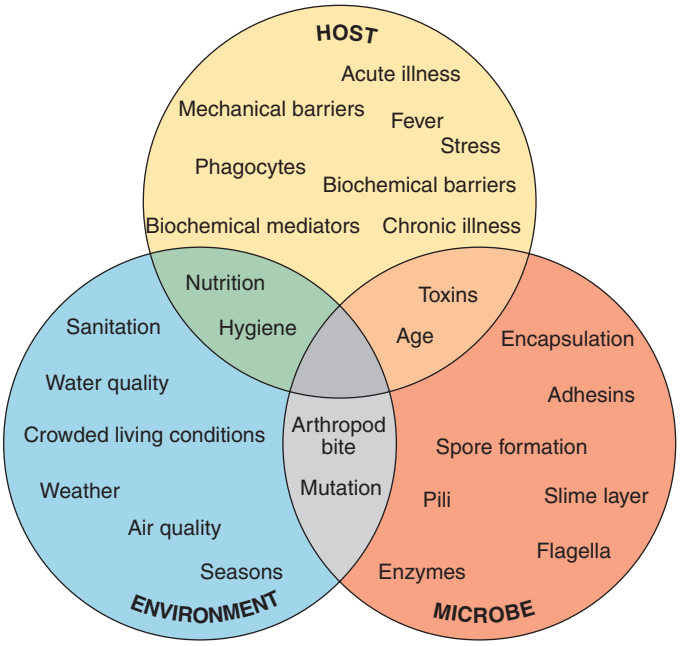


FIG 8.1 This depiction of the interactions of host, microbe, and environment provides a framework for understanding infectious processes.

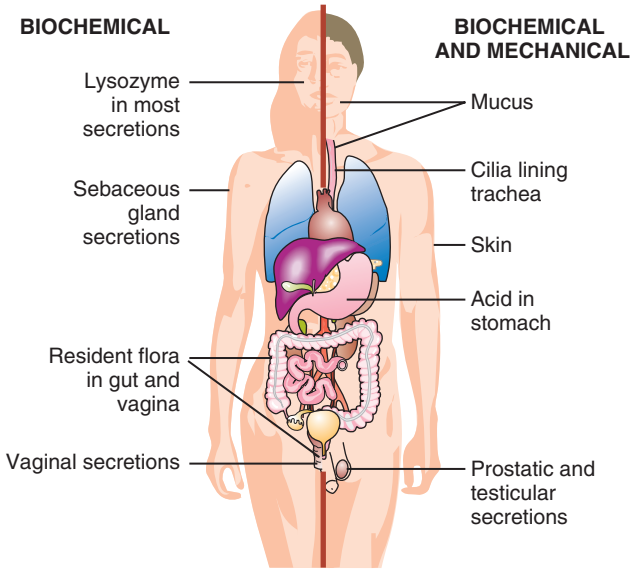


FIG 8.2 Some of the mechanical and biochemical barriers of the human body.

TABLE 8.2 Host Characteristics Influencing Infection

Exposure	Host Health
Animals, humans, insects, parasites	**Intact immune system**
Environmental	Absence of chronic disease such as diabetes
Water, toxins, pollutants, radiation, sewage	Absence of genetic abnormality
Hygiene	**Nutritional Status**
Toileting, handwashing, dental care, bathing	Adequate intake of proteins, vitamins, and minerals
Social Behaviors	**Antibiotic Exposure**
Illicit drug use, alcohol, smoking, risk taking: sexual, sports	Recent use, noncompletion of course
Travel	Development of resistance or allergy
Exposure to vectors in undeveloped countries	

of the skin inhibits the growth of bacteria and fungi. The mucous membranes are covered with sticky mucus that traps microorganisms so that they can be removed efficiently. In the lungs, the mucociliary transport system constantly sweeps trapped organisms up and out of the lungs. In the urinary tract, the frequent flow of urine helps mechanically sweep out bacteria from the urethra.

Biochemical barriers enhance the effectiveness of the mechanical barriers. The acidic environment of the skin, urine, and vagina inhibits bacterial growth. The secretion of acid into the stomach (with a pH of 1 to 2) results in the killing of many microorganisms. Saliva, mucus, tears, and sweat contain antimicrobial chemicals such as lysozyme, an enzyme that destroys cell walls of gram-positive bacteria. Lactoferrin is a mucosal protein that keeps bacterial replication low by reducing the availability of free iron needed for bacterial growth. Sebaceous gland secretions act as antifungals. Immunoglobulin A (IgA) is found in many of the body's secretions, where it acts to bind up and prevent entry of microorganisms through mucous membranes. (See Chapter 9 for a discussion about immunoglobulins.)

Removal or degradation of the body's mechanical and biochemical barriers creates a setting in which infection is likely. For example, burn victims who have lost portions of their skin barrier are at high risk for infection. Destruction and paralysis of the mucociliary transport system of the lung by smoke inhalation or by insertion of breathing tubes increases the likelihood that pathogens can cause pneumonia. Hospitalized patients who have incisions or intravenous and urinary catheters are at risk for infection because their skin barrier has been breached. When a urinary catheter is in place, flushing of bacteria from the urinary tract is impaired. Procedures or diseases that allow organisms to leak from the bowel into the peritoneal cavity are a particularly serious breach of a barrier that can result in life-threatening peritonitis.

Impaired Immune Function

Impaired function of the immune system is an important risk factor for the development of both opportunistic and pathogenic infections. Immunosuppressive drugs used in the treatment of cancer, autoimmune disorders, and inflammatory disorders are common culprits. However, many other factors can reduce a person's ability to resist infection, including poor nutritional status, young or old age, and chronic illness. A person who is at high risk for infection must be vigilant to maintain physical barriers and avoid exposures to pathogens through environmental controls, including avoidance of infected persons and diligent handwashing. Severely immunocompromised individuals may require isolation until their immune status improves.

Nutritional status. The World Health Organization (WHO) reports that inadequate nutrition is a general risk factor for disease and is a particularly important factor in tuberculosis (TB) infection. Protein-energy malnutrition is associated with defects in cell-mediated immunity (specific), impaired intracellular destruction by neutrophils, reduced complement activity, and decreased levels of secretory IgA. At the time of an illness, nutrition is negatively affected by decreased appetite, malabsorption, diarrhea, and diversion of nutrients for immune responses, further exacerbating a malnourished state. The presence of fever increases the metabolic rate, requiring more energy and micronutrients. Protein

malnutrition frequently occurs because of the lack of available clean and safe food.

Micronutrients are also important in immune function. Vitamin A contributes to maintenance of the epithelium, and vitamin E is an antioxidant that supports the development and function of T cells. Vitamin D supports innate and adaptive immunity and promotes immune protection against TB by increasing macrophage defenses. Decreased availability of zinc results in a reduction in neutrophil and natural killer cell function, complement activity, and lymphocyte activity. Iron deficiency is associated with decreased cell-mediated immunity and reductions in neutrophil action. Iron deficiency is one of the most common deficiencies worldwide.

Age. Age affects immune responsiveness as well as the likelihood of exposures to infectious agents. Viruses such as mumps, polio, or Epstein–Barr virus (EBV) cause less severe infection in infants, whereas others such as rotaviruses result in severe illness in infants. These age-related factors may reflect the availability of immune factors or the maturity of the immune system. T-cell function appears to be impaired in newborns, with the majority of the immune response provided by maternal IgG that crossed the placenta. As this immunity fades over approximately the first 6 months of life, these newborns are more at risk of serious infection. Those infants who are breast feeding will obtain secretory IgA, which does provide additional protection. Immunity tends to decline in the elderly, resulting in reduced antibody responses to new antigens. The immune system may become dysregulated in the elderly, as evidenced by the increased frequency of autoimmune diseases with age.

Chronic illness. Chronic illnesses such as diabetes, cancer, heart disease, and renal failure are associated with an increased risk of infection. Deaths in patients with chronic illnesses are frequently directly related to an infectious process. It is difficult to separate the various immune-suppressing components that accompany chronic illness because malnutrition and medications may play a part. Specific chronic illnesses may create a milieu in which pathogens take hold. For example, patients with hyperglycemia often develop genitourinary infections, in part because of the high glucose in the urine. Poor circulation to an area because of vascular disease can also limit the ability of the tissue to withstand infection and may impair the ability of immune cells to gain access to the area. The physiologic response to chronic stress is thought to contribute to risk for infection. Neurohormonal alterations lead to changes in the levels of cortisol and other stress hormones that may affect immune responsiveness (see Chapter 2).

Patients with poorly functioning or inadequate numbers of immune cells are particularly at risk for a wide variety of infectious diseases. Infection with opportunistic organisms that are usually not pathogenic is an indicator that a person is immunocompromised. The immunocompromised status might be a result of genetic disorders (such as hypogammaglobulinemia) or may be acquired (such as in patients who have undergone a splenectomy or have human immunodeficiency virus [HIV]). The prevalence of people with medication-induced immunosuppression is increasing. Antirejection medications such as corticosteroids, cyclosporine, and tacrolimus contribute to immuno-suppression in transplant patients. People who take corticosteroids and immunomodulators to suppress inflammatory and autoimmune diseases are also at high risk of infection. HIV disease is an important chronic condition that predisposes to a variety of other infections (see Chapter 12).

Immunization status. Immunization is the most effective means to reduce morbidity and mortality from vaccine-preventable diseases, such as hepatitis A, hepatitis B, influenza, and pneumococcal infections. There are two goals of immunization: The first is to confer immunity to a host by direct exposure to the altered pathogen without causing disease. The second is to decrease the number of susceptible hosts in the population, thereby limiting the possibility of transmission of the disease. By this method, known as *herd immunity,* the disease can be controlled or eliminated without immunizing everyone, as long as a high enough percentage of the population is immunized. The number of susceptible hosts in the population decreases as immunization rates increase. When a critical percentage of the susceptible population is immunized, disease outbreaks can be averted. If the percentage of adequately immunized individuals drops, epidemics may result.

Depending on the immunogenicity of the vaccine, several doses spaced at intervals may be necessary to allow for continued development of an antibody level sufficient to prevent disease. Recent outbreaks of whooping cough (pertussis) and measles have occurred in the United States in areas where the immunization rate has dropped. Perhaps predictably, when vaccines have been effective in preventing diseases for a number of decades, younger parents have not seen the consequences of these diseases and therefore may not feel compelled to vaccinate their children. The Centers for Disease Control and Prevention (CDC) is an important source for tracking infectious disease outbreaks and provides updates and recommendations for immunizations for adults, children, and travelers (www.cdc.gov).

Pathogen Characteristics

Pathogens possess certain characteristics that assist in their penetration and survival in the host despite the presence of an intact defense system. *Virulence* is the term used to describe the potential of microorganisms to cause disease in the host. Microorganisms with high virulence cause disease in human hosts even when their immune systems are intact. Low-virulence organisms have a reduced ability to cause disease and are less likely to cause illness unless other factors favor them, such as immune defects or loss of physical barriers. Bacterial virulence depends on the ability to adhere to host cells, to invade tissues, or to deliver toxins.

Adherence and Invasion

The ability to adhere to the contact surface is critical for the success of a microorganism. This ability is affected by the surface hydrophobicity, the net surface charge, the binding molecules on the bacteria (adhesions), and the interaction with the host cell membrane. Many bacteria, such as *E. coli,* have pili, which are hairlike structures that extend from the bacterial cell wall and help increase attachment of the bacteria to host cells (Fig. 8.3). Other bacteria, such as *Streptococcus pyogenes,* have hairlike appendages called *fimbriae* that extend from the cell surface and promote attachment. Most pathogenic bacteria remain outside of the cells, but some are able to enter host cells by binding to particular proteins on the cell surface, such as complement receptors (C3b). Intracellular bacteria may kill the host cell or may remain dormant and hidden from the immune system for a time.

Development of a microbial slime layer also facilitates adherence and improves survival of the pathogen. A thin layer of peptidoglycan is created that can participate in the development of pili or flagella on some organisms.

Bacterial Endotoxin

Some bacteria produce substances that are toxic to the host. Some of these bacteria secrete their toxins (see later), and others contain the toxin within the cell wall (endotoxin). Endotoxin is a lipopolysaccharide (LPS) in the membrane of gram-negative bacteria that is highly immunogenic and attracts immune cells to the area of infection and induces the production of inflammatory cytokines (Chapter 9).

FIG 8.3 Bacterial pili increase properties of adherence and may increase virulence. (From Rosenshine I, et al: A pathogenic bacterium triggers epithelial signals to form a functional bacterial receptor that mediates actin pseudopod formation. *The EMBO Journal,* 1996;15(11): 2613–2624.

Unfortunately, when endotoxin is released by immune destruction of the bacteria, the resulting immune activation can have profound effects on the host, causing septic shock, intravascular coagulation, and respiratory failure (see Chapter 20 for a discussion of shock).

Bacterial Exotoxins and Enzymes

Exotoxins are secreted by some types of bacteria and can cause havoc, even when the bacteria that made the toxin are no longer present. Neurotoxins produced by *Clostridium botulinum* and *Clostridium tetani* interfere with synaptic transmission at the neuromuscular junction, resulting in paralysis and respiratory failure. Like endotoxin (LPS), some secreted exotoxins are extremely good at activating the immune system to such a degree that the immune response can cause shock. For example, *Staphylococcus aureus* antigens can cause toxic shock syndrome.

Some bacteria produce tissue-degrading enzymes. These enzymes degrade certain body tissues, promoting continued disease. For example, *S. aureus* secretes coagulase, which coagulates plasma and contributes to the formation of fibrin walls around the lesions caused by these bacteria. This allows the bacteria to persist in the cellular environment. The coagulase also causes deposits of fibrin on the bacteria itself, leading to improved protection from phagocytosis.

Other bacteria produce hyaluronidase, which is an enzyme that breaks down hyaluronic acid. This acid is present in connective tissue. This action helps disseminate the bacteria through the tissues. Some bacteria are able to produce substances that are cytolysins. These cytolysins can dissolve red blood cells, tissue cells, or leukocytes. For example, group A streptococci produce streptolysin O, which lyses red blood cells.

Evasion of Immune Cells

Some bacteria develop an outside coating on their surface to prevent phagocytosis. The coating can be made of host cell components or a polysaccharide capsule. Some bacteria form interactive colonies that produce a polysaccharide matrix called *biofilm*. This slimy coating on solid surfaces can involve a single species or several species. These bacteria can be difficult to dislodge and may be hidden and protected from the host's immune mechanisms. An important and well-known example is the biofilm that forms on tooth enamel, commonly called *plaque*.

Endospore Formation

Several of the bacteria are capable of creating endospores. In response to the depletion of nutrients such as carbon, nitrogen, or phosphorus, the cell forms an internal spore. These spores are in a resting state that is markedly resistant to heat, chemical agents, and desiccation. When the environment is more favorable, spores are reactivated. Common bacteria with this ability include *Bacillus* and *Clostridium.*

Antimicrobial Resistance

At least eight mechanisms of bacterial resistance to drugs are known: **1)** Enzymatic inhibition: Bacteria can produce enzymes that inactivate drugs (e.g., β-lactamase can inactivate penicillins). **2)** Reduced permeability: Bacteria can reduce the permeability of their membranes to the antibiotic by down-regulation or loss of specific water-filled pores called *porins* that drugs use to enter the bacterial cell. **3)** Alteration of target site: Bacteria can create a modified target that is no longer as susceptible to the antibiotic. For example, a number of antibiotics bind to bacterial ribosomes to interrupt protein synthesis, but bacteria may change the binding site of the ribosome (by methylation, for instance) so that the drug no longer binds. **4)** Protection of target site: Bacteria may produce new proteins that protect or stabilize the target against the drug. **5)** Overproduction of target: Simply producing a large quantity of targets within the bacteria can make them resistant by binding up the drug to target, yet having a surplus of target that can maintain activity. **6)** Bacteria can employ an export mechanism that uses an ion gradient (H^+) to actively pump out the drug. **7)** The bacteria can develop a workaround to bypass the process that is affected by the drug. For example, the bacteria may develop a way to obtain needed substrates from the environment rather than producing them. **8)** Bacteria may develop a molecule (e.g., glycoprotein) to bind up the antibiotic before it has a chance to affect the bacterial cell processes. Usually these resistant properties develop by chance mutations in the organisms; however, some bacteria may be able to transmit their resistance to other bacteria through gene transfer.

Resistant infections occur because of natural, random genetic differences in the population of organisms that affect their susceptibility to antibiotics. The susceptible strains die, and the resistant strains emerge after exposure to antibiotics, creating a drug-resistant infection. Antibiotics do not cause mutations in bacteria and therefore don't *cause* resistance; however, excessive use of antibiotics and subtherapeutic dosing contribute to the emergence of resistant infections. Relatively resistant microorganisms survive a low-dose antibiotic course to become the dominant species and may then be transmitted to other individuals. Generally, when the antibiotic is no longer present, the resistant strains no longer have a growth advantage and will diminish.

The rapid development of antibiotic resistance in bacteria can occur when bacteria transmit their resistance genes to other bacteria. Bacteria are subjected to infections of their own in the form of plasmids and bacteriophages that can transfer genes between microorganisms. These mobile genetic elements can be transmitted through conjugation (bacterial sexual reproduction) or by infection of the bacteria (bacteriophages). The possibility of transfer of microbial resistance through mobile genetic elements is a greatly feared outcome that could produce virulent, multidrug resistant strains quite rapidly. More resistant bacteria in the environment because of antibiotic use increase the chance that mobile genetic elements could be transferred among them. Some common pathogens and their mechanisms of antimicrobial resistance are shown in Table 8.3. Susceptibility of a bacterial infection to particular antibiotics can be evaluated by culture and sensitivity testing. An example of common sensitivity testing results is shown in Box 8.1.

TABLE 8.3 Resistance Mechanisms Found in Common Bacterial Pathogens

Pathogen	Resistance Phenotype	Major Resistance Mechanism
Streptococcus pneumoniae	β-Lactams	Alteration of target enzymes (PBPs)
	Macrolides, lincosamides, streptogramin B	Alteration of ribosomal target sites (methylation of adenine residue in domain V of 23S rRNA—ermB); efflux (mefE)
	Tetracycline	Protection of ribosomal target site (tetM)
	Trimethoprim and sulfonamides	Alteration of target enzymes (dihydrofolate reductase–trimethoprim; dihydropteroate synthase—sul1, sul2 in sulfonamines)
	Fluoroquinolones	Alteration of target enzymes (DNA gyrase—gyrA mutations; topoisomerase IV—parC mutations)
Staphylococcus aureus	β-Lactams	Enzymatic inhibition (penicillinase production)
	Penicillin	Alteration of target enzyme—PBP2a (mecA)
	Methicillin, oxacillin, nafcillin, and cephalosporins (MRSA)	
	Glycopeptide GISA	Alteration of cell wall precursor targets (thickened cell wall binds drug, preventing it from reaching its target)
	GRSA	Alteration of cell wall precursor targets (plasmid-mediated transfer of vanA genes from VRE, resulting in d-ala-d-lac peptidoglycan precursors)
Enterococci	β-Lactams (ampicillin)	Alteration of target enzymes (PBP5 in *Enterococcus faecium*); enzymatic inhibition–rare (penicillinase in *E. fecalis*)
	Aminoglycosides	Altered ribosomal target site mutations, enzymatic inhibition (high-level resistance: aminoglycoside-modifying enzymes)
	Vancomycin	Alteration of cell wall precursor targets (high-level resistance: VanA, B, D phenotypes; low-level resistance: VanC, E, G phenotypes)
	Linezolid	Alteration of ribosomal target sites (G2576U mutation in domain V of 23S rRNA)
	Quinupristin-dalfopristin	Enzymatic inhibition; efflux; target modification *(E. faecium)*
Neisseria gonorrheae	Penicillins	PPNG: enzymatic inhibition (plasmid-acquired penicillinase); CRNG: altered target enzymes (PBPs)
	Fluoroquinolones	Alteration of target enzymes (DNA gyrase; topoisomerase IV); efflux (MtrR-CDE efflux system)
	Tetracycline	Protection of ribosomal target (tetM gene)
	Macrolides	Efflux; alteration in ribosomal targets (C2611T mutation in domain V of the 23S rRNA)
	MDR	Efflux (MtrR-CDE system: penicillin, tetracycline, macrolides)
Pseudomonas aeruginosa	β-Lactams	Enzymatic inhibition (AmpC cephalosporinases, extended-spectrum β-lactamases, metallo-β-lactamases); active efflux (MexAB); reduced outer membrane permeability (loss of OprD channel)
	Aminoglycosides	Enzymatic inhibition (aminoglycoside-modifying enzymes); efflux (MexXY); alteration of ribosomal targets (ribosomal methylation)
	Fluoroquinolones	Efflux (MexAB, CD, EF, XY, GH, VW); alteration of target enzymes (DNA gyrase mutations—gyrA)
	MDR	Overexpression of the MexA-MexB-OprM active efflux system (resistance to quinolones, tetracyclines, and trimethoprim)
Acinetobacter baumannii	β-Lactams	Enzymatic inhibition (AmpC cephalosporinases, plasmid-acquired β-lactamases of the TEM, SHV, CTX-M, PER, VEB families, metallo-β-lactamases of the IMP, VIM, SIM families, and OXA-type serine carbapenemases); alteration of target enzymes (PBPs); reduced outer membrane permeability; efflux pumps
	Aminoglycosides	Enzymatic inhibition (aminoglycoside-modifying enzymes); efflux pumps
	Quinolones	Efflux pumps
	Tigecycline	Efflux pumps
Stenotrophomonas maltophilia	β-Lactams	Impermeable outer membrane
		Enzymatic inhibition (inducible metallo-β-lactamases L1, L2)
	TMP-SMX	Alteration in sulfonamide target enzymes (sul1, sul2 genes—associated with plasmids or class 1 integrons)
	Fluoroquinolones	Alteration of target enzymes (DNA gyrase mutations); efflux pumps
	MDR	MDR efflux pump (smeDEF confers resistance to tetracycline, erythromycin, chloramphenicol, norfloxacin, ofloxacin)

TABLE 8.3 Resistance Mechanisms Found in Common Bacterial Pathogens—cont'd

Pathogen	Resistance Phenotype	Major Resistance Mechanism
Klebsiella pneumoniae	β-Lactams	Enzymatic inhibition (constitutive expression of penicillinases; extended-spectrum β-lactamases; KPC, NDM-1 carbapenemases); decreased outer membrane permeability
	Fluoroquinolones	Alteration of target enzymes (DNA gyrase mutations—gyrA); efflux; protection of target site (plasmid-mediated qnr genes)
	Aminoglycosides	Enzymatic inhibition (aminoglycoside-modifying enzymes); alteration of ribosomal targets (ribosomal methylation)
Bacteroides spp.	β-Lactams	Enzymatic inhibition (chromosomally encoded CepA cephalosporinases; metallo-β-lactamases); efflux (homologues of RND-pumps); alteration in drug targets (PBPs)
	Macrolides, lincosamides, streptogramin B	Alteration of ribosomal targets
	Tetracycline	Protection of ribosomal target (tetQ); efflux
	Quinolones	Alteration of target enzymes (DNA gyrase mutations—gyr A); efflux

CRNG, Chromosomally resistant *N. gonorrheae*; *CTX-M,* cefotaxime-M; *GISA,* glycopeptide intermediate *S. aureus*; *GRSA,* glycopeptide-resistant *S. aureus*; *IMP,* imipenem; KPC, *K. pneumoniae* carbapenemase; *MDR,* multidrug resistance; *MRSA,* methicillin-resistant *S. aureus*; *MtrR,* multiple transferable resistance; *NDM-1*, New Delhi metallo-β-lactamase–1; *PBPs,* penicillin-binding proteins; *PER, Pseudomonas* extended resistance; *PPNG,* penicillinase-producing *N. gonorrheae*; *RND,* resistance-nodulation-cell division; *rRNA,* ribosomal RNA; *SHV,* sulfhydryl variable; *SIM,* Seoul imipenemase; *TEM, Temoneira*; *TMP-SMX,* trimethoprim-sulfamethoxazole; *VEB,* Vietnam extended-spectrum β-lactamase; *VIM,* Verona integron-encoded metallo-β-lactamase; *VRE,* vancomycin-resistant enterococci.
From Bennett J, Dolin R, Blaser, M: *Mandell, Douglas, and Bennett's principles and practice of infectious diseases,* ed 8, Philadelphia, 2015, Saunders, p 247.

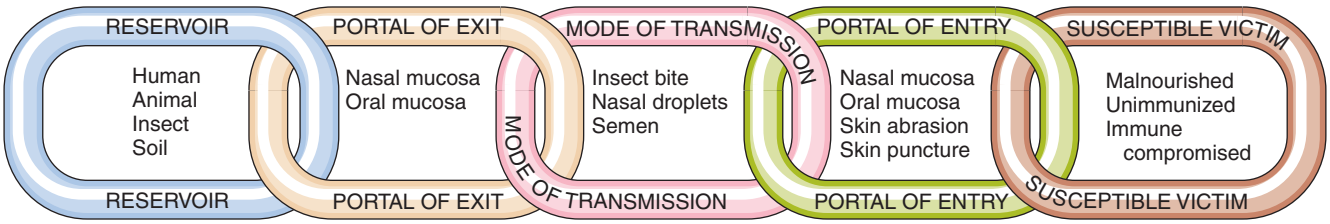


FIG 8.4 Chain of transmission of microorganisms from host to victim.

KEY POINTS

- A number of microorganisms are considered resident flora because they live on or in the host without causing disease. Resident flora and probiotics benefit the host by synthesizing molecules and inhibiting the growth of nonresident microorganisms. If the host's immune system is compromised, resident flora may become pathogenic and cause opportunistic infection.
- Microorganisms possess characteristics that enhance their pathogenic potential. Adherence is improved by the presence of adhesion molecules, slime layers, and pili. Escape from immune detection and destruction is enhanced by encapsulation, spore formation, mutation, use of flagella, and toxin production. Microorganisms that possess these advantageous characteristics are more virulent and thus more likely to cause disease.
- Drug resistance occurs when microorganisms undergo chance mutations that allow them to survive in the presence of an antibiotic. When the antibiotic is present, these resistant strains emerge to become the dominant species in an individual and may be transmitted to others, causing resistant infections.
- Immunizations alter the susceptibility of the host by stimulating the immune system to create antibodies to the pathogen.

BOX 8.1 Example of the Results of Bacterial Culture and Sensitivity Testing

Culture Urine Colony Count

Specimen Source	Urine
Result	> 100,000 Organisms/mL *Escherichia coli*
Report Status	Final

Susceptibility Testing

Cefazolin	Susceptible
Gentamicin	Susceptible
Levofloxacin	Susceptible
Minocycline	Susceptible
Nitrofurantoin	Susceptible
Oxacillin	Susceptible
Trimethoprim/Sulfa	Resistant
Vancomycin	Susceptible

TRANSMISSION OF INFECTION

The transmission of microorganisms requires an unbroken chain of events to enable one host to transmit the infection to another (Fig. 8.4). Pathogens must live and reproduce in a reservoir. The reservoir may be a human, as in the influenza virus; an animal, as in rabies; an insect, as in West Nile virus; or soil, as in enterobiasis (pinworm infestation). The pathogen must have a portal of exit and a mode of transmission from the reservoir to a susceptible host, where it finds a portal of entry. Some microorganisms can survive outside the reservoir for a time, whereas others must be transmitted from host to host by direct contact.

Routes of Transmission

There are three common routes of direct transmission of pathogens. The most common is through the exchange of body fluids (droplets) from kissing or sexual intercourse. In this type of transmission, droplets are spread from an infected host to the conjunctiva or mucous membranes of a second host. The second type of direct transmission occurs from the bite of an animal (e.g., rabies) or from contact with contaminated soil (e.g., systemic mycosis). The third type of direct transmission occurs through the placenta when diseases such as HIV are transmitted directly to the fetus. This is called *vertical transmission.*

Indirect transmission occurs as vehicle-borne, vector-borne, or airborne transfers. A vehicle-borne transfer occurs when an infectious agent is transported to the host. The vehicle could be food, water, clothing, plasma, or tissues. The agent does not require any development or multiplication and is delivered without change with the vehicle. Airborne transmission occurs with aerosols, including suspensions of particles (smaller than the size of droplets), which can be carried great distances from the source for transmission. Examples are outbreaks of measles and legionnaires disease. Vector-borne transmission can occur as a mechanical or biological transfer. The mechanical transfer occurs when an insect carries an infectious agent on its feet or proboscis. In this transfer the agent does not require multiplication or development before transfer. This is in contrast to biological vector-borne transport. In this type of transmission, the agent propagates and there is cyclic development before the arthropod can transmit the disease.

Control of the spread of infectious disease depends on breaking the chain of transmission in one or more places (Fig. 8.5). Therefore it is critical to understand the life cycle and transmission patterns (epidemiology) of different pathogens—a top priority of the CDC. An infectious disease that has a fairly constant presence in a community and changes little from year to year, such as the sexually transmitted disease *Chlamydia,* is classified as endemic. A significant increase in new infections in a certain population, such as a measles outbreak at a university, is termed an epidemic. An epidemic that has spread to a large geographic area is a pandemic. The bubonic plague, which caused 100 million deaths worldwide in the 1300s, was a pandemic.

Destroying nonhuman reservoirs and vectors of the pathogen can break the chain of transmission. For example, controlling the number of mosquitoes with insecticides and other biological means is a method used to curb the spread of malaria and West Nile virus. Immunization of domesticated animals against rabies eliminates one reservoir of potential rabies transmission. Distribution of clean needles in the intravenous drug user community is aimed at removing a common transmission vector (contaminated needles) for HIV, the causative pathogen responsible for acquired immunodeficiency syndrome (AIDS).

Blocking the portal of exit can also block transmission of the pathogen. Having patients with TB wear face masks while they move through the hospital and implementing respiratory isolation techniques to stop transmission are interventions aimed at blocking the portal of exit. Standard precautions are infection-control guidelines designed to block the pathogen's portal of exit, route of transmission, and portal of entry. Handwashing is one of the most effective ways to break the chain of transmission by blocking an important mode of transmission—contaminated hands.

Multiple environmental factors affect the prevalence and transmission of various infections and infestations. For example, parasitic infections are facilitated by hot and humid climates, overcrowded living conditions, the presence of insect vectors in bed linen or clothing, improper sewage disposal or treatment (such as the use of raw human sewage as fertilizer), the lack of clean water, and the consumption of contaminated raw or undercooked meat or vegetables.

Infections may be transmitted by inhalation of polluted dust or air. For example, the fungus *Coccidioides immitis,* which causes valley fever, is pandemic in the southwestern United States. Toxoplasmosis is caused by inhalation or ingestion of dirt, sand, or litter dust contaminated with cat feces that contain the causative protozoon *Toxoplasma gondii.* The plague is still a disease of concern in many countries. It is caused by the bacterium *Yersinia pestis,* which is carried by a rodent flea. There are multiple examples of foodborne illness such as *Salmonella,* with outbreaks related to improper processing, storage, or cooking of meats and raw produce.

Emerging Infectious Diseases

Pathogens can spread faster and in greater numbers due to global transport networks and world travel. Global growth, migration, and tourism are largely responsible for the threat of locally emergent diseases to escalate toward large-scale pandemics. For example, coronaviruses (CoVs) are general respiratory tract infections that infect a wide range of mammals and are symptomatic of the common cold. Recently, however, certain CoV infections, such as severe acute respiratory syndrome (SARS-CoV) in 2002 and Middle East respiratory syndrome (MERS-CoV)

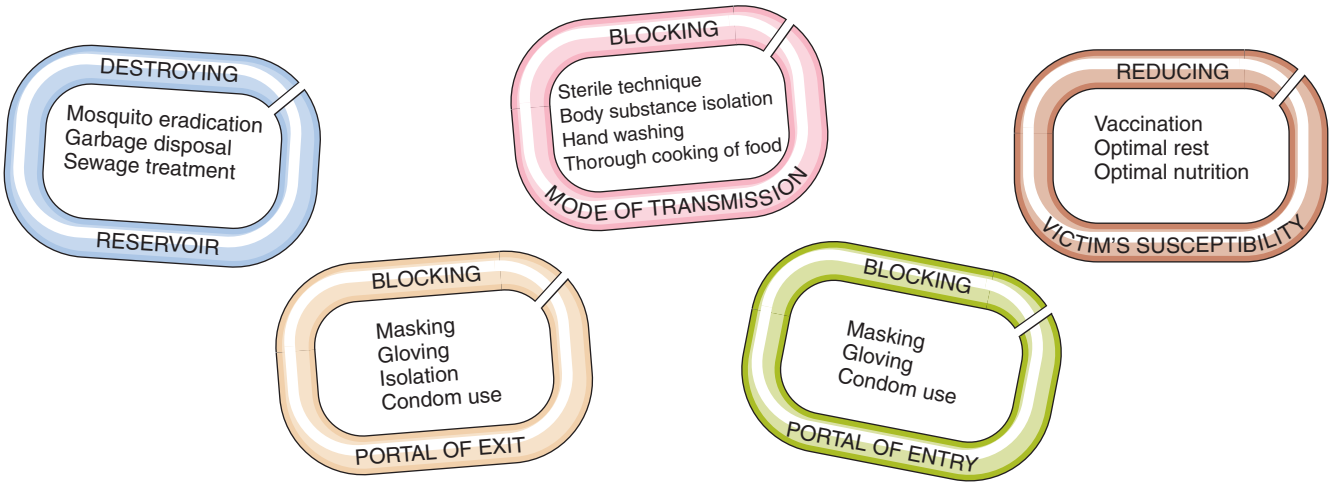


FIG 8.5 Breaking the chain of transmission of microorganisms from host to victim.

in 2012, emerged as serious diseases in humans. The SARS and MERS viruses emerged from zoonotic animal-to-human transmission and caused serious and sometimes fatal lower respiratory infections. The WHO reported that SARS infected 8096 people, resulted in 774 deaths, and was spread globally by human-to-human contact from 2002 until its eradication in 2004. MERS virus has claimed 291 lives of 837 confirmed cases since 2012 and has remained at large with unknown pathology in humans (WHO last reported July 23, 2014).

The single largest Ebola outbreak to date was documented in March 2014 in Guinea, West Africa. By August 8, 2014, the outbreak had spread to neighboring regions and was declared an epidemic and an international public emergency by the WHO. The WHO reported on September 13, 2015, that Ebola virus had claimed 11,306 lives of the 28,256 reported cases, 17 of which were treated outside Africa and 4 of whom died.

The Zika virus is a flavivirus spread by *Aedes* mosquitos and has no known treatment. Zika virus was mostly ignored until recently because of usually mild symptoms; however, in 2016 the Zika virus was correlated with significant developmental defects such as the severe brain malformation microcephaly. Microcephaly is a brain disorder that results in a smaller-than-normal head size and mental disability. It was also discovered that in addition to being carried by mosquitos, Zika can spread via sexual intercourse. In 2016 the CDC issued a comprehensive travel guide that warded off travel and set up warnings for the affected areas, especially for pregnant women, until more could be understood about the virus.

Modern medicine has eradicated many potentially fatal diseases with vaccination programs. Conversely, recent reports indicate that preventable disease outbreaks such as the measles, mumps, and whooping cough spiked in the United States and abroad, in part because of antivaccination movements. The CDC's National Center for Immunization and Respiratory Diseases recently reported the largest measles outbreak since its near eradication in 2000. A record 668 measles cases occurred throughout 27 states in 2014. The antivaccination movement was based on the speculative correlation of an increased diagnosis of autism with increased childhood vaccinations. This idea was promoted by fraudulent research that has since been retracted. However, significant segments of the population continue to avoid vaccinations for their children.

Weapons of Bioterrorism

Crude forms of weaponized infectious diseases have been around since 600 BC when plagued cadavers, filth, or pollution were used to weaken an enemy. Unfortunately, the use of biological warfare became more sophisticated throughout the nineteenth and twentieth centuries. Wartime efforts spurred the development and testing of weaponized pathogens. For example, "Unit 731" was a World War II Japanese biowarfare program that was believed to have executed more than 10,000 prisoners with weaponized *Bacillus anthracis, Neisseria meningitidis, Vibrio cholerae,* etc. Modern domestic bioterrorism in the United States has been largely isolated, extremist, or cult sponsored and resulted in a minimal death toll, but these events serve as a wake-up call for bioterrorism fortification, reducing risk of exposure, and improving public health surveillance.

KEY POINTS

- Transmission of disease requires a chain of events that includes passing of the pathogen from the reservoir of the infection through a portal of exit to a susceptible host through a portal of entry by a circumscribed mode of transmission.
- Environmental factors influence the likelihood of exposure and infection by microorganisms. Sanitation, air quality, living conditions, and climate are important factors.
- The spread of emergent infectious disease is further facilitated by a global society. For example, severe acute respiratory syndrome (SARS), Middle East respiratory syndrome (MERS), and Ebola outbreaks were pinpointed in isolated regions, yet human-to-human contact carried these pathogens elsewhere and resulted in worldwide epidemics.
- Dangerous pathogens designed for bioterrorism present proximate dangers to public health.

TYPES OF PATHOGENIC ORGANISMS

A large number of different types of microorganisms pose an infection risk for humans (Fig. 8.6). The chapter thus far has focused mainly on bacterial pathogens. Many bacterial infections are treatable with antibiotics, and there is much focus on them for that reason. Viral infections are difficult to treat effectively, and the primary focus has been on finding vaccines for them. Fungal and parasitic infections are also important contributors to human disease, particularly in immunocompromised hosts and in areas with contaminated water supply. Prion infection is from a protein and not an organism at all. Each of these infectious agents is discussed briefly, with a focus on common causes of infection.

Bacteria

Bacteria are single-celled, rigid-wall organisms that have no internal organelles (Fig. 8.7). Some live in the intestines of humans or other animals and participate in digestion. Others live in the soil and are responsible for its fertility. They degrade dead tissue into useful components for other organisms to use. Among the countless types of bacteria that exist, only a small percentage is known to be harmful to humans. Some bacteria that are associated with infections of various sites are shown in Box 8.2.

Bacteria have a variety of shapes, including cocci (spherical), bacilli (rod-shaped or comma-shaped rods, e.g., *Pseudomonas aeruginosa* vibrio), or spiral (twisted rod shaped, e.g., *Spirillum*). Those classified as cocci are also seen in clusters, chains, pairs, or tetrads. The majority of bacteria that cause disease in humans are one of these true bacteria. Filamentous bacteria may have branching structures that resemble fungi. *Mycobacterium tuberculosis* would be an example.

Spirochetes (the majority of which are anaerobic) possess a motile spiral filament (e.g., *Treponema pallidum*). Bacteria of the genus *Mycoplasma* do not have a rigid cell wall and are pleomorphic (many formed) in shape. They are some of the smallest of the bacteria; an example is *Mycoplasma pneumoniae,* which causes an atypical pneumonia. *Rickettsia* is a genus consisting of intracellular parasites that can have a variety of shapes. Vectors usually spread this group; for example, Rocky Mountain spotted fever is caused by *Rickettsia rickettsii* and transmitted by ticks. *Chlamydia* are also intracellular parasites, but have a more complex life cycle, with *Chlamydia trachomatis* as an example.

Bacteria are classified not only by morphology (shape) but also by the response to gram staining. Gram staining separates bacteria into gram-positive organisms, which appear dark purple under the microscope; gram-negative organisms, which appear pink; or acid-fast organisms, which resist staining but, once stained, resist discoloration. Further differentiation of bacteria is based on nutritional requirements (such as whether the organism is anaerobic or aerobic), on colony characteristics, and on resistance.

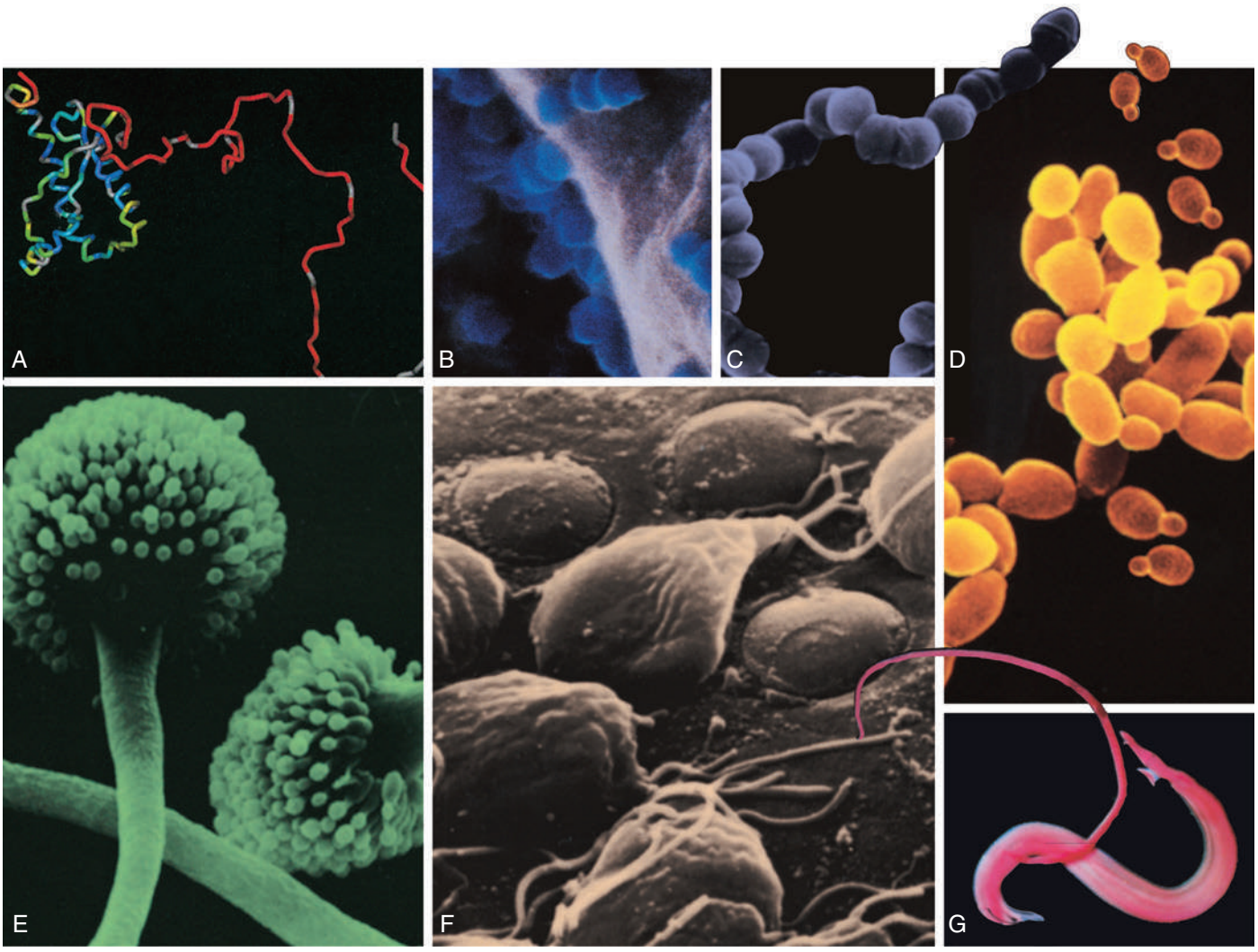


FIG 8.6 Examples of pathogenic organisms. **A,** Prion (infectious protein). **B,** Viruses (the human immunodeficiency virus [HIV] that causes AIDS). **C,** Bacteria (Streptococcus bacteria that cause strep throat and other infections). **D,** Fungi (yeast cells that commonly infect the urinary and reproductive tracts). **E,** Fungi (the mold that causes aspergillosis). **F,** Protozoa (the flagellated cells that cause traveler's diarrhea). **G,** Pathogenic animals (the parasitic worms that cause snail fever). (A, from Donne DG, et al: Structure of the recombinant full-length hamster prion protein PRp (29-231): the N terminus is highly flexible. *Proc Natl Acad Sci USA.* 1997;94:13452–13457. Copyright National Academy of Sciences, USA; B, from Lennart Nilsson C-G, from Patton KT, Thibodeau GA: *Anatomy & physiology,* ed 8, St Louis, MO, 2013, Mosby, p 26.)

Fig. 8.8 depicts examples of pathogenic bacteria and the areas that they commonly infect. Colonies are formed when bacteria penetrate initial defense mechanisms and multiply. In an attempt to contain and eliminate the invading bacteria, an acute inflammatory reaction occurs. Phagocytic cells such as neutrophils and macrophages are recruited to the area, where they ingest and destroy the microorganisms. If these responses are insufficient to contain the infection, the bacteria move through the body in natural currents of fluids (i.e., bloodstream, lymph system, or interstitial fluids). Bacteria may move through the lymph system to the lymph nodes, where they stimulate an immune response. If they are present in sufficient numbers to overwhelm the lymph nodes, circulating clumps of bacteria (emboli) can cause bacteremia and microabscesses. In severe cases, sepsis, hypotension, organ system failure, and death can occur (see Chapter 20).

Viruses

Viruses, the smallest known infective agents, range in size from 20 to 300 nm. They consist of a protein shell called the *capsid* and a core of genetic material made of either ribonucleic acid (RNA) or deoxyribonucleic acid (DNA). The capsid can be in many shapes, including helical, icosahedral, or large pleomorphic shapes. Some viruses also have a protective envelope surrounding the capsid that is obtained from the cell membrane of an infected host cell.

Viruses are classified as RNA or DNA viruses and as either single-stranded (ss) or double-stranded (ds). Some RNA viruses, called *retroviruses,* contain the enzyme reverse transcriptase and can convert their RNA into DNA that can be incorporated into the host's DNA. The HIV virus is an example of a retrovirus (see Chapter 12). A

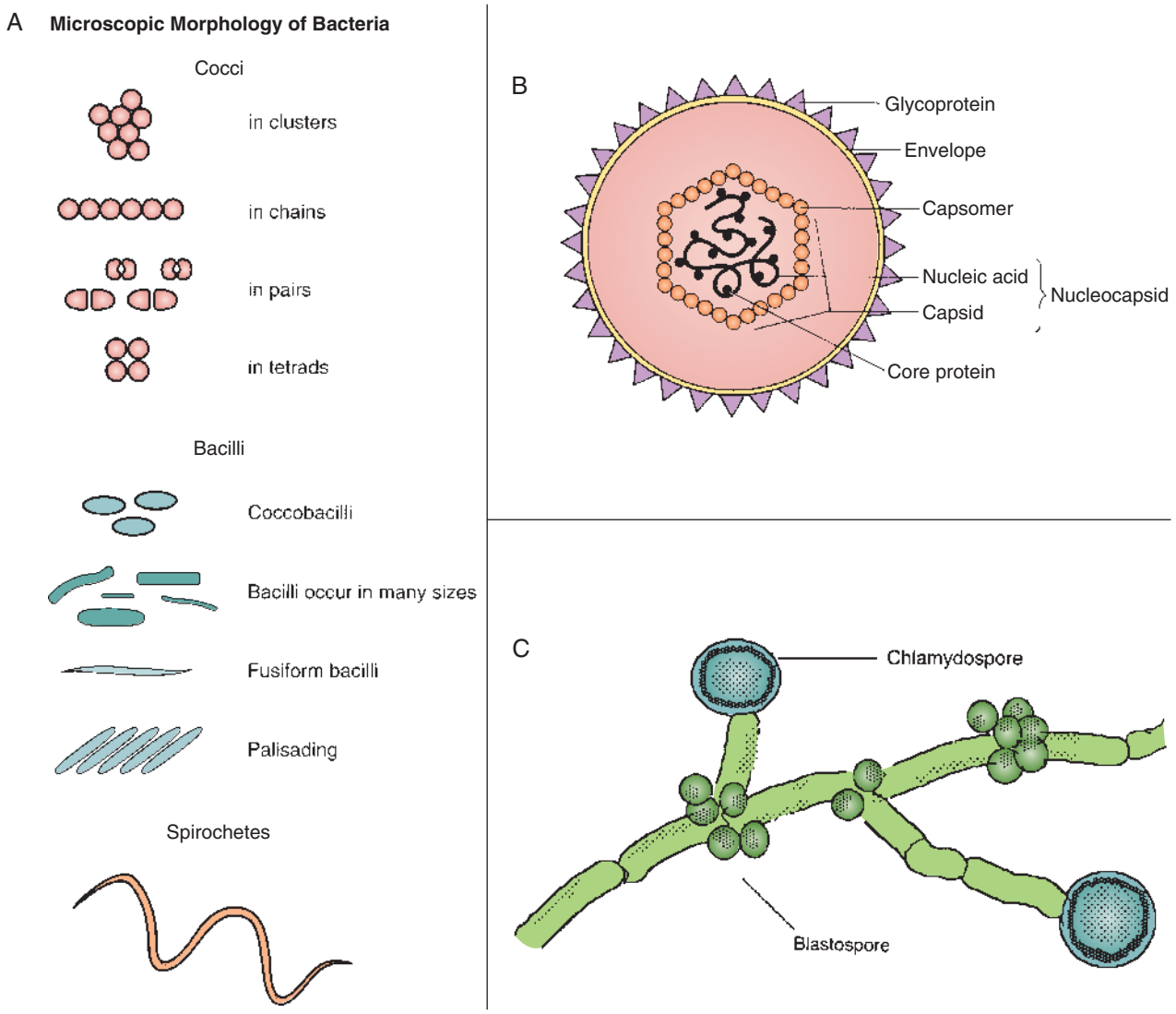


FIG 8.7 Types of microorganisms. **A,** Bacteria. **B,** Virus. **C,** Fungus. (A, from Mahon CR, et al: *Textbook of diagnostic microbiology,* ed 4, Philadelphia, 2011, Saunders; B and C, from Nisengard RJ, Newman MG: *Oral microbiology and immunology,* ed 2, Philadelphia, 1994, Saunders.)

comparison of viruses, transmission characteristics, and resulting disease processes is presented in Table 8.4.

DNA viruses (e.g., herpes simplex virus) enter the nucleus of the host cell and produce messenger RNA (mRNA) by commandeering the host cell's enzymes. Viral proteins are then formed from the mRNA, and the DNA of the virus is replicated by host polymerases. DNA and proteins are assembled into new viruses in the host cell. The RNA viruses replicate within the cytoplasm, and most produce mRNA that is then translated into proteins and genomic RNA, from which new viruses are created.

Transmission of a virus occurs from one infected person to another or from an animal reservoir (zoonotic infection). The steps of the viral life cycle include attachment to the target cell as the initial step. The virus then penetrates the cell membrane in various ways. Once inside the host cell, the virus uses the host cell's materials to produce new viral components. The virus may be released from the host cells by budding from the cell's surface. Viruses that do not manufacture an envelope are usually released by *lysing* the host cell, thus destroying it.

Fungi

Fungi are eukaryotic microorganisms with the ability to form complex structures with thick, rigid cell walls. They can grow as a mold with branched filaments or as a meshwork-type structure. Yeasts are a type of fungi with ovoid or spherical shapes. In contrast to bacteria, which have no organelles, the cytosol of fungi does contain organelles. Infections caused by fungi are called *mycotic infections,* or mycoses. Fungi cause infection first by colonizing the area. The fungus adheres to and proliferates on the site of infection. The next phase requires invasion of the epithelium. Anything that breaks the integrity of the skin (e.g., maceration) facilitates invasion. Polymorphonuclear leukocytes attempt to phagocytize and digest the invading fungi. Neutrophils, monocytes, and eosinophils can destroy fungi in the body.

Certain fungi live in the body as normal flora (e.g., *Candida*). When the body's defense mechanisms are compromised, they can overgrow and cause local or systemic infections. Patients who have been given antibiotics generally lose some of their resident flora along with the targeted pathogen. Fungi not affected by these antibiotics overgrow to fill the void niche. Some patients suffering from AIDS, leukemia,

BOX 8.2 Examples of Bacteria Associated With Specific Infections

Skin and Wound Infections
Staphylococcus aureus
Streptococcus pyogenes (group A)
Gram-negative bacilli
Treponema pallidum
Pseudomonas aeruginosa
Anaerobic streptococci
Clostridium spp.
Enterococcus
Bacteroides spp.

Brain and Meninges
Neisseria meningitidis
Haemophilus influenzae
Streptococcus pneumoniae
Streptococcus spp.
Escherichia coli
Gram-negative bacilli
Streptococcus pyogenes (group A)
Staphylococcus aureus
Mycobacterium tuberculosis
Listeria monocytogenes
Enterococcus (neonatal period)
Treponema pallidum
Leptospira
Streptococci (aerobic and anaerobic)
Bacteroides spp.

Ear, Nose and Throat
Streptococcus pneumoniae
Streptococcus pyogenes (group A)
Haemophilus influenzae
Gram-negative enteric bacilli
Pseudomonas aeruginosa
Anaerobic streptococci
Staphylococcus aureus
Neisseria gonorrheae
Bacteroides spp.
Fusobacterium
Spirochetes
Corynebacterium diphtheriae
Bordetella pertussis

Lungs and Pleura
Mycoplasma pneumoniae
Streptococcus pneumoniae
Haemophilus influenzae
Staphylococcus aureus
Klebsiella
Pseudomonas aeruginosa
Gram-negative bacilli
Streptococcus pyogenes (group A)
Mycobacterium tuberculosis
Chlamydia psittaci
Legionella pneumophila
Anaerobic streptococci
Bacteroides spp.
Coxiella burnetii
Fusobacterium
Enterococcus

Endocardium
Viridans group of streptococci
Staphylococcus aureus
Enterococcus
Other streptococci
Staphylococcus epidermidis
Gram-negative enteric bacilli
Pseudomonas aeruginosa

Peritoneum and Biliary Tract
Escherichia coli
Gram-negative bacilli
Enterococcus
Bacteroides fragilis
Anaerobic streptococci
Clostridium spp.
Streptococcus pneumoniae
Streptococcus pyogenes (group A)
Neisseria gonorrheae
Mycobacterium tuberculosis
Staphylococcus aureus

Kidney
Escherichia coli
Gram-negative bacilli
Staphylococcus aureus
Staphylococcus epidermidis
Mycobacterium tuberculosis

Genitourinary Tract
Neisseria gonorrhoeae
Chlamydia trachomatis
Trichomonas vaginalis
Ureaplasma urealyticum
Gram-negative enteric bacilli
Staphylococcus aureus
Mycobacterium tuberculosis

Bone (Osteomyelitis) and Joint
Staphylococcus aureus
Salmonella
Gram-negative enteric bacilli
Streptococcus pyogenes (group A)
Mycobacterium tuberculosis
Anaerobic streptococci
Pseudomonas aeruginosa

Joints
Neisseria gonorrheae
Streptococcus pneumoniae
Neisseria meningitidis
Haemophilus influenzae (in children)

alcoholism, drug abuse, and malnutrition or those being treated with immunosuppressive agents lack a well-functioning immune system that prevents fungi from overgrowing and are therefore more susceptible to fungal infections (opportunistic fungi). When the environment contains more nutrients on which the fungi can grow, such as the hyperglycemic bloodstream of a diabetic patient or the vaginal tract of a female taking antibiotics, fungi can overgrow and cause infections. *Pneumocystis carinii* was reclassified as a fungus in 2006 and renamed *Pneumocystis jiroveci*. This opportunistic infection is commonly associated with HIV disease.

Superficial mycoses, such as those caused by dermatophytes (e.g., tinea pedis), occur only on superficial, dead, keratinized tissue like hair, epidermis, and nails. Cutaneous fungi do not invade the tissues, but do result in an inflammatory response, as in *tinea pedis (athlete's foot)*. Subcutaneous mycoses occur when fungi are introduced into subcutaneous tissues and can be seen in ulcers or abscesses on the skin. Systemic infections are invasive to lungs and other organs (e.g., *Cryptococcus neoformans*). Systemic treatment is commonly used for immunocompromised patients or for patients with disseminated disease because skin infection is usually self-limited. Topical antifungal drugs may be used to treat these superficial infections. Table 8.5 summarizes some examples of fungal infections.

Systemic mycoses may occur in both healthy and immunocompromised hosts. Because the fungi causing systemic infections are usually found in soil, these infections tend to be endemic to certain regions where the fungus is found. Infection is caused by inhalation of dust containing the fungus. Because of the endemic nature of these fungi, large segments of the population in the area may have been exposed and infected without any symptoms. If symptoms develop, they are usually self-limiting and mild. However, for those with compromised immune systems, the disease may become severe. Examples of systemic mycoses are histoplasmosis, blastomycosis, and coccidiomycosis. *Histoplasma capsulatum* is a fungus that commonly occurs in soil in the central and eastern United States. *Histoplasma* also occurs in soil rich with chicken feces or bat guano. Humans and animals exposed to dust storms in endemic areas or contaminated with these feces are most likely to be infected. They may also have positive histoplasmin skin tests and may show calcified sites of infection in their lungs.

Parasites

Parasites establish themselves and benefit from another organism. They range in size from a small unicellular protozoan to large worms. Parasites are representative of four families of the animal kingdom: protozoa, or

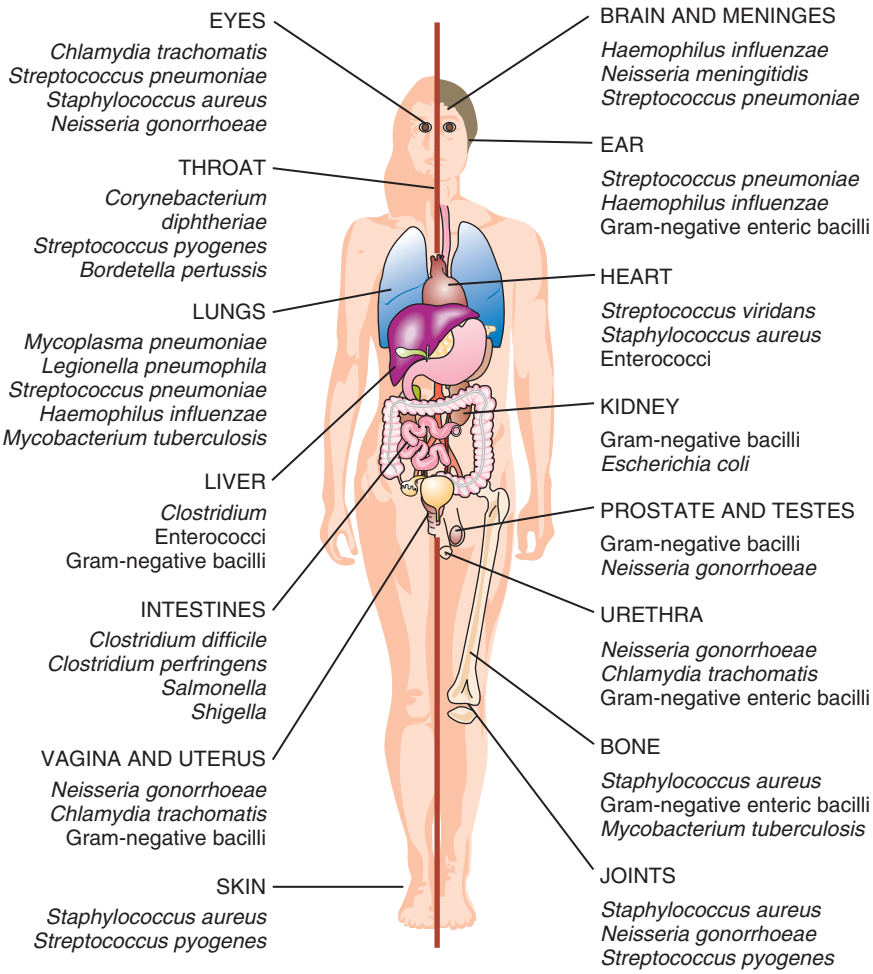


FIG 8.8 Examples of pathogenic bacteria classified according to the part of the human body that they commonly infect.

single-celled animals (Fig. 8.9); nemathelminths, or roundworms; platyhelminths, or flatworms; and arthropoda, or invertebrate animals with jointed appendages. These parasites live on or in the human body during some part of their life cycle. Parasites and protozoa are rarely transmitted by human contact; usually they are disseminated through a vector where the parasite or protozoan spends part of its life cycle. For example, malaria (*Plasmodium* sp.) is transmitted via mosquitoes. Many of the protozoal infections are transmitted through contaminated water or food and require ingestion (e.g., *Giardia lamblia*). Some parasites have specific surface glycoproteins that influence their ability to enter macrophages.

Host resistance depends on macrophages, neutrophils, eosinophils, and platelets that kill both protozoa and worm parasites. T cells are required to develop immunity against these organisms. The symptoms of parasitic infection depend on the area in which the infestation develops. Protozoan infestation (amebiasis) of the gastrointestinal tract may trigger cramping, abdominal pain, and bloody diarrhea. Infestation of the blood produces fever, chills, rigor, and later anemia, all of which are associated with malaria (*Plasmodium* infection). Acute pruritus and rash occur after infection of the skin with *Sarcoptes scabiei* (scabies).

Identification of the infectious agent is usually accomplished by visualization of the adult parasite; by direct observation of the area (inspection of the skin or hair); or by microscopic examination of blood, feces, or tissue samples. Table 8.6 summarizes various parasitic infections of humans, including the common name, location, symptoms, and mode of transmission.

KEY POINTS

- Microorganisms responsible for infections in humans include bacteria, viruses, fungi, and parasites.
- Bacteria are characterized according to shape (cocci, rods, spirals), reaction to stains (gram negative, gram positive, acid fast), and oxygen requirements (aerobic, anaerobic).
- Viruses are small pieces of genetic material (DNA, RNA) associated with proteins and lipids. Viruses are small intracellular pathogens that use the host's energy sources and enzymes to replicate. Viral replication may or may not destroy the host cell. DNA viruses may be incorporated directly into the host genome. RNA viruses serve as templates for the production of viral RNA and proteins.
- Retroviruses are RNA viruses that contain a special enzyme called *reverse transcriptase* that mediates the synthesis of a DNA copy of the RNA virus. The DNA can then be incorporated into the host genome and passed on to daughter cells when the cell divides.
- Fungal infections can be superficial (e.g., ringworm, athlete's foot), subcutaneous (e.g., sporotrichosis), or systemic (e.g., histoplasmosis). Systemic fungal infections tend to be more serious and usually do not occur unless the host's immune system is compromised.
- Parasites include protozoa, helminths (roundworms, flatworms), and arthropods. Manifestations of parasitic infections vary depending on the organism and site of infection. Common sites of parasitic infestation are the skin and gastrointestinal tract.

TABLE 8.4 Human Diseases Caused by Specific Viruses

Baltimore Classification	Family	Virus	Disease
dsDNA	Adenoviruses	Adenovirus	Acute febrile pharyngitis
	Herpes viruses	Herpes simplex type 1 (HSV-1)	Lesions in mouth, pharynx, conjunctivitis
		Herpes simplex type 2 (HSV-2)	Sores on labia, meningitis in children
		Herpes simplex type 8 (HSV-8)	Kaposi sarcoma
		Epstein–Barr virus (EBV)	Mononucleosis, Burkitt lymphoma
		Cytomegalovirus (CMV)	Mononucleosis, congenital infection
		Varicella-zoster virus (VZV)	Chickenpox, shingles
ssDNA	Papovaviruses	Papillomavirus	Warts, cervical carcinoma
dsRNA	Reoviruses	Rotavirus	Severe diarrhea
ssRNA+	Picornaviruses	Coxsackievirus	Nonspecific febrile illness, conjunctivitis, meningitis
		Hepatitis A virus	Acute hepatitis
		Poliovirus	Poliomyelitis
		Rhinovirus	Common cold
	Flaviviruses	Hepatitis C virus	Acute or chronic hepatitis, hepatocellular carcinoma
		Yellow fever virus	Yellow fever
		Dengue virus	Dengue fever
		West Nile virus	Meningitis, encephalitis
	Togaviruses	Rubella virus	Acute or congenital rubella
	Coronaviruses	SARS	Severe respiratory disease
	Caliciviruses	Norovirus	Gastroenteritis
ssRNA–	Orthomyxoviruses	Influenza virus	Influenza
	Paramyxoviruses	Measles virus	Measles
		Mumps virus	Mumps
		Parainfluenza	Croup, pneumonia, common cold
		Respiratory syncytial virus (RSV)	Pneumonia, influenza-like syndrome
	Rhabdoviruses	Rabies virus	Rabies
	Bunyaviruses	Hantavirus	Viral hemorrhagic fever
	Filoviruses	Ebola virus	Viral hemorrhagic fever
		Marburg	Viral hemorrhagic fever
	Arenavirus	Lassa virus	Viral hemorrhagic fever
ssRNA + with RT	Retroviruses	HIV	AIDS
dsDNA with RT	Hepadna viruses	Hepatitis B virus	Acute or chronic hepatitis, hepatocellular carcinoma

AIDS, Acquired immunodeficiency syndrome; *DNA,* deoxyribonucleic acid; *ds,* double-stranded; *HIV,* human immunodeficiency virus; *RNA,* ribonucleic acid; *RT,* reverse transcriptase; *SARS,* severe acute respiratory syndrome; *ss* single-stranded.

TABLE 8.5 Fungal Infections

Infection	Distribution	Vector	Symptoms
Cryptococcosis	Everywhere	Pigeon feces	Fever, cough, weight loss, pleuritic pain, CNS disturbances
Candidiasis	Normal flora	N/A	Mucocutaneous pain and pruritus at site of infection
Phycomycosis (mucormycosis)	Everywhere	Decayed matter, soil	Rhinocerebral mucormycosis: destruction of CN II, IV, V, VI; erosion of carotid artery; meningitis; brain abscess Pulmonary mucormycosis: dyspnea, chest pain, hemoptysis
Histoplasmosis	River valleys (e.g., California), southwestern USA (Arizona, Nevada)	Bird and bat feces	Flulike: cough, fever, myalgias, weight loss, anemia, leukopenia, thrombocytopenia, painful oropharyngeal ulcers
Coccidioidomycosis (San Joaquin Valley fever)	Semiarid USA (e.g., California), southwestern USA (Arizona, Nevada)	Dust, dirt	Cough, fever, pleuritic chest pain, weight loss, dyspnea, chest pain, CNS disturbances
Blastomycosis	Southeastern USA, south central USA, Midwestern USA, Great Lakes region	Unknown	Flulike: pleuritic chest pain, arthralgias, erythema nodosum, weight loss, fever, cough, chest pain
Aspergillosis	Everywhere	Decaying vegetation	Dyspnea, chest pain, hemoptysis, wheezing

CN, Cranial nerve; *CNS,* central nervous system; *GU,* genitourinary; *N/A,* not applicable.

TABLE 8.6 Parasitic Infections

Parasitic Agent	Common Name of Disease	Location of Infection	Symptoms	Mode of Transmission
Helminths (Worms)				
Nematodes (Roundworms)				
Ancylostoma duodenale	Hookworm	Blood vessels of gut	Anemia	Skin penetration
Ascaris lumbricoides	Giant roundworm	Small intestine, lungs	Pneumonitis (rare), intestinal obstruction (rare)	Oral (fecal contamination), autoinfection
Enterobius vermicularis	Pinworm	Cecum	Anal pruritus	Oral
Onchocerca volvulus	River blindness	Skin, eye	Blindness	Insect inoculation
Strongyloides stercoralis	Strongyloidiasis	Small intestine, lungs	Eosinophilia, urticaria, rash, abdominal pain, pneumonitis	Skin penetration, autoinfection
Trichinella spiralis	Trichinosis	Muscles	Muscular pain, eosinophilia, fever, periorbital edema	Oral (infected meat)
Trichuris trichiura	Whipworm	Intestine	Rectal prolapse	Oral (fecal contamination)
Wuchereria bancrofti	Filariasis	Lymphatics	Elephantiasis	Insect (mosquito)
Trematodes (Flukes)				
Clonorchis sinensis	Liver fluke	Liver	Biliary obstruction (rare)	Oral (raw fish)
Fasciola hepatica	Liver fluke	Liver	Fever, right upper quadrant abdominal pain, eosinophilia	Oral
Fasciolopsis buski	Intestinal fluke	Liver	Abdominal pain, diarrhea	Oral
Paragonimus westermani	Lung fluke	Lung, intestine	Eosinophilia, cough, chest pain, bronchitis	Oral (poorly cooked freshwater crab or crayfish)
Schistosoma haematobium	Blood fluke	Urinary tract	Acute: rash, fever, cough, chest pain, chills	Skin inoculation
Schistosoma japonicum	Blood fluke	Mesenteric blood vessels	Hepatomegaly, splenomegaly	Skin inoculation
Schistosoma mansoni	Blood fluke	Mesenteric blood vessels	Lymphadenopathy, eosinophilia	Skin inoculation
Cestodes (Tapeworms)				
Diphyllobothrium latum	Fish tapeworm	Intestine	Megaloblastic anemia	Oral (poorly cooked fish)
Taenia saginata	Beef tapeworm	Intestine	Mild abdominal pain	Oral (poorly cooked beef)
Taenia solium	Pork tapeworm	Intestine	Mild abdominal pain	Oral (poorly cooked pork)
Echinococcus granulosus	Hydatid cyst	Lung, liver	Cholestasis, liver congestion and atrophy, biliary obstruction	Oral (inoculation with sheep, cattle, or dog feces)
Protozoa				
Entamoeba histolytica	Amebic dysentery	Intestine	Bloody, mucoid diarrhea; colicky abdominal pain	Contaminated water, raw vegetables
Plasmodium spp.	Malaria	Liver, erythrocytes	High fever, chills, rigor, anemia, headache, malaise, chest pain, abdominal pain	Female *Anopheles* mosquito
Leishmania spp.	Kala azar; cutaneous leishmaniasis	Reticuloendothelial cells of body disseminates to spleen, liver, bone marrow, lymph glands	Chronic: abdominal discomfort, ascites, fever, weakness, pallor, weight loss, cough Acute: sudden fever, chills	All transmission accomplished through bite of sandflies after biting specific infected mammals
Trypanosoma spp.				
T. cruzi	Chagas disease	Bloodstream	Local inflammation, lymphadenopathy, muscular necrosis including myocardium (heart failure), esophagus, and colon (dilation); fever, malaise, anorexia, edema of face	Insects—hematophagous Triatoma (blood drinking)
T. brucei	African sleeping sickness	Bloodstream	Fever, malaise, headache, rash, CNS disturbances	Glossina flies (tsetse flies)

Continued

TABLE 8.6 **Parasitic Infections—cont'd**

Parasitic Agent	Common Name of Disease	Location of Infection	Symptoms	Mode of Transmission
Toxoplasma gondii	Toxoplasmosis	Throughout body	Acute: usually asymptomatic Immunosuppressed: encephalitis, myocarditis, pneumonitis Newborn: impaired vision, neurologic disorders	Eating raw or undercooked meat, poultry, or dairy foods; oral inoculation with cat feces
Giardia lamblia	Epidemic diarrhea	Intestine	Acute: self-limited diarrhea; occasionally malabsorption with weight loss	Fecal contamination of water; person to person
Trichomonas vaginalis	Trichomoniasis (vaginitis)	Vagina	Irritation, discharge	Sexually transmitted
Ectoparasites				
Pediculus humanus				
Var. *corporis*	Body louse	All hair-covered parts of body	Pruritus Nits at base of hair shaft	Person to person, by fomites
Var. *capitis*	Head louse	Head area		
Pediculus pubis	Pubic louse	Pubic area		
Sarcoptes scabiei (var. *hominis*)	Scabies	Skin	Pruritus, worse at night; linear burrows in folds of fingers, elbows, knees, axillae, pelvic girdle	Person to person
Maggots (larvae of dipterous flies)	Myiasis	Necrotic tissue	Depends on location of infestation	Dipterous flies
Chiggers (mites)		Skin	Intense pruritus, hemorrhagic papules	Inhabit dogs, rabbits, cats, rats; foul cheese, flour, house dust
Ticks		Skin	Can transmit tick paralysis, Lyme disease	Reside in wooded and grassy areas

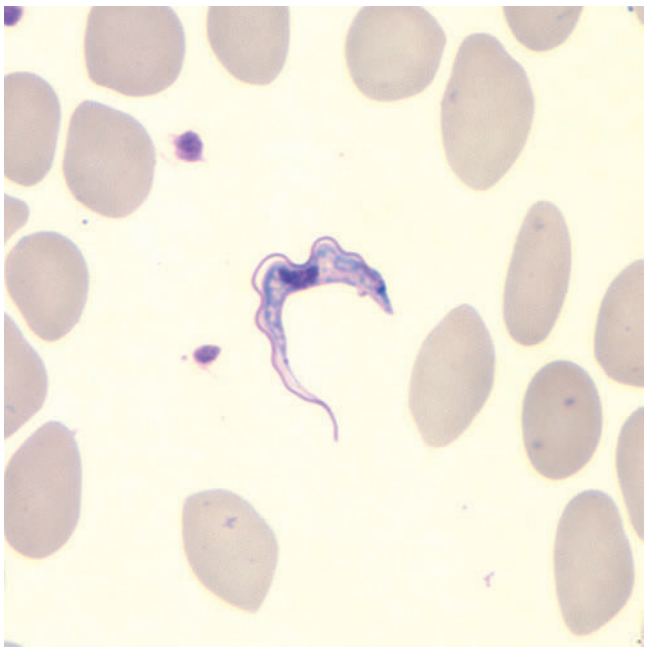


FIG 8.9 *Trypanosoma brucei* parasite in a blood smear. Giemsa-stained light photomicrograph. (Courtesy Blaine Mathison, Centers for Disease Control and Prevention, Atlanta.)

SUMMARY

The host–microbe relationship is determined by the characteristics of both the microorganism and the host. Many microorganisms are commensal and make up the human microbiome that is thought to protect from pathogens. Multiple host factors such as the integrity of barriers to transmission, nutritional status, age, and drug regimen all have an impact on this relationship and may increase the risk of infection with pathogens. Infection is an important cause of morbidity and mortality in health care facilities. Health care professionals have a key role in the prevention, surveillance, and early detection of infectious processes in hospital and community settings. The identification of high-risk individuals who are more susceptible to infection will assist in earlier detection and better management to improve outcomes.

RESOURCES

Commensal and Infectious Organisms

Adlerberth I, Wold AE: Establishment of the gut microbiota in Western infants. *Acta Paediatr* 98:229–238, 2009.

Collado MC, Bäuerl C, Pérez-Martínez G: Defining microbiota for developing new probiotics. *Microb Ecol Health Dis* 23:35–39, 2012.

Jacobi CA, Malfertheiner P: *Escherichia* coli Nissle 1917 (Mutaflor): new insights into an old probiotic bacterium. *Dig Dis* 29:600–607, 2011.

McCance KL, Huether SE: *Pathophysiology: the biologic basis for disease in adults & children*, ed 7, St Louis, MO, 2015, Mosby.

Qin J, Li R, Raes J, et al: A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464:59–65, 2010.

Ryan K, Ray CG, Ahmad N, et al: *Sherris medical microbiology*, ed 6, New York, 2014, McGraw-Hill.

Vallès Y, Gosalbes MJ, de Vries LE, et al: Metagenomics and development of the gut microbiota in infants. *Clin Microbiol Infect* 18:21–26, 2012.

WHO: *Guideline: Nutritional care and support for patients with tuberculosis.* Geneva: World Health Organization; 2013. Available at: http://apps.who.int/iris/bitstream/10665/94836/1/9789241506410_eng.pdf.

Emerging Infectious Diseases

CDC: *Guidelines for Evaluation of US Patients Suspected of Having Ebola Virus Disease.* August 1, 2014. Available at: http://emergency.cdc.gov/han/han00364.asp. (b) WHO. *Ebola Situation Report.* Update, September 16, 2015. Available at: http://apps.who.int/ebola/sites/default/files/atoms/files//who_ebola_situation_report_16-09-2015.pdf?ua=1.

CDC: *Measles cases and outbreaks*, February 23, 2015. Available at: http://www.cdc.gov/measles/cases-outbreaks.html.

CDC: *Zika virus*, June 14, 2016. Available at: https://www.cdc.gov/zika/index.html.

Delves PJ, Martin SJ, Burton DR, Roitt IM: *Roitt's essential immunology*, ed 12, Boston, 2011, Wiley-Blackwell Publishing.

Riedel S: Biological Warfare and Bioterrorism: A Historical Review. *Proc (Bayl Univ Med Cent)* 17(4):400–406, 2004.

WHO: *Middle East respiratory syndrome coronavirus (MERS-CoV)*, Update, July 23, 2014: Available at: http://www.who.int/csr/don/2014_07_23_mers/en/.

9

Inflammation and Immunity

Jacquelyn L. Banasik

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- What are the major organs and cellular components of the body's defense against foreign antigens?
- How do immune cells communicate through cell-to-cell interactions and through secreted cytokines?
- How do innate and adaptive immune mechanisms differ?
- How do macrophages, granulocytes, and lymphocytes work together to locate, recognize, and eliminate pathogens?
- What is the role of MHC class I and II proteins in cell-mediated immunity?
- Why is an immune response usually more effective on subsequent exposure to an antigen than after the first exposure?
- How do noncellular immune system components, including antibodies, complement, and clotting factors, aid the immune response?

CHAPTER OUTLINE

COMPONENTS OF THE IMMUNE SYSTEM, 159
Epithelial Barriers, 159
Mononuclear Phagocyte System, 159
Lymphoid System, 160
Primary Lymphoid Organs, 160
Secondary Lymphoid Organs, 161
Tonsils, 161
Spleen, 161
Lymph Nodes and Lymphatics, 161
Peyer Patches, 162
Leukocytes, 162
Neutrophils, 162
Eosinophils, 163
Basophils and Mast Cells, 164
Monocytes and Macrophages, 164
Dendritic Cells, 165
Lymphocytes, 165
Natural Killer Cells, 166
T Lymphocytes, 166
B Lymphocytes, 167
Chemical Mediators of Immune Function, 167
Complement, 167
Kinins, 169
Clotting Factors, 169
Cytokines and Chemokines, 169
INNATE DEFENSES AND INFLAMMATION, 169
Inflammation, 171
Increased Vascular Permeability, 171
Emigration of Leukocytes, 172
Phagocytosis, 172
Chronic Inflammation, 174
Healing, 174
Inflammatory Exudates, 175
Systemic Manifestations of Inflammation, 175
SPECIFIC ADAPTIVE IMMUNITY, 176
Major Histocompatibility Complex, 176
Antigen Presentation by MHC, 176
MHC Class I Presentation, 177
MHC Class II Presentation, 177
Mechanisms of Cell-Mediated Immunity, 178
T Helper Cells ($CD4^+$), 178
Cytotoxic T Cells ($CD8^+$), 181
Mechanisms of Humoral Immunity, 181
Antigen Recognition by B Cells, 181
Antibody Structure, 183
Class Switching and Affinity Maturation, 185
Antibody Functions, 186
Passive and Active Immunity, 187
Passive Immunity, 187
Active Immunity, 187
INTEGRATED FUNCTION AND REGULATION OF THE IMMUNE SYSTEM, 189
Integrated Response to Microbial Antigen, 189
Integrated Response to Viral Antigen, 191
Regulation of Immune Function, 192

The immune system is a complex network of cells and tissues that work together to protect the body against foreign invaders. The wide variety of potential pathogens requires a defense system that is diversified and adaptable. Several types of white blood cells (WBCs) are important in localizing, recognizing, and eliminating foreign substances. These immune cells are strategically situated in diverse locations so that pathogens may be detected quickly. The dispersed nature of these defensive cells necessitates a complex system of intercellular communication to effectively mobilize reinforcements to areas of need. A tremendous amount of information has accumulated about how immune cells communicate and the processes that enable them to migrate to particular locations. The impact of this research goes far beyond the traditional immune disorders such as immunodeficiency diseases and hypersensitivity reactions. The immune system has been implicated in the pathogenesis of disorders as diverse as atherosclerosis, myocardial infarction, shock, diabetes, and stroke. Therefore an understanding of immune function is fundamental to the study of a wide variety of diseases. This chapter describes the organs and cells that constitute the immune system, the mechanisms of action of innate and adaptive defenses, and the communication processes whereby immune cells achieve a coordinated response. Underreactions and overreactions of the immune system, immune system malignancies, and human immunodeficiency virus disease are described in Chapters 10, 11, and 12, respectively.

COMPONENTS OF THE IMMUNE SYSTEM

The structures of the immune system include (1) skin and mucous membranes; (2) the mononuclear phagocyte system; (3) the lymphoid system, including spleen, thymus gland, and lymph nodes; and (4) bone marrow. All these structures are inhabited by different types of WBCs (leukocytes) that mediate inflammation and immunity. Leukocytes are responsible for locating and eliminating pathogens and foreign molecules. They are aided in their task of bodily defense by a number of chemical mediators, including complement, kinins, clotting factors, cytokines, and chemokines.

Components of the immune system are often categorized into specific or innate defenses according to the mechanisms whereby antigens are recognized. Innate defenses require no previous exposure to mount an effective response against an antigen, and a wide variety of different antigens are recognized. Natural killer (NK) cells and phagocytic cells such as neutrophils and macrophages are mediators of innate defenses. In contrast, specific defenses respond more effectively on second exposure to an antigen (adaptive) and are highly selective in the ability to recognize antigens. B lymphocytes (B cells) and T lymphocytes (T cells) are the agents of specific immunity.

Although separating immune components into specific and innate systems is helpful for studying inflammation and immunity, it is an artificial division because they function in a highly integrated manner. The approach used in this chapter is to first describe the major components of the immune system, discuss innate and specific adaptive defenses separately, and then summarize the integrated function of the entire system and its regulation.

EPITHELIAL BARRIERS

The skin and mucous membranes are sometimes called the "first line of defense" because they are frequently the initial sites of microbial invasion. Intact epithelia in skin and mucous membranes provide mechanical and chemical barriers that prevent microorganisms from gaining access to the body's tissues. The skin epithelium produces antimicrobial peptides called *defensins* that can kill a wide variety of bacteria and fungi. The intestinal epithelium produces another form of bactericidal peptide called *cryptocidins* that prevent bacteria from colonizing the intestinal wall. Resident microorganisms may aid in providing this line of defense by making conditions inhospitable for pathogens (see Chapter 8). Disruption of the normal epithelial barriers increases the likelihood that pathogens will successfully establish an infection. Physical trauma (e.g., burns, lacerations, erosions) and biochemical alterations (e.g., pH changes, increased glucose concentration, decreased enzyme production) predispose to infection. Pathogens that breach the skin or mucous membranes are generally first detected by cells of the mononuclear phagocyte system. These cells are thought to originate from monocytes produced in the bone marrow. Specialized antibody-secreting cells also locate to the mucous membranes where they produce antibodies of the immunoglobulin A (IgA) class. IgA antibodies bind antigens on the mucosal surface and prevent them from entering more deeply into the tissues.

MONONUCLEAR PHAGOCYTE SYSTEM

The mononuclear phagocyte system (previously called the *reticuloendothelial system*) is composed of dendritic cells, monocytes, and macrophages that are widely distributed throughout the body. Monocytes from the circulating blood migrate to organs and tissues to become macrophages. Macrophages are found throughout the body and are assigned various names according to the tissues in which they are located, such as alveolar macrophages in the lungs, microglial cells in the brain, Kupffer cells in the liver, and histiocytes in connective tissue (Fig. 9.1). Tissue dendritic cells are a monocyte-derived cell type that specializes in capturing and presenting antigens to T cells. Dendritic cells are strategically located in subcutaneous and submucosal tissues.

Macrophages and dendritic cells are often the first immune system cells to encounter a pathogen or foreign antigen after it has entered the

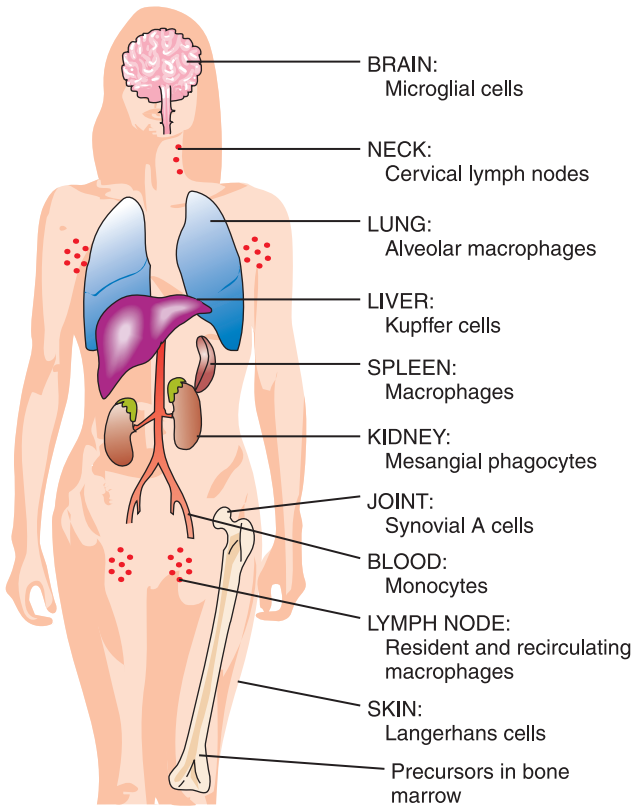


FIG 9.1 Cells of the mononuclear phagocyte system.

body, and they are instrumental in communicating news of the invasion to other immune cells. This communication is accomplished through secretion of chemical signaling molecules called *cytokines* and by presentation of captured antigen to the specific, adaptive immune cells. Dendritic cells specialize in antigen presentation and are able to migrate quickly to lymphoid tissues when they have captured antigens. Macrophages have many other roles in the immune response in addition to their sentry function. Macrophages are powerful phagocytes, each capable of ingesting numerous microbes. Macrophages are called on to clean up the area in which dead neutrophils and inflammatory debris have accumulated after an inflammatory reaction, and they have a role in wound healing.

LYMPHOID SYSTEM

The primary lymphoid organs are the bone marrow and thymus gland, which are the structures where lymphocytes develop. All types of lymphocytes are produced from stem cells in the bone marrow (Fig. 9.2). T lymphocytes then migrate to the thymus for development, whereas B lymphocytes and NK cells stay in the marrow to develop. NK cells are a population of lymphocytes that lack both T-cell and B-cell markers. NK cells are produced and released from the bone marrow and function in innate immune responses. NK cells are found mainly in the circulation and spleen. Once mature, T and B lymphocytes migrate to the secondary lymphoid organs where they await activation by antigens. Secondary lymphoid organs include the tonsils, spleen, lymph nodes, and Peyer patches (Fig. 9.3).

Primary Lymphoid Organs

Bone marrow is contained in all the bones of the body. The primary function of bone marrow is hematopoiesis, or the formation of blood cells. There are two kinds of bone marrow: red and yellow. Hematopoiesis is carried out by red (functioning) marrow. By adulthood, red marrow is confined to the pelvis, sternum, ribs, cranium, ends of the long bones, and vertebral spine. Yellow or fatty bone marrow is found in the remaining bones. It normally does not contribute to hematopoiesis in adults, but

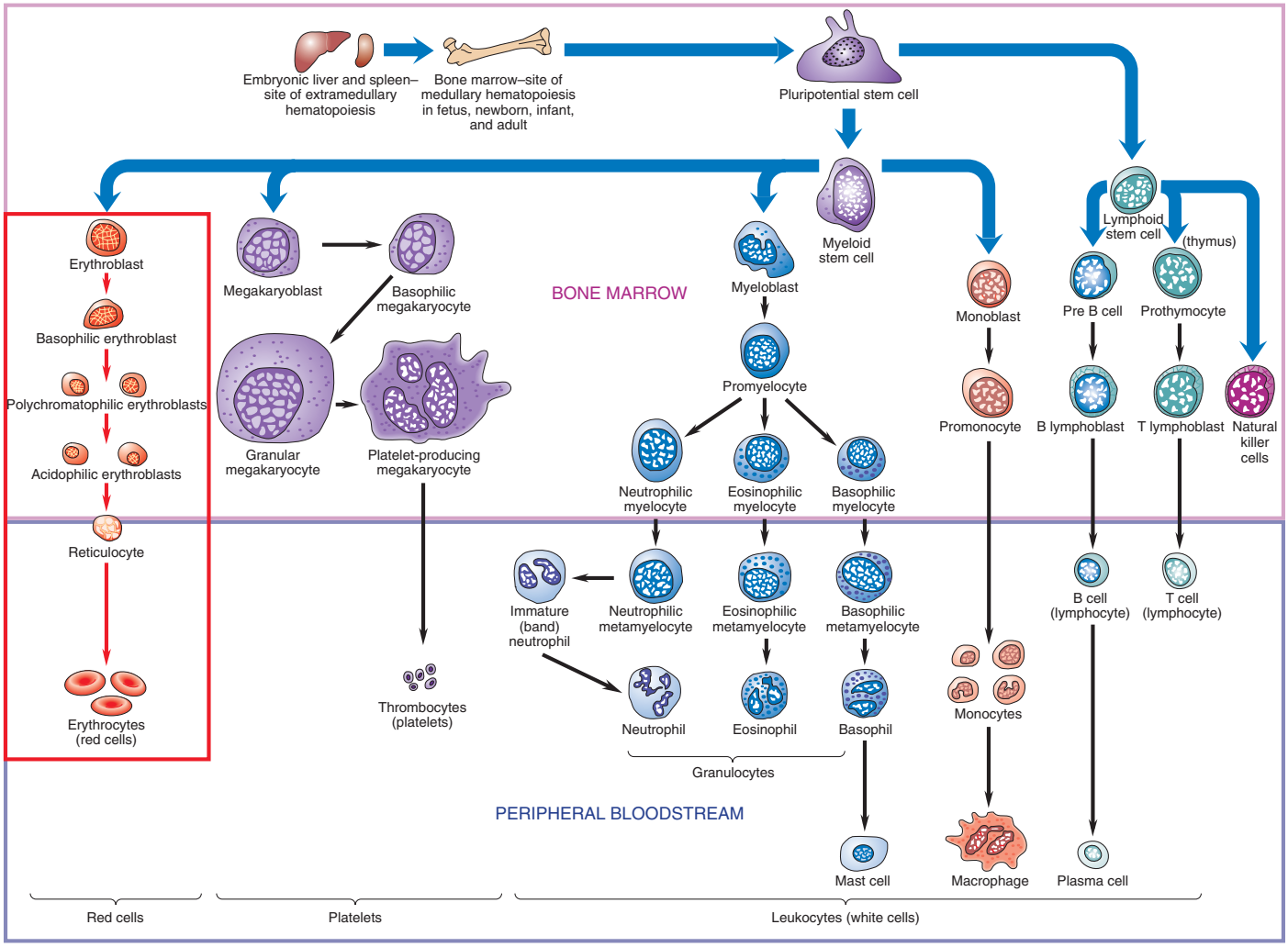


FIG 9.2 Maturation of human blood cells showing pathways of cell differentiation from the pluripotent stem cell to mature granulocytes, monocytes, lymphocytes, thrombocytes, and erythrocytes. Production begins in embryo blood islands in the yolk sac. As the embryo matures, production shifts to the liver, spleen, and bone marrow. In an adult, nearly all hematopoiesis occurs in the bone marrow. The two major differentiation pathways are the myeloid pathway and the lymphoid pathway. The lymphoid pathway produces lymphocytes, whereas the myeloid pathway produces granulocytes, monocytes, platelets, and red blood cells.

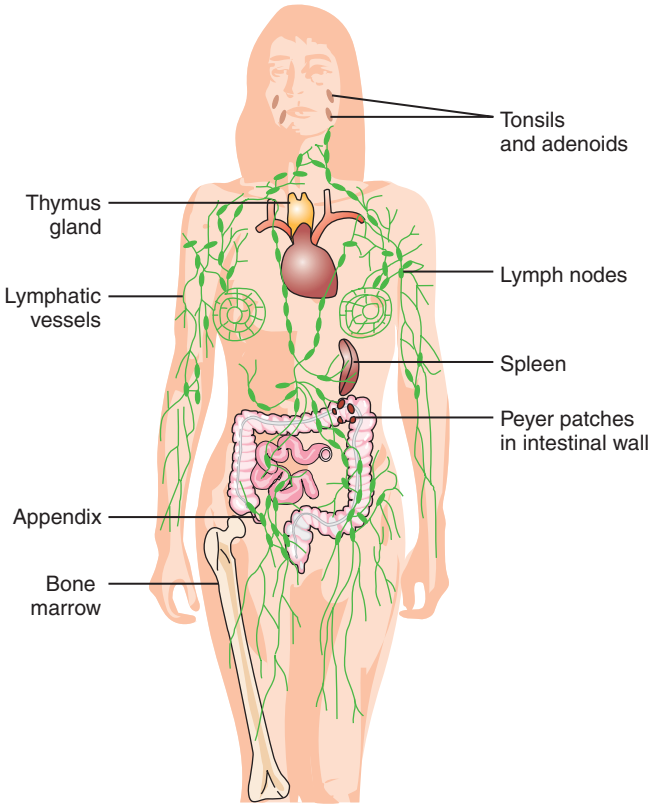


FIG 9.3 Principal organs of the lymphoid system.

can be recruited to become red marrow again under conditions of increased need for hematopoiesis.

B cells are produced and develop in the bone marrow. B cells migrate from the outer edges toward the center of the bone marrow as they develop. Pre–B cells are subjected to a highly selective quality control process, and less than 25% of the developing B cells are allowed to survive. During migration through the bone marrow, immature B cells are exposed to self antigens. B cells that do not bind to self antigens continue to develop into mature B cells. If immature B cells encounter self antigens to which they bind while still in the bone marrow, a series of events is triggered to induce self-tolerance. The immature B cell is stimulated to reactivate its genetic recombination machinery in an attempt to produce a new B-cell receptor (BCR) that does not bind to self antigens. If this receptor-editing attempt fails to alter binding sufficiently, the immature B cell will undergo apoptosis in the bone marrow. Mature B cells that leave the bone marrow to colonize secondary lymphoid organs are called *naive* B cells because they have not yet encountered antigen.

T cells develop in the thymus, which is located in the anterior mediastinum overlying the heart. Pre–T cells initially enter the outer aspect (cortex) of the thymus lobules, and many die while they migrate to the center (medulla) of the thymus. The selection process for T cells is even more rigorous than that for B cells; only about 5% of the cells entering the thymus survive to reenter the circulation and colonize secondary lymphoid organs. The thymus is relatively large at birth and steadily atrophies after puberty. The thymus produces interleukin-7 (IL-7), a cytokine that promotes T-cell proliferation.

Secondary Lymphoid Organs

Once mature, lymphocytes leave their primary lymphoid organs and travel through the blood to localize in peripheral, or secondary, lymphoid tissues, including lymph nodes, spleen, tonsils, and Peyer patches in the intestine. These naive T cells and B cells express specific chemokine receptor proteins on their cell surfaces that allow them to migrate, or "home," to specific locations in lymph tissue. Most mature T lymphocytes are in constant circulation through lymphatic tissues and the bloodstream. It has been estimated that a lymphocyte makes a circuit from the blood to tissues to lymphatics and back to the bloodstream once or twice per day. Antigens can be carried to the naive cells in the lymph nodes by the specialized antigen-presenting dendritic cells. When exposed to an appropriate antigen, T cells and B cells migrate toward each other within the lymph nodes and begin to proliferate. Activated T cells may then migrate to lymph vessels and travel to the bloodstream, where they are dispersed throughout the system. The majority of B cells stay in the lymph node, where they mature into antibody-secreting plasma cells. Lymphocyte recirculation and homing are regulated by binding interactions between various types of cell adhesion molecules, including selectins, integrins, chemokines, and addressins.

Tonsils

Tonsils are aggregates of lymphoid tissue located in the mouth and pharynx. The tonsils are strategically located at the entrance to the digestive and respiratory tracts, where they are likely to encounter microorganisms. Unlike lymph nodes, tonsils have no afferent (incoming) lymphatic vessels. They do have efferent lymphatic drainage so that activated lymphocytes from the tonsils can migrate to other lymphoid organs. Tonsils normally make an important contribution to immune function; however, they may occasionally become chronically infected, and surgical removal (tonsillectomy) is then helpful.

Spleen

The spleen is located under the diaphragm on the left side of the body. It measures about 12 cm in length, which makes it the largest of the lymphoid organs. The spleen provides an important filtering function for blood. The tissue structure of the spleen is similar to that of lymph nodes. It is surrounded by a capsule of connective tissue and filled with a meshwork of red pulp and localized masses of lymphocytes called *white pulp.* Within the red pulp are many blood-filled sinuses lined with macrophages. Macrophages filter out foreign substances and old red blood cells. Lymphocytes located in the white pulp are in a strategic position to come into contact with bloodborne antigens. Lymphocytes thus activated in the spleen can migrate to other lymphoid organs via efferent lymphatics. Like the tonsils, the spleen does not have afferent lymphatic vessels.

Lymph Nodes and Lymphatics

The lymphatic vessels begin with small, closed-ended lymphatic capillaries in direct contact with the interstitial fluid surrounding cells and tissues. Lymphatics pick up fluid and proteins that escape the bloodstream and return them to the circulation by way of the right lymphatic and thoracic ducts. Along the way from lymphatic capillaries to the thoracic ducts, lymph flows through specialized structures called *lymph nodes.* Lymph nodes are found primarily in the neck, axilla, thorax, abdomen, and groin. They often become tender and palpable when responding to foreign invaders. Projections of connective tissue called *trabeculae* divide the interior of the lymph node into compartments (Fig. 9.4). Lymph nodes contain large numbers of B and T lymphocytes and macrophages. B cells are the predominant cell type in the cortical follicles, whereas T cells predominate in the area just under the cortex called the *parafollicular.* The central region, or medulla, is populated by macrophages, B cells, and plasma cells (antibody-secreting B cells). Lymph fluid flows through the nodes in a way that allows these immune cells to filter, detect, and react to foreign material.

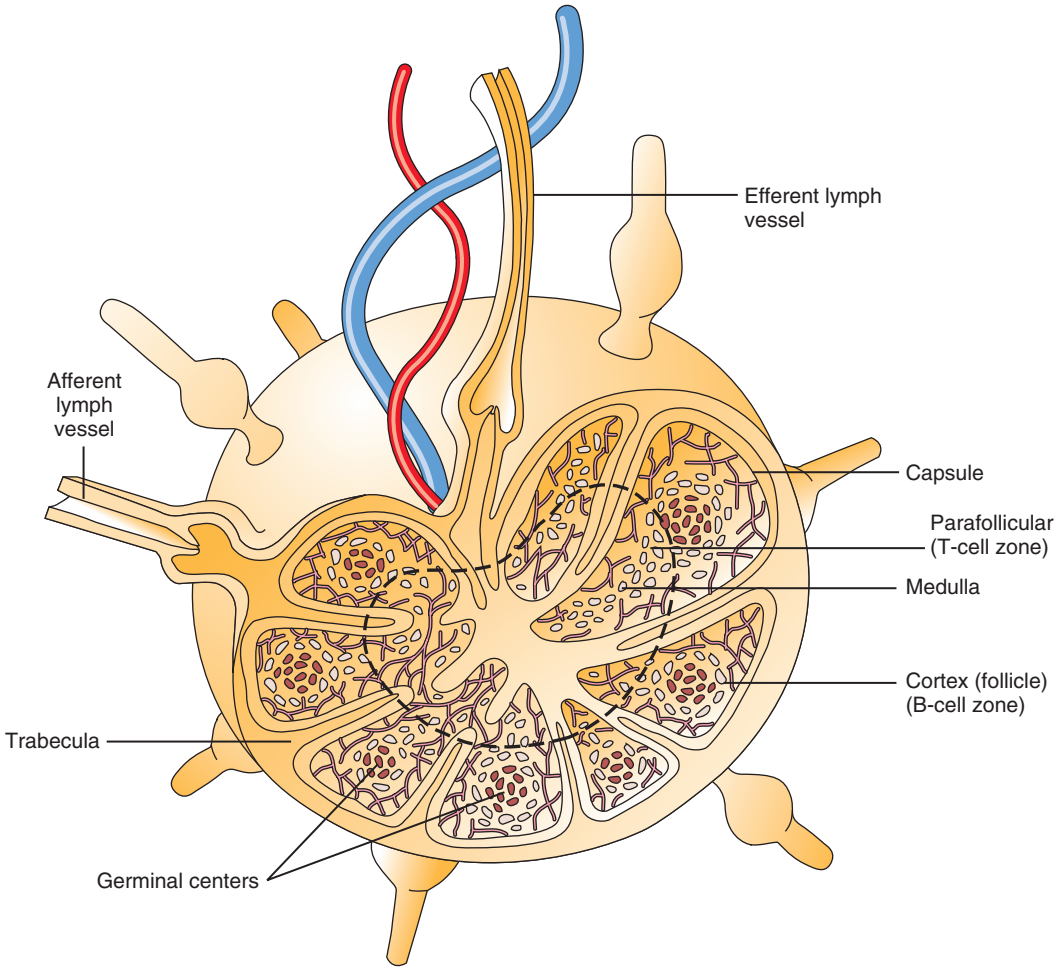


FIG 9.4 Schematic drawing of a typical lymph node showing afferent and efferent lymph vessels, as well as B-cell and T-cell zones.

Peyer Patches

Aggregates of lymphoid tissue can be found scattered throughout the body, particularly in the gastrointestinal, respiratory, and urogenital tracts. These structures are analogous to lymph nodes, but they are not encapsulated and contain primarily B cells. Because of their location, these structures have been termed *mucosa-associated lymphoid tissue* or *gut-associated lymphoid tissue*. These structures, also called *Peyer patches,* are of particular importance in producing antibodies to microorganisms that tend to invade mucosal tissue.

LEUKOCYTES

Leukocytes, or WBCs, are the primary effector cells of the immune system. Each of the different types of leukocytes found in blood has a special job to perform. All leukocytes, as well as red blood cells and platelets, are formed from stem cells in the bone marrow. Stem cells can produce daughter cells that differentiate along several different pathways to become mature cell types (see Fig. 9.2). The first major differentiation step produces either a lymphoid stem cell or a myeloid stem cell. Lymphoid stem cells further differentiate to form B and T lymphocytes and NK cells. Myeloid stem cells can produce a variety of cell types, including red blood cells, platelets, monocytes, dendritic cells, and granulocytes. Monocytes that migrate from the blood into tissues are called *macrophages.* Granulocytes are further divided into neutrophils, eosinophils, and basophils. Basophils are precursors of the mast cells located in tissues.

Development of these cell types is influenced by hormonal signaling molecules called *cytokines.* Cytokines are produced locally in the bone marrow and by various other cells. Certain cytokines stimulate stem cell growth, proliferation, and differentiation into particular cell types. The WBC count and differential are commonly measured laboratory tests used to evaluate WBC production. A normal WBC count and differential are shown in Table 9.1. The general features of each of the WBC types are summarized in the following sections.

Neutrophils

Neutrophils are circulating granulocytes that are also known as *polymorphonuclear leukocytes* (polys or PMNs). They account for 60% to 80% of the total WBC count. Neutrophils normally have two to five nuclear lobes and coarse, clumped chromatin. Neutrophils arise from bone marrow stem cells and undergo several stages of maturation. As illustrated in Fig. 9.2, these stages, from least to most mature, are myeloblast, promyelocyte, metamyelocyte, band cell, and mature segmented neutrophil.

Neutrophils stored in the bone marrow outnumber, by about tenfold, the quantity of circulating neutrophils. An adult produces more than 1×10^{11} neutrophils each day. These stored neutrophils are released into the circulation, where they have a half-life of 4 to 10 hours. Neutrophils that are not recruited into tissues within about 6 hours undergo programmed

cell death (apoptosis). Neutrophils are early responders to an acute bacterial infection and arrive in large numbers very quickly. They are phagocytes that engulf and degrade microorganisms. Circulating neutrophils have receptors on their cell surfaces that enable them to bind to endothelial cells in areas of inflammation. These receptors, called *L-selectins,* and chemokine receptors allow neutrophils to adhere and roll along the capillary surface. Other interactions between neutrophil integrin receptors and extracellular matrix then facilitate movement of neutrophils through the capillary wall and into the tissue. Neutrophils are attracted to areas of inflammation and bacterial products by chemotactic factors such as complement fragments and cytokines. This process is discussed in more depth in the section titled "Inflammation."

TABLE 9.1 Leukocyte Proportions and Functions

Type	Percentage*	Role in Inflammation
Neutrophils	60–80	First to appear after injury, phagocytosis
Lymphocytes	20–30	Immune response
Monocytes (macrophages)	3–8	Phagocytosis
Eosinophils	1–6	Allergic reactions, parasite infection
Basophils	0–2	Contain histamine, mediate type I allergic reactions, initiate inflammation

*Total white cell count, 4000 to 10,000/μL.

Initially during an acute infection, neutrophilia, or an increase in the number of circulating neutrophils, occurs as the bone marrow releases stored neutrophils. As neutrophils are consumed and demand exceeds production, an increase in the number of immature (band) neutrophils occurs. Bands are identified by their lack of nuclear segmentation. This increase in band cells is referred to as a "shift to the left of normal" (Fig. 9.5). Traditionally, the band count has been used to differentiate bacterial from viral infections, and a greater shift to the left is viewed as a more severe infection.

Neutrophils produce potent chemical mediators that enable them to destroy microorganisms. Numerous toxins released by neutrophils have been identified, including oxidizing free radicals, defensins, and proteolytic enzymes, such as elastase. Because of the ability to generate free radicals and release enzymes, neutrophils can cause extensive damage to normal tissue during their inflammatory response.

Eosinophils

Eosinophils are circulating granulocytes that have two nuclear lobes and stain brilliant red-orange with eosin. They constitute 1% to 6% of the total WBC count. Eosinophils mature in the bone marrow (3 to 6 days) and circulate in the blood for about 30 minutes. They have a half-life of 12 days in tissue. Eosinophils arise from myeloid stem cells and undergo a maturation process similar to that of neutrophils.

Eosinophils are particularly associated with an increase in number during allergic reactions and infection by intestinal parasites. The role of eosinophils in allergic reactions is less well characterized than that of mast cells. Eosinophils are recruited into areas of inflammation by the chemokine eotaxin, which is produced by epithelial cells. Eosinophils release inflammatory chemicals, such as lysosomal enzymes, peroxidase,

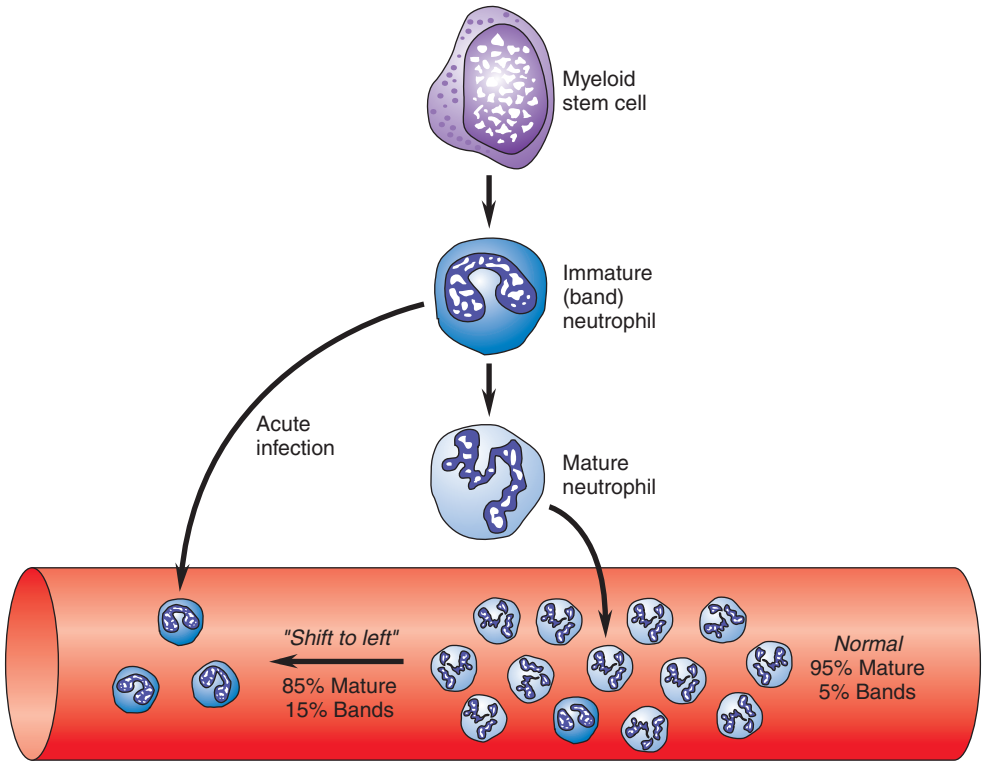


FIG 9.5 Inflammatory cytokines stimulate the release of more immature neutrophils, called bands, from the bone marrow. An increased ratio of bands to mature neutrophils is termed a "shift to the left." This clinical term evolved from the practice of listing bands to the left of mature cells on the laboratory report sheet. A shift to the left is commonly seen with acute bacterial infections.

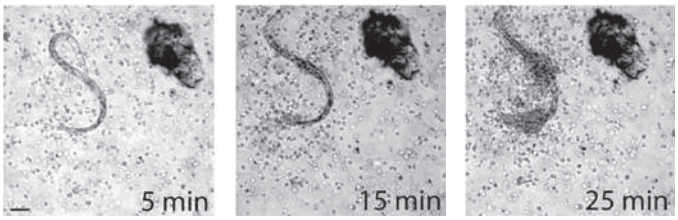


FIG 9.6 Small dots are eosinophils migrating toward and attacking the worm over time. (From Patnode ML, Bando JK, Krummel MF, et al: Leukotriene B4 amplifies eosinophil accumulation in response to nematodes. *J Exp Med* 211(7):1281-8, 2014.

major basic protein, and cationic protein. The primary function of eosinophils is to kill parasitic helminths (worms). Helminths are too large to be phagocytosed by neutrophils or macrophages, and their exterior is resistant to attack by complement or mast cell products. Eosinophils produce specialized molecules such as major basic protein and eosinophil cationic protein, which may be more effective against helminths (Fig. 9.6). Eosinophils recognize helminths that have been opsonized (coated) with IgE antibody. They bind to the IgE and then release their stored chemicals onto the surface of the opsonized helminth. Parasitic infections are a significant problem in much of the world, with one-third of the population being affected.

Basophils and Mast Cells

Basophils are granulocytes characterized by granules that stain blue with basophilic dyes. Basophils account for 0% to 2% of the total leukocyte count. Basophils are structurally similar to mast cells. Mature basophils circulate in the vascular system, whereas mast cells are found in connective tissue, especially around blood vessels and under mucosal surfaces. When stimulated by cytokines, mature basophils can migrate to connective tissue, but once in the tissue, basophils (then called *mast cells*) do not reenter the circulation.

The average basophil life span is measured in days, whereas mast cells can live for weeks to months. Mast cells and basophils have IgE receptors that allow them to bind and display IgE antibodies on their cell surfaces. When an appropriate stimulus occurs, such as antigen binding to the IgE antibodies, mast cells and basophils release granules (degranulate) containing proinflammatory chemicals.

Mast cell and basophil granules contain histamine, platelet-activating factor, and other vasoactive amines that are important mediators of immediate hypersensitivity responses (Fig. 9.7). Degranulation of mast cells and basophils begins the inflammatory response that is characteristically associated with allergic reactions. Mast cells and basophils are also involved in wound healing and chronic inflammatory conditions (see Chapter 10).

Monocytes and Macrophages

Monocytes and macrophages, like granulocytes, originate from bone marrow stem cells of the myeloid lineage. Monocytes are immature macrophages and account for about 5% of the total WBC count. Monocytes circulate in the bloodstream for about 3 days before they enter tissue to become macrophages. As described earlier, macrophages are found in widespread locations as part of the mononuclear phagocyte system.

Phagocytosis by macrophages is similar to that by neutrophils except that neutrophils are short-lived and die in the process of fighting infection. Macrophages, in contrast, may live for months to years and can migrate in and out of tissue. Macrophages are more efficient phagocytes than neutrophils and can ingest several times as many microorganisms

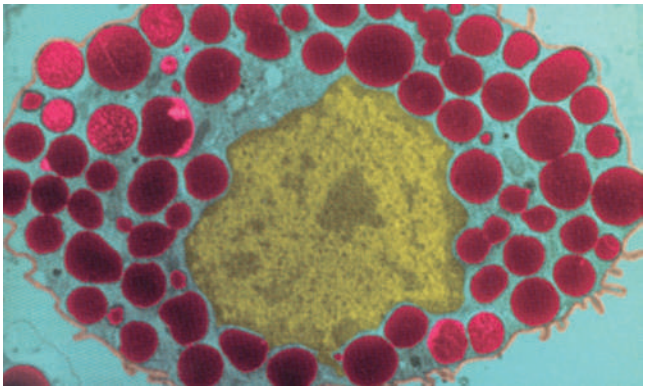


FIG 9.7 Micrograph of a mast cell showing a large yellow nucleus and numerous packets containing histamine, which are colored red. (From Roitt IM, Brostoff Male DK: *Immunology,* ed 3, St Louis, 1993, Mosby.)

FIG 9.8 Scanning electron micrograph of a macrophage (red) attaching to and phagocytizing bacteria (yellow). (From Nairn R, Helbert M: *Immunology for medical students,* ed 2, St Louis, 2007, Mosby, p 3. With permission from Juergen Berger, Max-Planck Institute, and the Science Photo Library.)

(Fig. 9.8). Macrophages are capable of cell division and may proliferate at the site of inflammation.

Macrophages are covered with a variety of receptor proteins on their cell surface (Fig. 9.9). Some of these receptors help macrophages locate antigens that have been coated by antibodies. These receptors are called *Fc receptors* because they bind to the part of an antibody called the *constant fragment,* or *Fc.* Macrophages also have receptors for the complement component C3b. Complement, like antibodies, can coat an antigen and make it more recognizable to macrophages. Coating of antigen by antibodies or by complement is called opsonization. Macrophages have receptors that help them recognize bacteria directly. These innate pattern-recognition receptors bind to particular molecules prevalent in the bacterial cell wall. For example, mannose receptors and numerous toll-like receptors on macrophages allow them to recognize

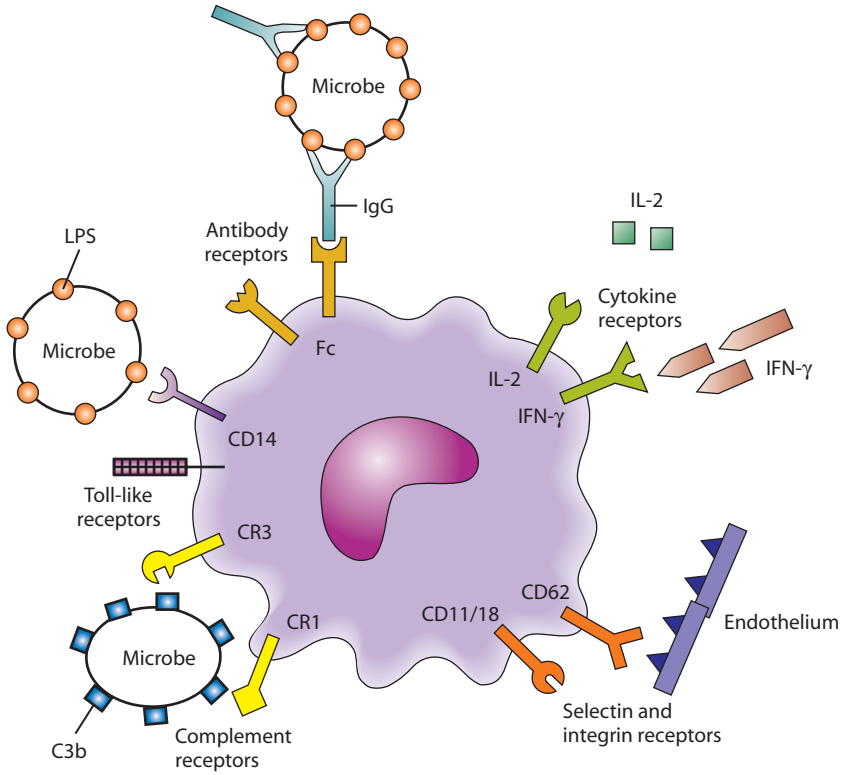


FIG 9.9 Macrophage surface receptors. Macrophages display receptors for a number of extracellular molecules that enhance their function such as cytokines, complement, selectins, integrins, and antibody (Fc). Toll-like receptors recognize patterns of microbial components and trigger intracellular signaling cascades in the macrophage. *IFN-γ,* Interferon-γ; *IL,* interleukin; *LPS,* lipopolysaccharide.

common microbial structures (see Fig. 9.9). Other receptors, called selectins, chemokine receptors, and integrins help macrophages adhere to capillary walls and enter and move through tissue. Integrin receptors bind to proteins in the extracellular matrix and help macrophages target, or "home," to certain areas.

In addition to their phagocytic function, macrophages have an important secretory function. Some of the substances secreted by macrophages are cytokines, which help coordinate the activities of other immune cells (Fig. 9.10). Macrophage cytokines include IL-1, IL-6, IL-12, and tumor necrosis factor-α (TNF-α). These cytokines promote inflammation, as well as the activity of other WBCs, including neutrophils and lymphocytes (see the section titled "Cytokines and Chemokines").

Macrophages secrete a number of proteins that are important in wound healing. Some of these proteins are enzymes that degrade tissue (e.g., collagenase, elastase, plasminogen activator), whereas others stimulate the growth of new granulation tissue (e.g., fibroblast growth factor, angiogenic factors).

A third function of macrophages, in addition to phagocytosis and secretion, is antigen presentation. For T cells to recognize antigens, these antigens must first be processed and presented on the surface of an antigen-presenting cell such as dendritic cells, macrophages, or B cells. Macrophages accomplish this task by first engulfing the antigen, then processing it into smaller pieces, and finally combining the antigen fragments with special membrane proteins. The antigen complexes are then displayed on the macrophage cell surface, where T lymphocytes (T helper cells) can recognize and become activated by them. Antigen presentation is explored in more detail in the section titled "Specific Adaptive Immunity."

Dendritic Cells

Dendritic cells are derived in the bone marrow from the same progenitor cells that produce monocytes and macrophages and are structurally and functionally similar to macrophages. Dendritic cells derive their name from an unusual shape that has extensive projections from the surface (Fig. 9.11). Dendritic cells are located throughout the body as part of the mononuclear phagocyte system. They function primarily as antigen-presenting cells, capturing antigen in tissues and then migrating to lymphoid areas to present antigen to T cells. Some types of dendritic cells produce chemical messengers called *type I interferon* (IFN-α and IFN-β) in response to viral infections. Type I interferons suppress the viral replication machinery in nearby cells and help stop the local spread of the virus.

Lymphocytes

The three major types of lymphocytes are NK cells, T cells, and B cells. NK cells function in innate immunity, whereas B and T lymphocytes are the cells responsible for specific, adaptive immunity. B and T cells have the capacity to proliferate into "memory cells," which provide long-lasting immunity against specific antigens. NK, T, and B cells are derived from a common lymphoid stem cell in the bone marrow that is stimulated to proliferate by bone marrow–derived cytokines, including IL-7. T cells then migrate to the thymus, where they mature. B cells remain in the bone marrow during their maturation phase. NK cells are released into the circulation. Together NK, B, and T lymphocytes compose approximately 20% of the total WBC count. Mature NK cells circulate and populate the spleen, whereas T and B cells migrate to secondary lymphoid organs.

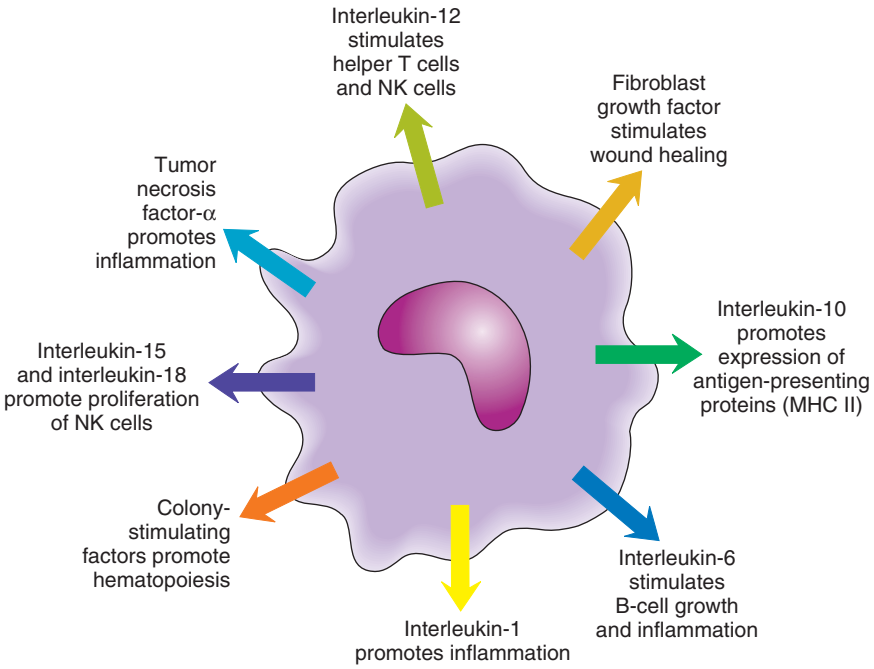


FIG 9.10 Macrophages are of central importance in initiating inflammation and recruitment of other leukocytes to areas of need. Macrophages secrete a variety of cytokines that induce inflammation and chemotaxis. Some macrophage cytokines stimulate the growth and differentiation of other white blood cell types.

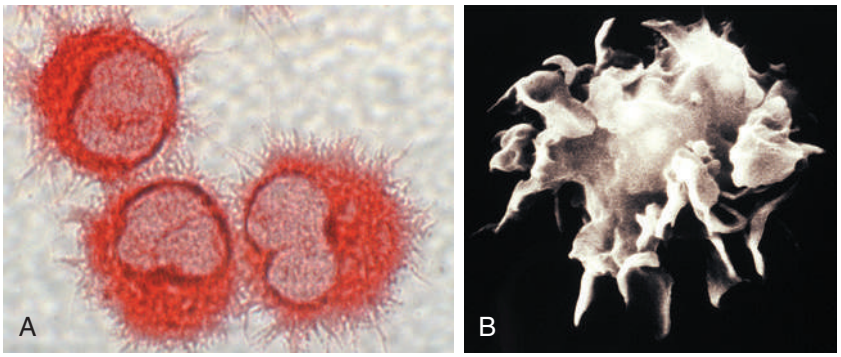


FIG 9.11 Dendritic cell morphology. **A,** Light micrograph of resting dendritic cells from the bone marrow. **B,** Scanning electron micrograph of a mature dendritic cell showing extensive projections of the cell membrane. (From Abbas AK et al: *Cellular and molecular immunology,* ed 8, Philadelphia, 2015, Saunders, p 112. Courtesy of Dr. Y-J Liu, MD, Anderson Cancer Center, Houston, TX.)

Structurally, lymphocytes are small, round cells with a large, round nucleus. Despite their relatively uniform appearance, lymphocytes can be sorted into a number of subpopulations based on characteristic surface proteins called *cluster of differentiation (CD) markers.* More than 350 different CD markers have been identified thus far, with different immune cell types displaying different combinations on their cell surfaces. Lymphocytes have many complex and differentiated functions, and only the major lymphocyte subtypes are discussed in this chapter.

Natural Killer Cells

NK cells have no B- or T-cell markers and are not dependent on the thymus for development. NK cells are considered innate immune cells because they can effectively kill tumor cells and virally infected cells without previous exposure. NK cells kill their target cells by a mechanism similar to that used by cytotoxic T cells. Unlike T and B cells, NK cells can respond to a variety of antigens and are therefore not specific for a particular antigen. Like neutrophils and macrophages, NK cells recognize antibody-coated target cells with their Fc receptors. This process is called *antibody-dependent cell-mediated cytotoxicity*. NK cells also target virally infected cells and tumor cells. They are thought to be able to recognize virally infected cells through innate pattern-recognition receptors; however, only those cells that lack certain normal self proteins on their cell surface (major histocompatibility complex I, or MHC I, proteins) are targeted for killing. Cells that display normal MHC I on their cell surfaces are protected from NK cell cytotoxicity, but will be susceptible to killing by cytotoxic T cells that recognize viral antigen displayed on the MHC I proteins.

T Lymphocytes

Two major classes of T lymphocytes can be differentiated by the presence or absence of CD4 and CD8 surface proteins (Fig. 9.12). T cells that possess CD4 proteins ($CD4^+$) are called *T helper cells.* T helper cells interact with antigens presented on the surface of specialized antigen-presenting cells such as dendritic cells, macrophages, and B cells. T

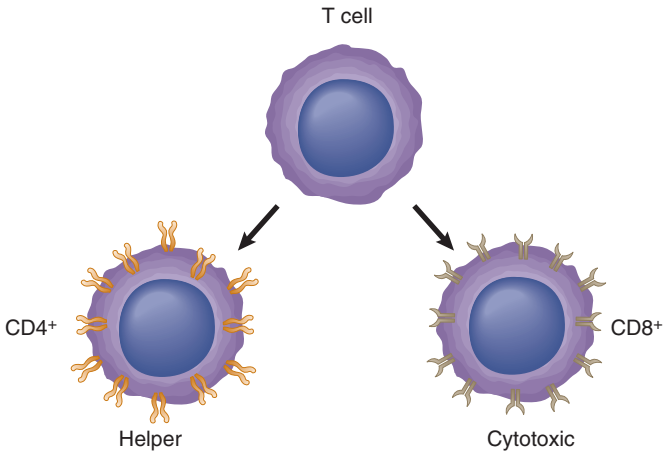


FIG 9.12 Two major classes of T lymphocytes can be differentiated by CD markers on the cell surface. T helper cells have CD4 markers, whereas cytotoxic T cells have CD8 markers.

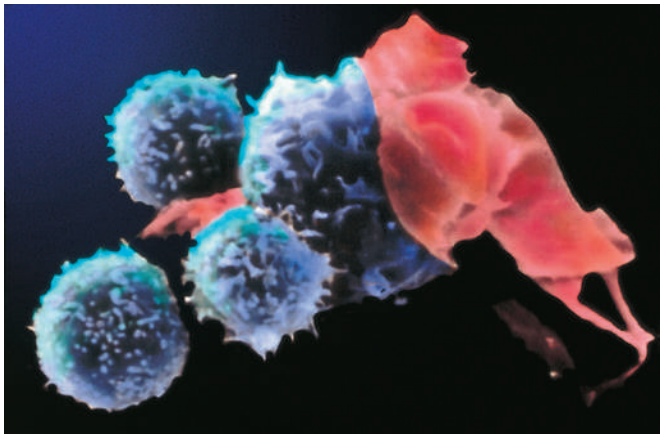


FIG 9.13 Scanning electron micrograph of activated T cells *(blue)* and a tumor cell *(red).* (From Nairn R, Helbert M: *Immunology for medical students,* ed 2, St Louis, 2007, Mosby, p 3. With permission from BSIP Lecaque and the Science Photo Library.)

helper cells can be further divided into subclasses based on the types of cytokines that they secrete. These subtypes are described in the section "Mechanisms of Cell-Mediated Immunity."

The presence of CD8 protein ($CD8^+$) on a T lymphocyte characterizes it as a cytotoxic T cell. Cytotoxic T cells recognize antigen presented in association with surface proteins that can be found on all nucleated cells of the body (MHC I). When a $CD8^+$ T cell recognizes a foreign antigen on a cell, the antigen-presenting cell is killed, thus the name *cytotoxic T cell.* $CD8^+$ cells are particularly effective at destroying virally infected cells, foreign cells, and mutant cells (Fig. 9.13). Proliferation of activated cytotoxic T cells is enhanced by T helper cell cytokines, particularly IL-2.

B Lymphocytes

B cells are distinguished from other lymphocytes by their ability to produce antibodies and by the presence of antibody-like receptors (BCRs) on their cell surfaces. Each B cell carries many copies of identical BCRs and is able to respond to one specific antigen epitope (Fig. 9.14). Naïve B cells require "help" from T helper cells to respond efficiently to protein antigens. B cells bind and internalize the protein antigen and then process and present it to T helper cells. T cells that recognize the presented peptides bind to and are activated by the B cell. T-cell help is provided

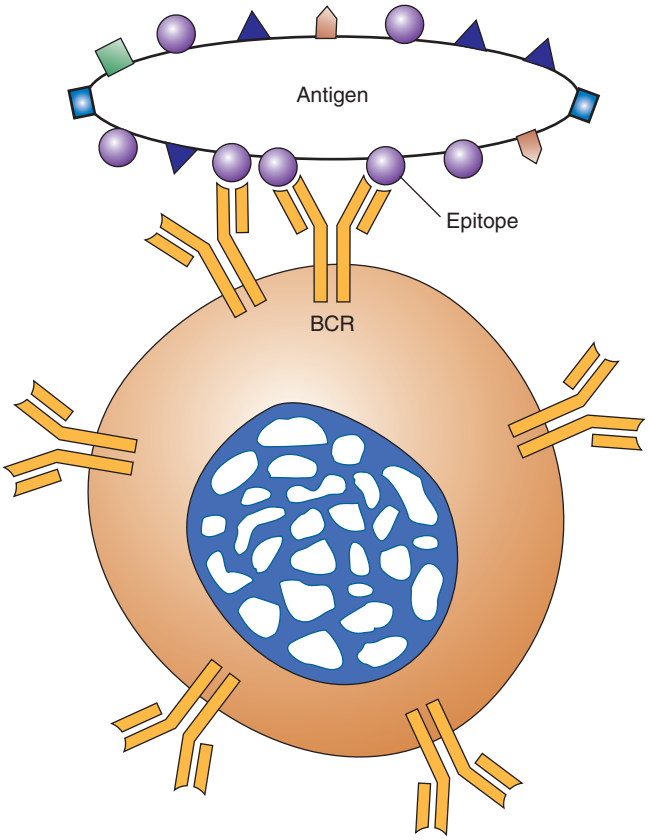


FIG 9.14 Typical B cell showing a number of identical B-cell receptors (BCRs) on the cell surface. Each BCR is capable of binding to two identical antigen epitopes.

to the B cell by physical cell-to-cell contact through coreceptor binding, as well as through secreted cytokines. Some B cells can respond to nonprotein antigens such as bacterial carbohydrate and lipid molecules. B-cell responses to nonprotein antigens are T-cell independent because T cells respond only to peptide antigens. Exposure to antigens stimulates B cells to mature into antibody-secreting plasma cells and memory cells. B-cell memory cells form a reserve of cells that can quickly mount an immune response on subsequent exposure to the same antigen. Memory cells are able to survive for months to years, whereas most antibody-secreting plasma cells live for only a few days. Plasma cells are able to secrete antibodies at a rate of about 2000 per second per cell. Memory B cells and plasma cells develop in germinal centers located in secondary lymphoid organs, including the lymph nodes and spleen. A few long-lived plasma cells inhabit the bone marrow and continue to produce a low level of antibody, which provides immediate protection on second exposure to the same antigen. The mechanisms of specific adaptive immunity are explored later in the chapter.

CHEMICAL MEDIATORS OF IMMUNE FUNCTION

Complement

The complement system consists of about 30 plasma proteins that interact to enhance inflammation, chemotaxis, and lysis of target cells. Complement proteins are synthesized in the liver and by macrophages and neutrophils. They circulate in the blood in an inactive form. Activation of the complement cascade occurs via three different pathways: classical, alternative, and lectin. In all three pathways, the inactive complement proteins are converted to their active form in a sequence

of reactions. Major actions of complement proteins include cell lysis, facilitation of phagocytosis by opsonization, inflammation, and chemotaxis (Fig. 9.15).

The classical pathway is usually triggered by IgG or IgM antibody–antigen complexes. The alternative pathway can be initiated on first exposure to an antigen. Lipopolysaccharide, in bacterial cell walls, and bacterial endotoxin are effective triggers of the alternative pathway. In the classical pathway, an antibody hooked onto an antigen combines with C1, the first of the complement proteins. This step sets in motion a domino effect called the *complement cascade* (Fig. 9.16). The alternative pathway begins with the activation of C3. The alternative pathway can be activated on first exposure and is part of the innate immune response. The lectin pathway also can be triggered on first exposure. Lectin is a circulating biomolecule that binds mannose on bacterial cell walls and triggers complement activation at C2 and C4. C3 spontaneously degrades into active C3b fragments in plasma. If microbial cell surfaces are present, the C3b fragment can bind directly to the microbe. Two other complement proteins, factors B and D, combine with C3b to initiate the alternative pathway. C3 is the most important and plentiful of the complement proteins. C3 divides into two fragments called *C3a* and *C3b.* C3a is a proinflammatory protein that causes histamine release from mast cells, contraction of smooth muscle, and increased endothelial cell permeability. C3b initiates the next step in the cascade by cleaving C5 into its active fragments: C5a and C5b. Complement protein fragment C5a is both a powerful inflammatory chemical and a potent chemotactic agent. C5a chemotaxis stimulates neutrophils and monocytes to migrate to the inflamed tissue. C5a also activates neutrophils by triggering their oxidative activity and increasing their glucose uptake.

The C5b fragment combines with C6, C7, C8, and multiple units of C9 to form a large porelike structure (C5b6789) called the *membrane*

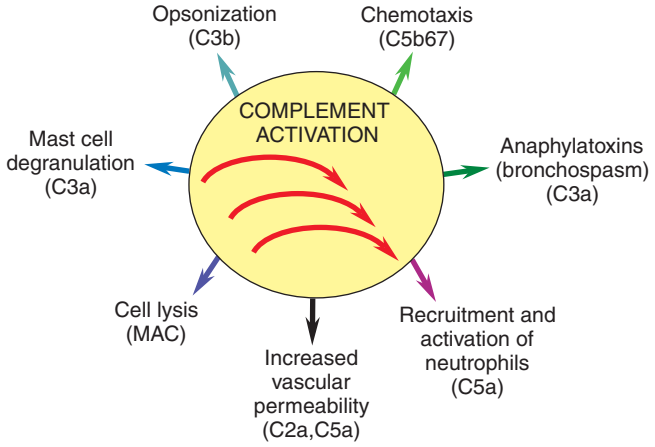


FIG 9.15 Activation of the complement cascade results in the production of products that perform a variety of functions to augment the immune response. *MAC,* Membrane attack complex.

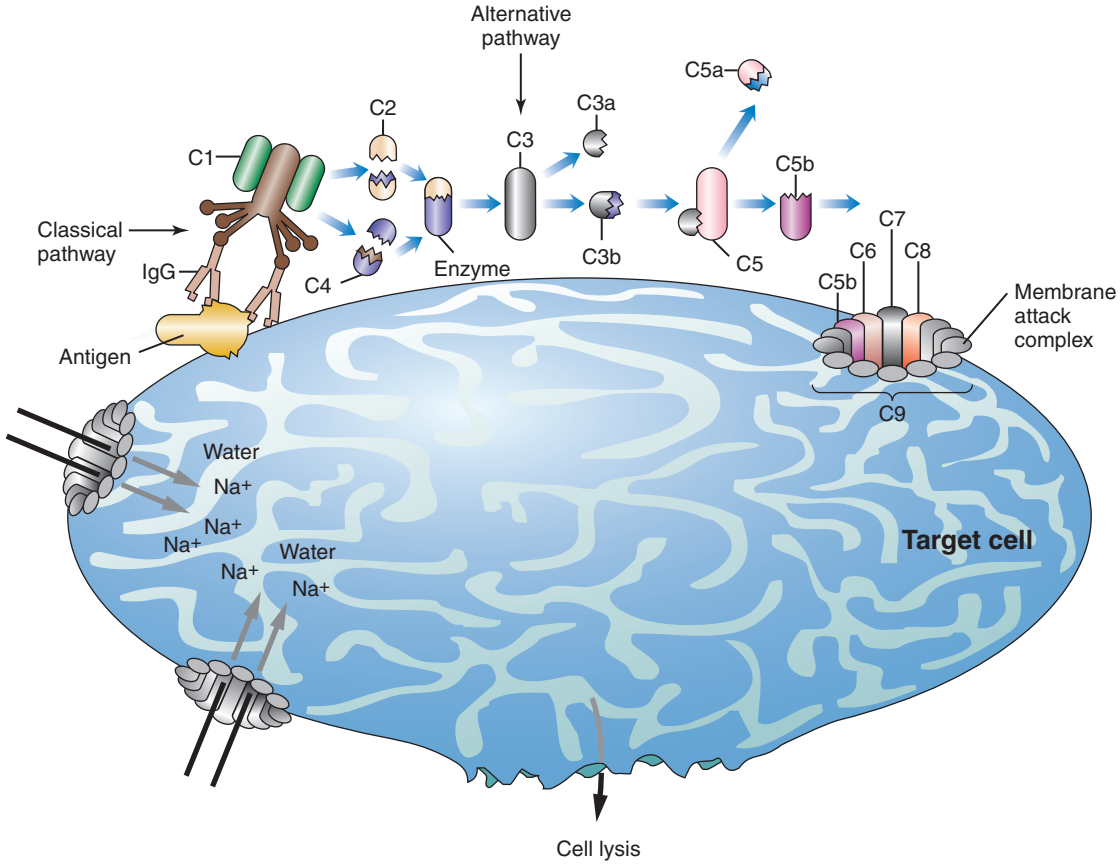


FIG 9.16 Complement cascade. The cascade is activated by the first complement molecule, C1, that binds an antigen–antibody complex. This event begins a domino effect, with each of the remaining complement proteins performing its part in the attack sequence. The end result is a hole in the membrane of the offending cell and destruction of the cell. Activation of the complement cascade results in the formation of membrane attack complexes that insert in the cell membrane. These porelike structures allow sodium and water influx that causes the cell to swell and rupture.

attack complex. The membrane attack complex has a direct cytotoxic effect by attacking cell membranes and disrupting the lipid bilayer. This action allows free movement of sodium and water into the target cell, which causes it to rupture (see Fig. 9.16). The complement system is a potent inflammatory and cytotoxic system that is carefully regulated by inhibitory factors. Normal host cells produce membrane and inhibitory plasma proteins that prevent complement binding to their surface (e.g., C1 inhibitor, protein S).

Kinins

Bradykinin and kallidin are two of the many kinins present in the body. Kinins are small polypeptides that cause powerful vasodilation. They are especially active in the inflammatory process. The kinin system is linked to the clotting system via the Hageman factor (XII) and is activated with the activation of clotting. The first step in this process is the conversion of factor XII to factor XIIa (Fig. 9.17). Factor XIIa converts a substance known as *prekallikrein* to kallikrein. Kallikrein converts precursor substances known as *kininogens* to kinins. The most prevalent is kallidin, which is then converted to bradykinin. Activated kinins cause increased vascular permeability, vasodilation, and smooth muscle contraction. Kinins are also responsible for pain, which is one of the classic signs of inflammation.

Clotting Factors

The blood coagulation cascade's major purpose is to stop bleeding. It is also intimately involved in inflammation and triggering of the kinin system. The key linkage between the inflammatory response and clotting system is activated factor XII (Hageman factor) (see Fig. 9.17). (The blood coagulation cascade is discussed in detail in Chapter 14.) Activation of the coagulation cascade results in the formation of insoluble fibrin strands, which provide an effective barrier to the spread of infection. Clot formation also activates the fibrinolytic cascade, which splits fibrin proteins. Some of the fibrin degradation products are chemotactic signals for neutrophils.

Cytokines and Chemokines

Cytokines are polypeptide signaling molecules that affect the function of other cells by stimulating surface receptors. Cytokines function in a complex intercellular communication network. WBC cytokines have previously had many names, including monokines, lymphokines, and interleukins, depending on their cell of origin. The number of known cytokines is large and growing; they can be grouped according to their source and function (Table 9.2).

These cytokines generally function as chemotactic factors (chemokines), antiviral factors, mediators of inflammation, hematopoietic factors, or activation signals for specific types of WBCs. Cytokines function to enhance and coordinate both innate and specific immune defenses. They are discussed in more detail in the sections that follow.

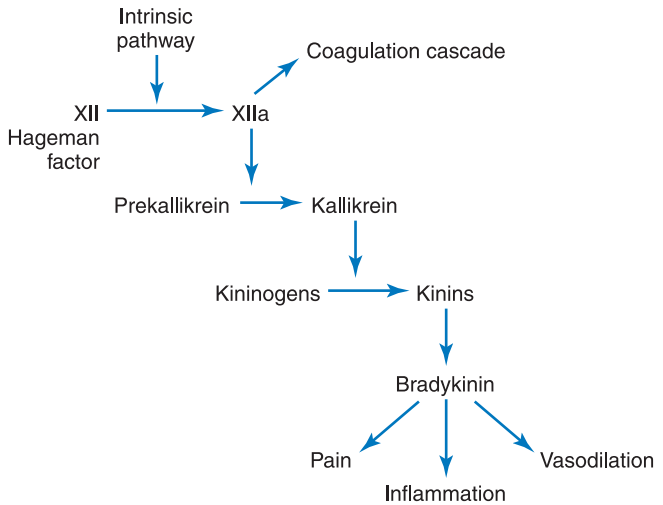


FIG 9.17 Common linkage of the kinin and coagulation systems through the activation of factor XII (Hageman factor). *XIIa,* Activated factor XII.

KEY POINTS

- The primary lymphoid organs are the thymus and bone marrow. T cells develop in the thymus, whereas B cells develop in the bone marrow. Mature lymphocytes then migrate to secondary lymphoid structures, including the spleen and lymph nodes.
- Blood cells are produced in the bone marrow in response to specific hematopoietic growth factors. Granulocytes (neutrophils, basophils, eosinophils) and monocytes (macrophages) are phagocytic cells that provide innate protection. Lymphocytes (B cells, T cells) are specific cells that react only to particular antigens. Natural killer (NK) cells are lymphocytes that lack T-cell and B-cell markers and function in innate immune responses. Other blood components produced by bone marrow are erythrocytes and platelets.
- Neutrophils are the most numerous WBCs in blood. A large storage pool lies in the bone marrow and can be mobilized in response to antigen. Neutrophils are the predominant WBC type mobilized in early infection. They migrate to the area by following chemotactic factors and perform phagocytic functions. During acute bacterial infection, larger numbers of immature neutrophils (bands) are released into the blood, which is termed a "shift to the left." Chronic infections may produce a shift to the right with more segmented neutrophils than normal.
- Monocytes located in tissue are called *macrophages*. Monocytes and macrophages are distributed in strategic locations throughout the body, including the skin, lungs, gastrointestinal tract, liver, spleen, and lymph. Macrophages are powerful phagocytes and are predominant in late inflammation.
- T lymphocytes, the major effectors of cell-mediated immunity, interact with specific antigens on cell surfaces. They are important in immunity against foreign, infected, or mutant cells. In addition, they secrete cytokines that boost the immune response of B cells and other cell types. T cells are composed of two main subtypes called CD4 (helper) and CD8 (cytotoxic). B lymphocytes are the major effectors of antibody-mediated immunity.
- The complement system consists of about 30 plasma proteins that interact in a cascade fashion to produce important mediators of inflammation and immunity. The cascade can be activated by microbial antigens (alternative pathway) or by antigen–antibody complexes (classical pathway).
- Cytokines are peptide factors released by immune cells. They have many functions, including as inflammatory mediators, chemotaxins, intercellular communication signals, growth factors, and growth inhibitors. Macrophages and lymphocytes are important sources of immune cytokines.

INNATE DEFENSES AND INFLAMMATION

Inflammation occurs when cells are injured, regardless of the cause of the injury. It is a protective mechanism that also begins the healing process. The inflammatory response has three purposes: (1) to neutralize and destroy invading and harmful agents, (2) to limit the spread of harmful agents to other tissue, and (3) to prepare any damaged tissue for repair. Inflammatory reactions increase capillary permeability such that phagocytic cells, complement, and antibodies can leave the bloodstream and enter tissues where they are needed.

Five cardinal signs of inflammation have been described: (1) redness (rubor), (2) swelling (tumor), (3) heat (calor), (4) pain (dolor), and

TABLE 9.2 Selected Immune Cytokines and Their Functions

Cytokine	Origin	Function
IFN-α	Macrophages and induced by RNA or DNA viruses and by single- or double-stranded polyribonucleotides	Inhibits virus replication, toxic to cancer cells, stimulates leukocytes, facilitates NK cell activity, produces fever, increases B- and T-cell activity
IFN-β	Fibroblasts and induced by RNA or DNA viruses and by single- or double-stranded polyribonucleotides	Inhibits virus replication, toxic to cancer cells, facilitates NK cell activity, produces fever
IFN-γ	T cells (T_H1 and $CD8^+$) and NK cells	Promotes antigen expression via MHC I and II, activates macrophages, inhibits cell growth, induces myeloid cell lines, promotes B cell switch to IgG
IL-1	Mononuclear phagocyte	Stimulates T_H17 cells and macrophages, induces acute phase reaction of inflammation, produces fever; similar to TNF and endogenous pyrogen
IL-2	T helper cells (T_H1)	Promotes growth of T cells, enhances function of NK cells, assists T-cell maturation in thymus and B-cell proliferation
IL-3	T cells, endothelial cells, fibroblasts, other cells	Induces proliferation and differentiation of other lymphocytes, pluripotent stem cells, mast cells, and granulocytes
IL-4	T helper cells (T_H2)	Promotes T-cell/B-cell interactions, promotes synthesis of IgE by B-cell and T_H2 cell growth, promotes mast cell and hematopoietic cell growth
IL-5	T helper cells (T_H2)	Promotes growth and differentiation of B cells to secrete IgA, induces differentiation of eosinophils
IL-6	Mononuclear phagocytes, T cells and nonlymphoid cells (e.g., endothelium)	Promotes immunoglobulin secretion by B cells, induces fever, promotes release of inflammation factors from liver cells, promotes differentiation of hematopoietic stem cells and nerve cells
IL-7	Stromal cells in bone marrow	Stimulates immature lymphocytes to divide to produce B and T cells
IL-8	Macrophages	Enhances inflammation and chemotaxis (CXCL chemokine)
IL-9	T_H2 cells	Enhances growth of T helper cells, B cells, and mast cells
IL-10	T_{REG} cells and macrophages	Inhibits activation of macrophages and dendritic cells, inhibits IL-12 production
IL-11	Stromal cells in bone marrow	Stimulates platelet production
IL-12	Macrophages, dendritic cells	Enhances T_H1 cell activities and release of IFN-γ by T cells and NK cells
IL-13	T_H2 cells	Stimulates B-cell growth and IgE production, suppresses macrophages
IL-14	T cells	Induces B-cell proliferation
IL-15	Macrophages and various other cell types	Similar actions to IL-2, enhances proliferation of T cells (CD8) and NK cells
IL-16	$CD8^+$ T cells	$CD4^+$ cell chemotaxis, suppresses viral replication of HIV
IL-17	$CD4^+$ T cells	Stimulates production of colony-stimulating factors and chemokines
IL-18	Macrophages in response to microbes	Increases NK-cell proliferation and secretion of IFN-γ by T_H1
IL-19	Macrophages	Stimulates macrophage IL-1 secretion
IL-20	Monocytes	Stimulates hematopoietic stem cells
IL-21	T_H2 and T_H17 and T_{FH} cells	Activates B cells, stimulates production of NK cells
IL-22	T_H17 cells	Epithelial cells, increased barrier function, defensin production
IL-23	Macrophages and dendritic cells	Differentiation of T_H17 cells
IL-24	Monocytes, T cells	Monocyte inflammatory cytokine production
IL-25	T cells, mast cells and macrophages	Stimulates production of cytokines by T_H2 cells (IL-4, IL-5, IL-13)
IL-26	T cells, monocytes	Uncertain
IL-27	Macrophages and dendritic cells	Inhibits T_H1 cells
IL-31	T_H2 cells	Uncertain
IL-33	Endothelial cells, fibroblasts	T_H2 cell development
TNF-α	Macrophages	Induces leukocytosis, fever, weight loss, inflammation, necrosis of some tumors; stimulates lymphokine synthesis; activates macrophages; toxic to viruses and tumor cells
TNF-β	T_{REG} cells	Inhibits B-cell and T-cell proliferation
G-CSF, M-CSF, GM-CSF	Macrophages, T cells, fibroblasts	Stimulates granulocyte and monocyte production in bone marrow
TGF-β	T cells, macrophages	Inhibits T cells, B cells, and macrophages
Chemokines		
CXCL 1–16	Macrophages and various cells in tissues	Recruitment of neutrophils, macrophages, lymphocytes
CCL 1–28	Macrophages and various cells in tissues	Recruitment of neutrophils, macrophages, lymphocytes

GM-CSF, Granulocyte-macrophage colony-stimulating factor; *HIV,* human immunodeficiency virus; *IFN,* interferon; *IL,* interleukin; *NK,* natural killer; *TNF,* tumor necrosis factor; *TGF,* transforming growth factor.

(5) loss of function (functio laesa). The suffix *-itis* is commonly used to describe conditions associated with inflammation. For example, appendicitis, tendonitis, and nephritis refer to inflammation of the appendix, tendon, and kidney, respectively.

Inflammation can be caused by many conditions. Any injury to tissue will evoke an inflammatory response. Injury can arise from sources outside the body (exogenous) or from sources inside the body (endogenous). Surgery, trauma, burns, and skin injury from chemicals are all examples of exogenous injuries. Endogenous injuries may result from tissue ischemia such as myocardial infarction or pulmonary embolism.

Inflammation and infection are commonly confused because they often coexist. Under normal conditions, infection is always accompanied by inflammation; however, not all inflammation involves an infectious agent. For example, inflammation can occur with sprain injuries to joints, myocardial infarction, sterile surgical incisions, thrombophlebitis, and blister formation as a result of either temperature extremes or mechanical trauma.

Inflammation may be categorized as either acute or chronic. Acute inflammation is short in duration, lasting less than 2 weeks, and involves a discrete set of events. Chronic inflammation tends to be more diffuse, extends over a longer period, and may result in the formation of scar tissue and deformity.

INFLAMMATION

Events in the inflammatory response to bacterial antigens and tissue injury include (1) vasodilation and increased vascular permeability, (2) recruitment and emigration of leukocytes, and (3) phagocytosis of antigens and debris. The general inflammatory response is outlined in Fig. 9.18.

Innate immune cells, including neutrophils, macrophages, and some tissue cells, are equipped with pattern recognition receptors that allow them to bind to foreign particles and produce inflammatory cytokines that are released to initiate inflammatory cascades. Pattern recognition receptors, such as mannose receptors and Toll-like receptors, have been known to be present on the cell membrane for some time, but now their presence inside of cells has been described. When these internal receptors detect foreign matter, such as viral RNA or DNA, an internal cascade of events can assemble a group of proteins into a functional unit called an *inflammasome*. Caspase-1 is an important component of the inflammasome that increases the production and release of interleukin-1, an important proinflammatory cytokine that causes fever and contributes to recruitment and activation of immune cells.

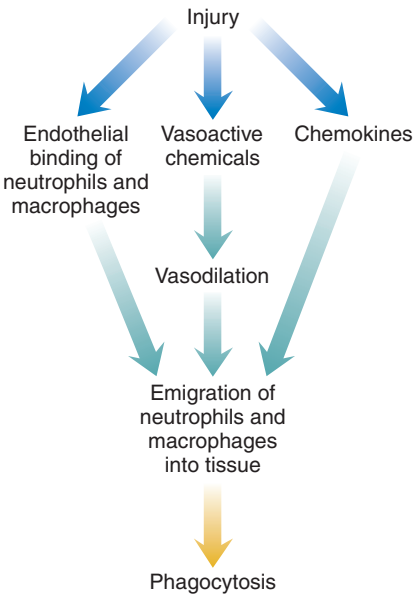


FIG 9.18 Tissue injury stimulates the release of a number of chemical mediators that promote vasodilation, chemotaxis, and binding of neutrophils and macrophages to area capillaries. These events facilitate the emigration of neutrophils and macrophages into the tissue, where they begin phagocytosis.

Increased Vascular Permeability

Prostaglandins and leukotrienes are important mediators of inflammation (Table 9.3). Mast cells are an important source of these inflammatory chemicals. Mast cells in the area of injury degranulate and release packets of histamine and other inflammatory chemicals. One of the early actions of these mediators is to vasodilate and cause endothelial cells to begin retraction and rounding up, thus increasing capillary permeability. The greater volume of blood increases the amount of pressure within the blood vessels (hydrostatic pressure). The increased pressure along with increased permeability pushes fluid out of the blood vessels and into the surrounding tissue, contributing to local swelling. Because of the dilated blood vessels and open capillaries, more blood is carried to the injured area and contributes to the redness, pain, heat, and swelling of inflammation (Fig. 9.19).

Histamine is an early mediator of this inflammatory response. It is such a potent vasodilator that it can cause significant reductions in blood pressure when released in excessive amounts. Histamine also causes bronchial constriction and mucus production. Histamine receptor–blocking agents are widely used in allergic reactions, such as skin reactions and hay fever, to suppress these inflammatory actions of histamine.

Prostaglandins and leukotrienes are phospholipid compounds formed from arachidonic acid. The prostaglandins involved in inflammation contribute to vasodilation and increased permeability (Fig. 9.20). Prostaglandin D_2 also acts as a chemotactic factor and stimulates neutrophil emigration. Prostaglandins cause pain by enhancing the sensitivity of pain receptors. They arise from the cyclooxygenase pathway and can be inhibited by drugs that block enzymes in this pathway, such as aspirin.

Five types of leukotrienes are generated from the lipoxygenase pathway: A_4, B_4, C_4, D_4, and E_4. Leukotriene B_4 is a potent chemotactic agent that causes aggregation of leukocytes; leukotrienes C_4, D_4, and E_4 are inflammatory and cause smooth muscle contraction, bronchospasm, and increased vascular permeability. Leukotriene receptor–blocking agents can be used to inhibit the inflammatory actions of these chemicals.

During the early phase of tissue inflammation, platelets move into the site and adhere to exposed vascular collagen. The platelets release fibronectin to form a meshwork trap and stimulate the intrinsic clotting cascade to help reduce bleeding. Platelets release a number of peptide growth factors, including platelet-derived growth factor and insulin-like growth factor. Platelet-derived growth factor stimulates fibroblast cell proliferation, and insulin-like growth factor type 1 is a potent vascular endothelial cell chemotactic factor. Triggering of the blood coagulation cascade also occurs and leads to the formation of a fibrin clot. Usually, early clot formation occurs within several minutes. Fibrin is also deposited in the lymph system, where it causes lymphatic blockage. Lymphatic blockage "walls off" the area of inflammation from the surrounding tissue and delays the spread of toxins.

The vascular changes that occur soon after injury are beneficial to the injured tissue because irritating or toxic agents are diluted by the fluid that leaks out of the blood vessels into surrounding tissue. In addition, when the fluid leaves the blood vessels, the remaining blood becomes viscous (thick) and circulation is slowed, facilitating neutrophil emigration.

TABLE 9.3 Mediators of Acute Inflammation

Mediator	Vasodilation	Increased Permeability: Immediate	Increased Permeability: Sustained	Chemotaxis	Opsonin	Pain
Histamine	+	+++	–	–	–	–
Serotonin (5-HT)	+	+	–	–	–	–
Bradykinin	+	+	–	–	–	++
Complement 3a	–	+	–	–	–	–
Complement 3b	–	–	–	–	+++	–
Complement 5a	–	+	–	+++	–	–
Prostaglandin (E_2)	+++	+	+?	–	–	–
Leukotrienes (B_4, D_4)	–	+++	+?	+++	–	–
Lysosomal proteases	–	–	++*	–	–	–
Oxygen free radicals	–	–	++*	–	–	–

Data from Roitt I et al: *Immunology,* ed 6, St Louis, 2001, Mosby.
*Proteases and oxygen-based free radicals derived from neutrophils are believed to mediate a sustained increase in permeability by means of their damage to endothelial cells.

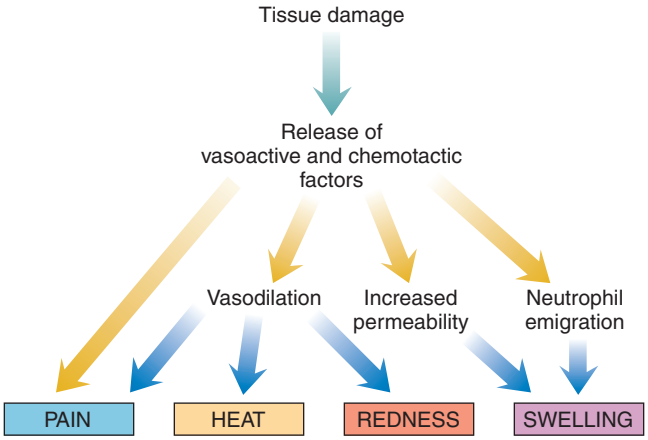


FIG 9.19 Cardinal signs of acute inflammation result mainly from vasodilation and increased vascular permeability.

Emigration of Leukocytes

As blood flows through areas of inflammation, neutrophils move to the sides of the blood vessels and roll along the endothelium of the vessel wall. This process is referred to as *margination* or *pavementing.* Normally, neutrophils slide past the capillary endothelial cells and do not stick. Injured tissue triggers the expression of adhesion molecules on the surface of endothelial cells, and the adhesion molecules bind to receptors on neutrophils (Fig. 9.21). These receptors, called *selectin and chemokine receptors,* help neutrophils stick and roll along the capillary endothelial surface. Binding to and subsequent movement through the capillary wall are accomplished by another group of receptors called *integrins.* Chemokines present on the endothelium enhance the binding affinity of integrins so the neutrophil can attach firmly to the vessel wall. The process of passing through the blood vessel walls and migrating to the inflamed tissue is referred to as *emigration* or *diapedesis.* Diapedesis begins within a few minutes to hours of injury. Even though the spaces between endothelial cells lining the vessels are much smaller than the neutrophils, neutrophils are able to slide through a small portion at a time.

Neutrophils are attracted to the inflamed tissue by a process called *chemotaxis.* Biochemical mediators that attract neutrophils include bacterial toxins, degenerative products of the inflamed tissue, the C5a complement fragment, and other substances. Neutrophils are thus guided through the tissue to an area of injury by these chemicals. Because neutrophils are highly mobile, they are first on the scene to begin phagocytosis and production of collagenase to degrade dead tissue. Monocytes are slightly slower to arrive at an area of inflammation but use a similar process of emigration to gain entry to the area of tissue injury.

Eosinophils and NK cells may also respond to the site of inflammation. Eosinophils are rich in chemical mediators such as hydrolases and peroxidases, which may contribute to the inflammatory process. NK cells are most effective in recognizing virally infected cells and opsonized microbes.

Phagocytosis

Once neutrophils and monocytes (macrophages) enter the tissue, they begin the process of phagocytosis (Fig. 9.22). These cells produce a wide variety of enzymes that digest protein structures. Some of these enzymes include lysozyme, neutral proteases, collagenase, elastase, and acid hydrolases. Neutrophils and macrophages specialize in collagen and extracellular matrix degradation. Peptide bonds are cleaved in the extracellular matrix by collagenase, elastase, proteinase, and gelatinase. If the microbe is small enough to be internalized, it will be captured by the phagocyte and endocytosed into a phagosome. The phagosome then merges with a lysosome containing degradative enzymes. Large antigens may trigger the neutrophil to release its degradative enzymes extracellularly, causing damage to local tissues.

Oxidizing agents, the most destructive of the inflammatory cell products, are formed as a result of the phagocyte oxidase enzyme system on the membrane of the lysosome. Neutrophils are capable of synthesizing and assailing microorganisms with these oxidizing agents, which include the following oxygen radicals: superoxide (O_2^-), hydrogen peroxide (H_2O_2), and hydroxyl ions (OH^-). Oxidizing agents directly attack cell membranes and thereby increase permeability. Nitric oxide products may also be produced by inducible nitric oxide synthase (iNOS) and function in concert with oxygen radicals to attack microbial molecules (see Fig. 9.22).

Because acute inflammation can cause severe tissue damage, it is not surprising that a system of inactivators is present. An important inhibitor of inflammatory damage is α_1-antiprotease. Antiproteases are made in the liver and circulate continuously in the bloodstream. α_1-Antiprotease inhibits the destructive proteases released from activated neutrophils. A deficiency of antiproteases can predispose an individual to inflammatory tissue destruction.

Neutrophils have a limited capacity to phagocytose foreign and inflammatory debris. Once the neutrophil leaves the circulation to fight

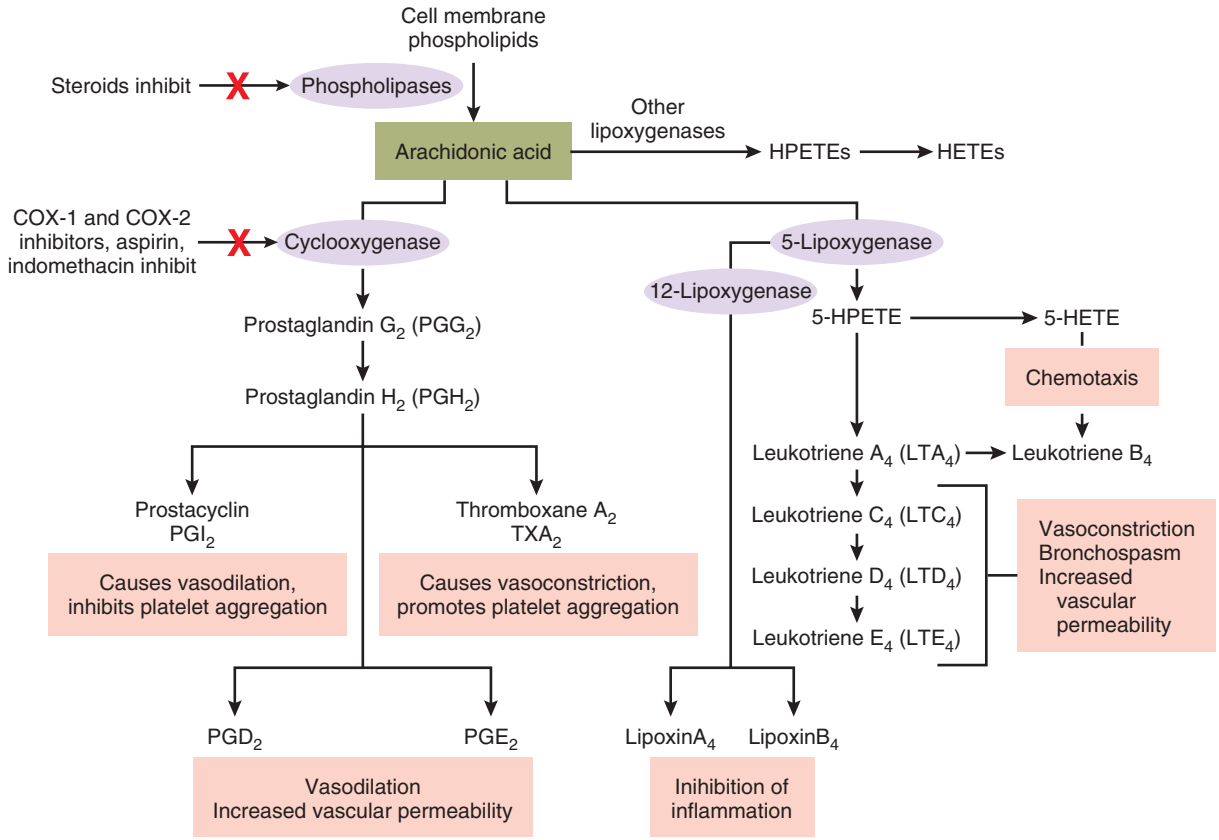


FIG 9.20 Generation of prostaglandins, thromboxane, and leukotrienes from arachidonic acid and roles in inflammation. *HETEs,* Hydroeicosatetraenoic acids; *HPETEs,* hydroperoxyeicosatetraenoic acids.

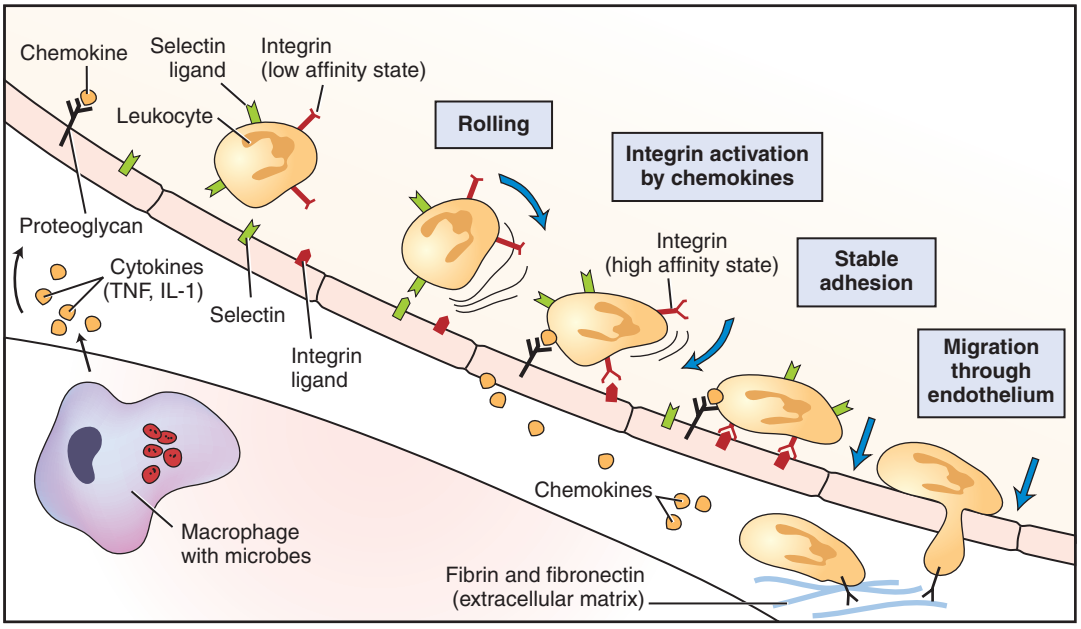


FIG 9.21 Emigration of neutrophils from the bloodstream into tissue is mediated by receptor interactions with the capillary endothelium. With inflammation and injury, endothelial cells begin to express binding molecules on their cell surfaces (selectins). Leukocytes also have selectins, which can bind to endothelial adhesion proteins. The selectin interactions cause the leukocytes to stick and roll. Chemokines on the surface of endothelial cells interact with neutrophils (and macrophages) to increase the binding affinity of integrin receptors on leukocytes. Firm attachment and diapedesis through the capillary wall is facilitated by integrins that allow the neutrophils to bind to endothelial cells and extracellular matrix and then pull themselves into the tissue. *IL,* Interleukin; *TNF,* tumor necrosis factor.

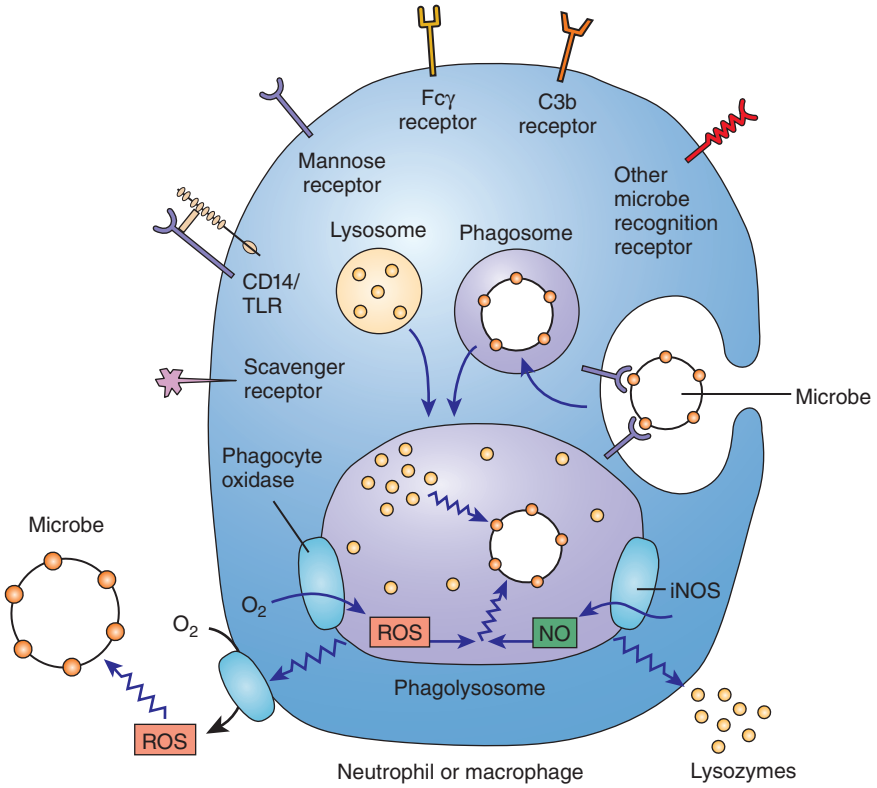


FIG 9.22 Neutrophils and macrophages have a number of different receptors on their surface that enable them to bind to components of microbes or to opsonins like IgG and complement. Bound microbes are internalized into phagosomes that fuse with lysosomes containing numerous enzymes. Some of these enzymes degrade proteins (proteolytic), and others such as oxidase and inducible nitric oxide synthase (iNOS) produce free radicals that attack molecular bonds. When phagocytes are strongly stimulated or microbes are too large to internalize, the lysosomal enzymes may be activated or released at the cell surface, causing tissue damage and inflammation. *ROS,* Reactive oxygen species.

an infection, it is unable to return and will die at the site. When phagocytosis is incomplete, a collection of dead neutrophils, bacteria, and cellular debris, called *pus,* may form at the site. Macrophages are left with the job of removing spent neutrophils and preparing the site for healing. A predominance of monocytes and macrophages in an inflamed area signals the beginning of chronic inflammation.

Chronic Inflammation

Macrophages are essential for wound healing because of their phagocytic and debridement functions. Macrophages produce proteases that help in removing foreign protein from the wound. Macrophages also release tissue thromboplastin to facilitate hemostasis and stimulate fibroblast activity. Macrophages secrete other peptide growth factors such as angiogenic factor, which encourages the growth of new blood vessels. Macrophages also phagocytose spent neutrophils and their degradation products so they do not interfere with healing. Prolonged inflammation may impair healing and result in an accumulation of macrophages, fibroblasts, and collagen, called a *granuloma.* Granulomas are usually evident on examination of tissue biopsy as clusters of macrophages surrounding particulate matter or resistant microbes such as *Mycobacterium tuberculosis.* Fibrosis and scarring are evident because normal parenchyma is replaced with fibrous tissue.

HEALING

Healing of tissues after inflammation can occur in different ways over time. Usually the reconstructive phase begins 3 to 4 days after injury and persists for 2 weeks. The major cells involved in this phase include fibroblasts, endothelial cells, and myofibroblasts.

Fibroblasts are found all over the body and are thought to originate in mesenchymal primitive tissue. They synthesize connective tissue and are able to migrate. Fibroblasts are stimulated to make collagen, proteoglycans, and fibronectin by a variety of growth factors. Macrophages secrete lactate and release growth factors that stimulate fibroblasts. Fibroblasts respond to contact and density inhibition and thereby facilitate orderly cellular growth. Myofibroblasts develop at the wound edge and induce wound contraction.

Endothelial cells grow into the connective tissue gel stimulated by angiogenic substances. They usually develop capillary beds from existing vessels. The new capillaries can transport nutrients for tissue repair and wound healing. However, because the new capillaries are leaky, they contribute to continuing edema.

Regeneration of damaged tissue into the preexisting tissue type requires survival of the basement membrane and tissue stem cells. Some cell types regenerate constantly; among these types are the epithelial cells of the skin and mucous membranes, bone marrow cells, and lymphoid cells. Cells of the liver, pancreas, endocrine glands, and renal tubules are also able to regenerate when necessary. However, some cell types, such as neurons and muscle cells, regenerate poorly. The maturation phase of wound healing occurs several weeks after the injury and may last for 2 years or more. It is characterized by wound remodeling by fibroblasts, macrophages, neutrophils, and eosinophils. Wound remodeling is the process of collagen deposition and lysis with debridement of the wound edges. During this phase the wound changes color from

bright red to pink to whitish. As long as a wound is pink, the maturation phase is not completed.

INFLAMMATORY EXUDATES

Exudate is fluid that leaks out of blood vessels, combined with neutrophils and the debris from phagocytosis. Exudates may vary in composition, but all types have similar functions, including (1) transport of leukocytes and antibodies, (2) dilution of toxins and irritating substances, and (3) transport of the nutrients necessary for tissue repair.

Serous exudate is watery, has a low protein content, and is similar to the fluid that collects under a blister. This type of exudate generally accompanies mild inflammation. With mild inflammation, the permeability of the blood vessels is not greatly changed. As a result, only some protein molecules escape from vessels, and serous exudate, with a low protein content, develops. Small amounts of red blood cells may leak into the serous fluid with capillary injury, resulting in a pink-tinged drainage called *serosanguineous* drainage.

With greater injury, more inflammation occurs and the blood vessels become more permeable. Because of this increased permeability, more protein can pass through the vessel walls. Fibrinogen, a large protein molecule, can pass through a highly permeable blood vessel wall. *Fibrinous exudate* is sticky and thick and may have to be removed to allow healing; otherwise, scar tissue and adhesions may develop. However, in some instances fibrinous exudate may be beneficial. In the case of acute appendicitis, fibrinous exudate may actually wall off and localize the infection and prevent its spread.

Purulent exudate is called *pus.* Purulent exudate generally occurs in severe inflammation accompanied by bacterial infection and is primarily composed of neutrophils, protein, and tissue debris. Large pockets of purulent exudate, called *abscesses,* must generally be removed or drained for healing to take place.

Hemorrhagic exudate has a large component of red blood cells. This type of exudate is usually present with the most severe inflammation. Hemorrhagic exudate occurs with severe leakage from blood vessels or after necrosis or breakdown of blood vessels.

SYSTEMIC MANIFESTATIONS OF INFLAMMATION

Inflammation is associated with both localized and systemic signs and symptoms. The localized symptoms, described previously, occur with both acute and chronic inflammation. Depending on the magnitude of injury and the resistance of the individual, localized inflammation can lead to systemic involvement. Systemic responses include fever, neutrophilia (increased blood neutrophil count), lethargy, and muscle catabolism. Three macrophage-derived cytokines—IL-1, IL-6, and TNF-α—are responsible for most of the systemic effects of inflammation.

TNF-α and IL-1 act on the brain to raise body temperature, induce sleep, and suppress appetite. By raising the set point for body temperature, these cytokines induce conservation of heat through vasoconstriction, as well as increased heat production through shivering. An increase in body temperature is assumed to improve the immune response; however, the mechanism is unclear. IL-1 is responsible for stimulating the release of neutrophils from bone marrow storage sites, thus producing neutrophilia. All three cytokines act on skeletal muscle to enhance protein catabolism, which provides an available pool of amino acids for efficient antibody production by plasma cells.

The liver is another important target for IL-1 and TNF-α. These cytokines induce the liver to release a number of proteins collectively called *acute phase proteins,* which include complement components, clotting factors, and protease inhibitors (Fig. 9.23). Two of the most important acute phase proteins are C-reactive protein (CRP) and serum amyloid A. CRP binds to phospholipids on bacterial cell membranes and acts as an opsonin to facilitate phagocytosis.

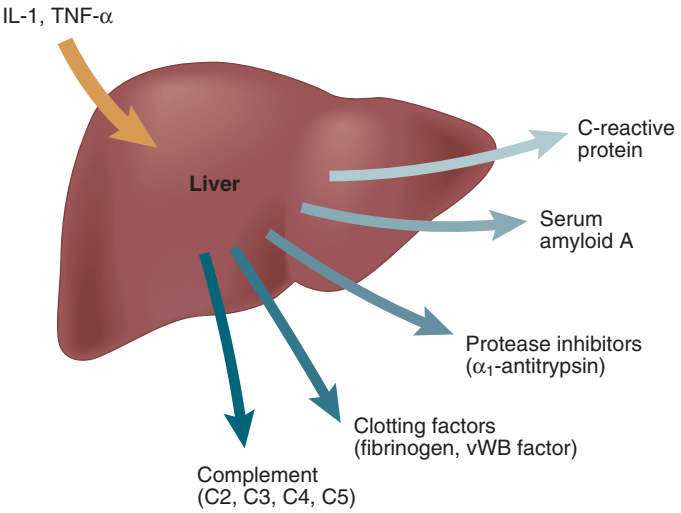


FIG 9.23 The liver is a target for three important cytokines: interleukin-1 (IL-1) and tumor necrosis factor-α (TNF-α). In response to these cytokines, the liver releases a number of proteins, collectively called acute phase proteins. *vWB,* von Willebrand factor.

When the liver releases acute phase proteins, the level of fibrinogen in the serum is increased. Fibrinogen coats the surface of red blood cells and reduces their charge so that they aggregate more readily. A blood test called the *erythrocyte sedimentation rate* (ESR, "sed rate") provides a simple measure of the level of inflammation in an individual. Thus an elevated ESR indicates the presence of inflammation in the body. The greater the inflammation, the faster the red blood cells precipitate to the bottom of a test tube and the higher the ESR. The ESR is a nonspecific but clinically useful indicator of inflammation. Serum CRP activity is also used as a nonspecific indicator of inflammation in a manner similar to the ESR.

KEY POINTS

- Previous exposure to foreign antigens is not required for the activation of innate immune defenses. Inflammation is an important aspect of innate immunity that involves localization of harmful agents and the movement of phagocytic cells to the area. Classic manifestations of inflammation are redness, swelling, heat, pain, and loss of function.
- Inflammatory chemicals such as histamine, prostaglandins, and leukotrienes are released from injured tissues, mast cells, macrophages, and neutrophils. These chemicals increase vascular permeability, vasodilate, and attract immune cells to the area (chemotaxis).
- Phagocytes migrate to the inflamed area, collect at the side of the vessel, and squeeze through into the tissue. Emigration of neutrophils and macrophages is facilitated by chemokines, selectins, and integrins present on the surface of endothelial cells and leukocytes. Neutrophils arrive in large numbers in acute bacterial infection and begin active phagocytosis. Neutrophils and macrophages produce proteolytic enzymes and oxidizing agents to destroy and digest antigens. With chronic inflammation, macrophages and lymphocytes predominate.
- Healing is mediated by growth factors released from platelets and immune cells that stimulate fibroblasts to divide and manufacture extracellular matrix proteins. Endothelial cells respond to angiogenic growth factors by forming capillary networks.
- Inflammatory exudate functions to transport immune cells, antibodies, and nutrients to the tissue and dilute the offending substances. Serous exudate

is watery and low in protein; fibrinous exudate is thick, sticky, and high in protein; purulent exudate contains infective organisms, leukocytes, and cellular debris; and hemorrhagic exudate contains red blood cells.

- Systemic manifestations of inflammation include fever, neutrophilia, lethargy, muscle catabolism, increased acute phase proteins (CRP), and increased ESR. These responses are attributable to the IL-1, IL-6, and TNF-α released from macrophages and inflamed tissues.

SPECIFIC ADAPTIVE IMMUNITY

The specific immune system uses remarkably effective and adaptive defense mechanisms capable of recognizing foreign invaders, destroying them, and retaining a memory of the encounter such that an even more effective defense (adaptive) will be achieved after subsequent exposure. As previously described, B and T lymphocytes are the cellular mediators of specific adaptive immunity. B cells are said to provide "humoral" immunity because the antibodies they produce are found in body fluids, or "humors." T cells provide "cell-mediated" immunity because they recognize antigen presented on the surface of cells. To achieve immunity against specific antigens, B and T lymphocytes must be capable of recognizing an enormous range of foreign antigen yet not be reactive to self tissues.

Differentiation between self and nonself requires a complex lymphocyte development process in which self-reactive lymphocytes are destroyed and potentially useful lymphocytes are preserved. The MHC proteins have a primary role in enabling lymphocytes to react to foreign antigen while remaining tolerant to self antigen. Self-tolerance is not always effectively maintained, and impairment in self-tolerance can result in the development of autoimmune disorders (see Chapter 10).

MAJOR HISTOCOMPATIBILITY COMPLEX

A cluster of genes on chromosome 6 is known as the major histocompatibility complex (MHC). In humans, the MHC is also known as the human leukocyte antigen (HLA) complex. The proteins made by these genes are displayed on the surface of body cells and mark them as "self." The MHC contains three classes of genes: I, II, and III (Fig. 9.24). Class I and II genes code for proteins that display, or "present," antigens on the surface of cells. Antigen presentation is a vital first step in the initiation of an immune response. T lymphocytes cannot recognize foreign antigens unless they are displayed on MHC proteins on the surface of a cell. Class III genes code for a variety of proteins, many of which are important to inflammatory reactions, including several complement proteins.

A great deal of polymorphism is found in the MHC class I and II genes, which means that it is highly unlikely that one individual will have exactly the same MHC genotype as another individual. For example, three gene loci for MHC class I proteins (A, B, C) are located on each chromosome 6, and an individual inherits one chromosome from each parent for a total of six MHC class I genes. Each of these genes has many different forms (alleles) such that each of the six is likely to be different (Fig. 9.25). Related individuals will generally be more similar but not identical (unless identical twins). The "matching" of MHC gene expression is an important consideration for tissue and organ transplantation. The closer the match is, the less likely that the host will reject the transplant. An individual also receives six MHC class II genes that are expressed on specialized antigen-presenting cells, such as dendritic cells, macrophages, and B cells. Because of the potential for mixing and matching of class II MHC gene products, an individual may express 10 to 20 different MHC class II proteins. The MHC class I and II proteins on the surface of cells display both self and foreign antigens

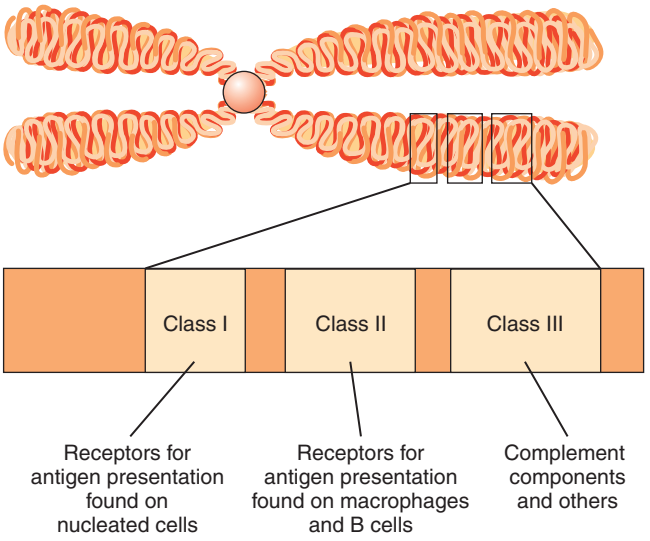


FIG 9.24 Major histocompatibility complex genes are categorized into three main groups known as class I, II, and III. Class I and II genes code for antigen-presenting proteins, whereas class III genes code for a heterogeneous group of proteins, many of which serve immune functions.

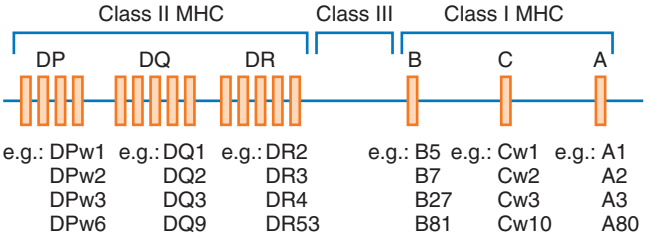


FIG 9.25 Each individual receives six class I major histocompatibility complex (MHC) genes, including pairs of A, B, and C genes. One member of the pair is inherited from each parent. MHC class I genes are expressed in all nucleated cells of the body. Each individual also receives six class II MHC genes, three from each parent. However, class II proteins are composed of two polypeptide chains such that an individual may have 10 to 20 different MHC class II protein molecules. Class II MHC proteins are expressed on the surface of specialized antigen-presenting cells like macrophages, dendritic cells, and B cells. The structure of an individual's MHC proteins is assessed to determine the "tissue type" when matching for tissue transplantation procedures.

for inspection by T cells. Cells displaying foreign antigens stimulate an immune response, whereas those displaying self antigens do not because self-reacting T cells are not allowed to survive in the thymus. Genetic diversity in MHC gene expression is believed to be important to the preservation of a species because new pathogens are likely to encounter at least some individuals with MHC genotypes that can recognize and eliminate these pathogens.

ANTIGEN PRESENTATION BY MHC

Nucleated cells in the body are capable of expressing MHC class I proteins on their cell surfaces, whereas only certain specialized cells, primarily dendritic cells, macrophages, and B cells, are able to express MHC class II proteins. Cytotoxic T cells are able to recognize foreign antigen bound to MHC class I proteins, whereas T helper cells recognize foreign antigen bound to MHC class II proteins. T cells are screened during development in the thymus so that they recognize self MHC

proteins and do not react to self peptides displayed by self MHC proteins. This concept is explored further in the section "Mechanisms of Cell-Mediated Immunity." The sources of antigen, mechanism of antigen processing, and T-cell response to antigen are quite different for MHC I and MHC II reactions.

MHC Class I Presentation

Nucleated cells continuously produce MHC class I proteins on the rough endoplasmic reticulum (ER), where they are combined with various peptide fragments that are present in the cytoplasm. These peptides result from degradation of normal intracellular proteins. The MHC I–peptide complexes are cycled to the cell surface for inspection by T cells. Normal MHCs displaying normal cellular proteins are ignored by T cells. If abnormal proteins are produced in the cell, then the MHC I–peptide complex will be recognized as foreign and an immune response will occur. The peptide antigens presented on MHC I are of intracellular origin. Because viruses are able to gain access to cells directly, viral protein is a common source of foreign MHC class I antigens. Abnormal intracellular proteins produced by mutant cells may also be presented on MHC I, thus targeting them for immune destruction. Before intracellular proteins can be presented at the cell surface, they must be processed and transported to the ER, where they are combined with newly synthesized MHC class I protein (Fig. 9.26). Peptide fragments are generated in the cytoplasmic proteasomes and escorted through the ER by special transporters called *transporters associated with antigen processing* (TAPs). The TAPs are located near the MHC I complexes on the ER membrane and target the peptides to the MHC I–binding cleft. The MHC I–binding cleft can accommodate peptide fragments of 8 to 11 amino acids (Fig. 9.27). The MHC I–antigen complexes then travel to the cell membrane, where they are displayed. Recognition of foreign antigen in association with the MHC I protein on the cell surface targets the presenting cell for destruction by cytotoxic T cells. When the cytotoxic T cell binds to the MHC I–antigen complex (and coreceptors), it is stimulated to release enzymes and pore-forming proteins (perforins) that lyse the target cell. Cytotoxic T cells can only recognize an antigen if it is physically bound to an MHC class I molecule. Cytotoxic T cells are thus said to be MHC class I restricted.

MHC Class II Presentation

MHC class II proteins are used to present antigens obtained from extracellular sources. Extracellular antigens must first be engulfed by the antigen-presenting cell. Cells of the monocyte–macrophage lineage, dendritic cells, and B cells are responsible for presenting antigen by MHC II. Macrophages and dendritic cells obtain foreign antigens by phagocytosis and are thus able to process and present a large number of different antigens. They are said to be "nonspecific" for this reason. B cells, on the other hand, are particular about the antigens that they engulf. The antigen must specifically bind to the BCR to be ingested by a B cell. Each B cell has only one type of BCR and therefore processes and presents only one specific antigen. The specificity of the BCR corresponds to the antibody that the activated B cell will produce. The process of B-cell activation is explored in the section titled "Mechanisms of Humoral Immunity."

After the antigen-presenting cell has ingested an antigen, it is degraded into fragments within the cellular phagosomes (endocytic vesicle). MHC II proteins are synthesized on the rough ER and pick up an antigen

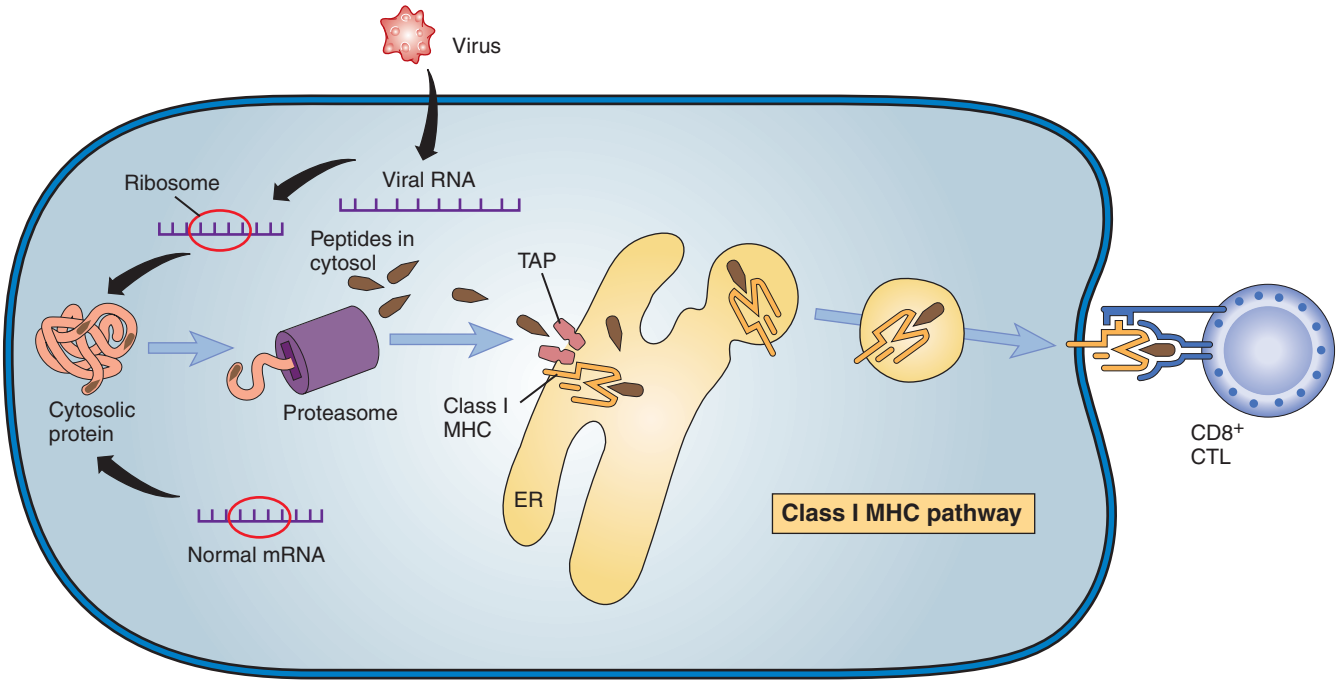


FIG 9.26 Nearly all nucleated cells of the body are able to process and display antigen in association with major histocompatibility complex (MHC) class I protein. The antigens come from the intracellular compartment, and a common source of foreign antigen is viral infection. The viral proteins made within the cell's cytoplasm are processed into peptide fragments in the proteasome and then enter the endoplasmic reticulum (ER) through TAP transporters. There they combine with MHC class I proteins. The MHC class I–antigen complex then shuttles to the cell surface within a vesicle. When the vesicle combines with the plasma membrane, the MHC class I–antigen complex is displayed on the cell surface. *CTL,* Cytotoxic T lymphocyte; *TAP,* transporter associated with antigen processing. (Redrawn from Abbas AK et al: *Cellular and molecular immunology,* ed 8, Philadelphia, 2015, Saunders, p 124.)

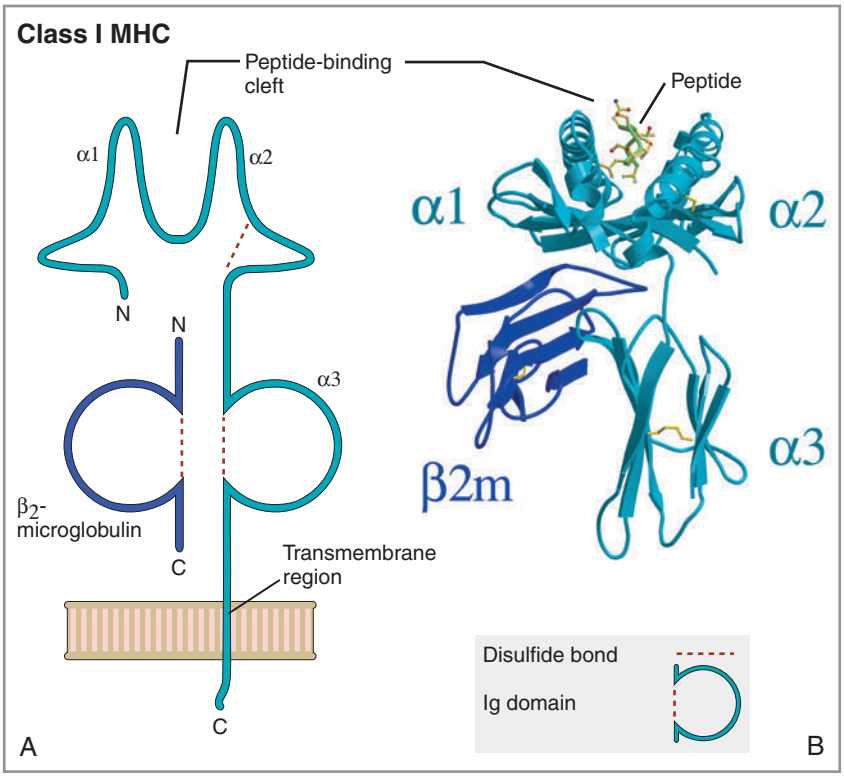


FIG 9.27 Schematic **(A)** and ribbon **(B)** diagrams of the class I major histocompatibility complex molecule. Note that the peptide-binding cleft is formed from one polypeptide chain that restricts the size of peptide in the pocket to 8 to 11 amino acids. (From Abbas AK et al: *Cellular and molecular immunology,* ed 8, Philadelphia, 2015, Saunders, p 120.)

from the phagosome on their way to the plasma membrane (Fig. 9.28). The class II MHC molecule is formed by two protein chains, and the binding cleft is more flexible than that of MHC I proteins (Fig. 9.29). Peptides displayed by MHC class II proteins range in size from 10 to 30 amino acids. The MHC II–antigen complexes are then displayed at the cell surface where T helper cells can detect them. T helper cells can only recognize a foreign antigen if it is physically bound to an MHC II protein. T helper cells are thus said to be MHC II restricted. Naive T cells located in lymph nodes are usually presented with antigen by dendritic cells displaying foreign antigens on their MHC II proteins. Dendritic cells populate the body surfaces and mucous membranes. When they engulf antigen, they break their tissue attachments and migrate to lymph nodes where they interact with T helper cells.

MECHANISMS OF CELL-MEDIATED IMMUNITY

T cells are able to recognize foreign antigen displayed on the surface of antigen-presenting cells through specialized receptors called *T cell receptors* (TCRs). Each T cell has tens of thousands of identical TCRs on its cell surface. Each T cell is thus able to recognize and respond to only a single antigenic epitope. This property is what makes T cells specific. The binding specificity of the TCR is determined by recombination, rearrangement, and random mutations within the genes that code for the TCR-binding domain. Billions of different TCR amino acid sequences are possible, thus providing a tremendous diversity of potential antigen-binding specificities. This diversity increases the likelihood that one or more T cells will have the right TCRs to allow recognition of any of the various pathogens that may gain access to the body. The drawback to this random approach is that many TCRs will be useless or may bind self antigens. A rigorous selection process occurs in the thymus such that self-reactive T cells are eliminated. This selection process requires at least two steps. In the first, T cells must demonstrate an ability to recognize self MHC proteins displayed on the surface of specialized thymic cells. Portions of the TCR must make appropriate contact with the MHC protein, or the T cell will not be able to respond to antigens presented on the cell surface. Expression of either CD4 or CD8 on the T cell helps determine which class of MHC the T cell must fit. T cells that do not have functional TCRs undergo apoptosis in the thymus. The second requirement is that the TCR does not bind tightly to MHC proteins that are displaying normal self-derived peptides. Tight binding to self peptides also triggers the cell to initiate apoptosis. T cells that pass these tests migrate to secondary lymphoid tissues to await foreign antigens. Exposure of a T cell to its corresponding antigen results in expansion of the T cell into a clone of cells that all recognize the same antigen. This process ensures that useful T cells are maintained in the body as memory cells, whereas T cells that do not encounter antigen will not proliferate. Members of the T-cell clone migrate to lymphoid organs throughout the body, where they can respond rapidly should the same antigen reenter the system. The life span of mature T cells is long, but the numbers of memory cells in a clone will decline over time. However, intermittent exposure to the antigen is likely to occur and will stimulate proliferation and maintain immunity.

The two major types of T cells, T helper cells and cytotoxic T cells, react quite differently to activation of their TCRs by antigen and are therefore described separately in the following sections.

T Helper Cells ($CD4^+$)

T helper cells recognize antigen in association with MHC class II molecules. The CD4 protein is needed to enable T helper cells to bind the MHC II protein, whereas the TCR recognizes the specific antigen

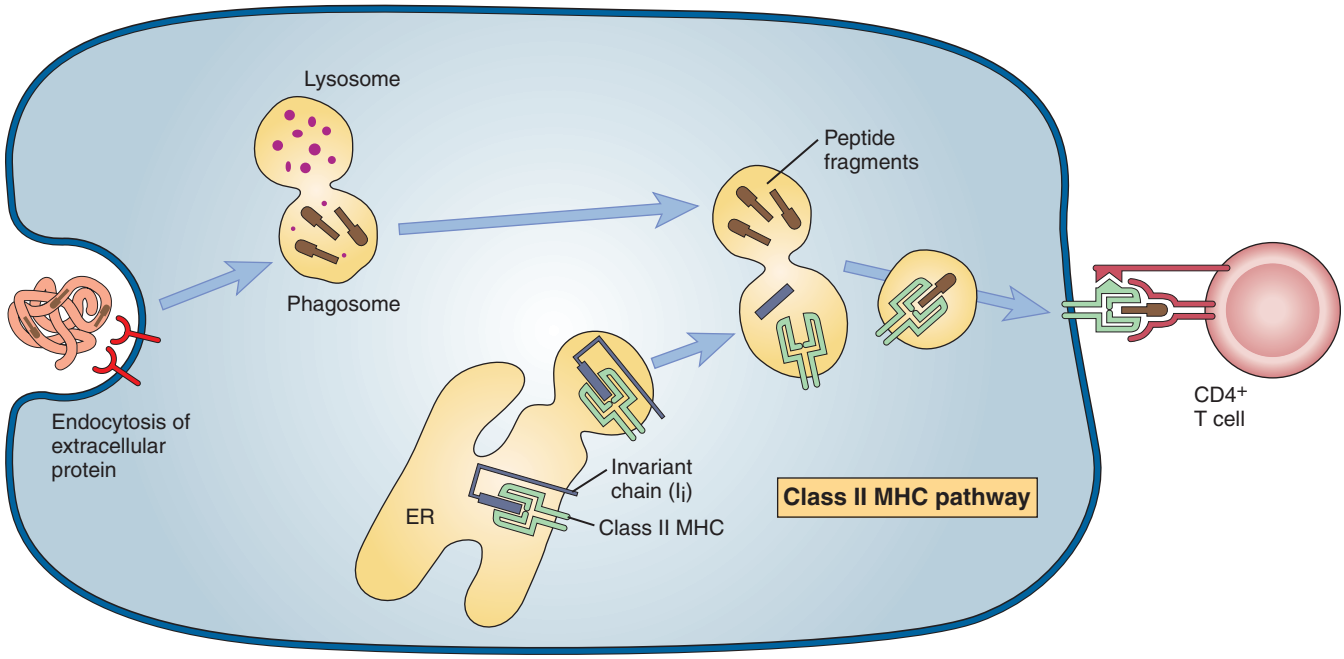


FIG 9.28 Only specialized cells are able to obtain extracellular antigen for processing and presentation in association with major histocompatibility complex (MHC) class II protein. These cells are primarily dendritic cells, macrophages, and B cells. The antigen is first engulfed into a vesicle called a phagosome, which fuses with a lysosome. Enzymes within the phagosome break the protein into pieces. MHC II molecules are synthesized on the endoplasmic reticulum (ER) and then transported to the phagosome in a vesicle. The binding cleft of the MHC II protein is complexed with a blocking protein to prevent it from retrieving peptide before it reaches the phagosome. The phagosome and vesicle fuse, and the MHC II loses its blocking protein and picks up an antigen peptide. The complex then migrates to the cell surface and combines with the cell membrane. The MHC II–antigen complex is then displayed on the cell surface. (Redrawn from Abbas AK et al: *Cellular and molecular immunology,* ed 8, Philadelphia, 2015, Saunders, p 124.)

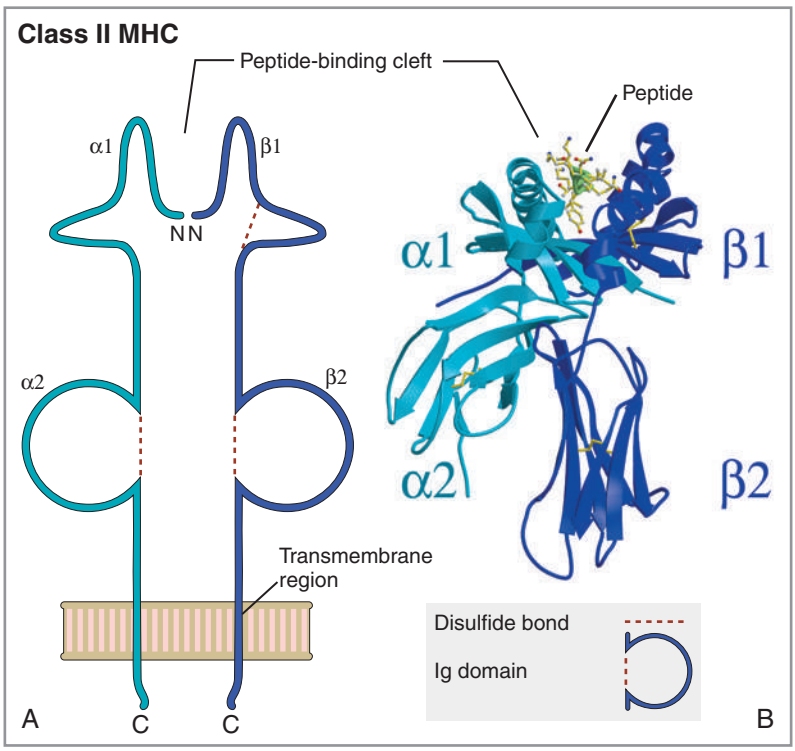


FIG 9.29 Schematic **(A)** and ribbon **(B)** diagrams of the class II major histocompatibility complex (MHC) molecule. Note that the peptide-binding cleft is formed from two separate polypeptide chains, which allows the size of peptide in the pocket to be 10 to 30 amino acids. (From Abbas AK et al: *Cellular and molecular immunology,* ed 8, Philadelphia, 2015, Saunders, p 121.)

being presented (Fig. 9.30). Binding of the TCR to its corresponding antigen generates a signaling cascade in the cytoplasm of the T helper cell. The TCR is linked to this signaling cascade through another protein called *CD3*. Stimulation of CD3 results in the activation of enzymes (kinases) in the cytoplasm that mediate the production of two second messengers: inositol trisphosphate (IP_3) and diacylglycerol (DAG) (see Chapter 3). Inositol trisphosphate initiates a rise in the concentration of intracellular calcium ions, which also act as second messengers to change cell behavior. Other protein kinases turn on the genes for cytokines (e.g., IL-2, IFN-γ, and others), IL-2 receptors, and other cell surface proteins. As previously mentioned, the subtypes of T helper cells (i.e., T_H1, T_H2, T_H17, T_{FH}, and T_{REG}) secrete somewhat different amounts and types of cytokines. These cytokines provide the "help" that T helper cells give to other cells of the immune system. For example, IL-2 activates helper and cytotoxic T cells, NK cells, and macrophages; and IFN-γ is a potent activator of macrophages. IL-2 and IFN-γ are the main cytokines secreted by T_H1 cells. The cytokines secreted by T_H2 cells have stimulatory effects on B cells (e.g., IL-4, IL-5, IL-13). In addition, when a B cell is serving as the antigen-presenting cell, T helper cells provide specific B-cell help through direct cell-to-cell contact by receptor proteins. The T_H17 lymphocytes secrete IL-17, a cytokine that stimulates neutrophil recruitment and prompts fibroblasts and other cells to secrete more proinflammatory signals. T_{FH} lymphocytes reside in the lymph node follicles where they produce cytokines that stimulate B cells. T_{REG}

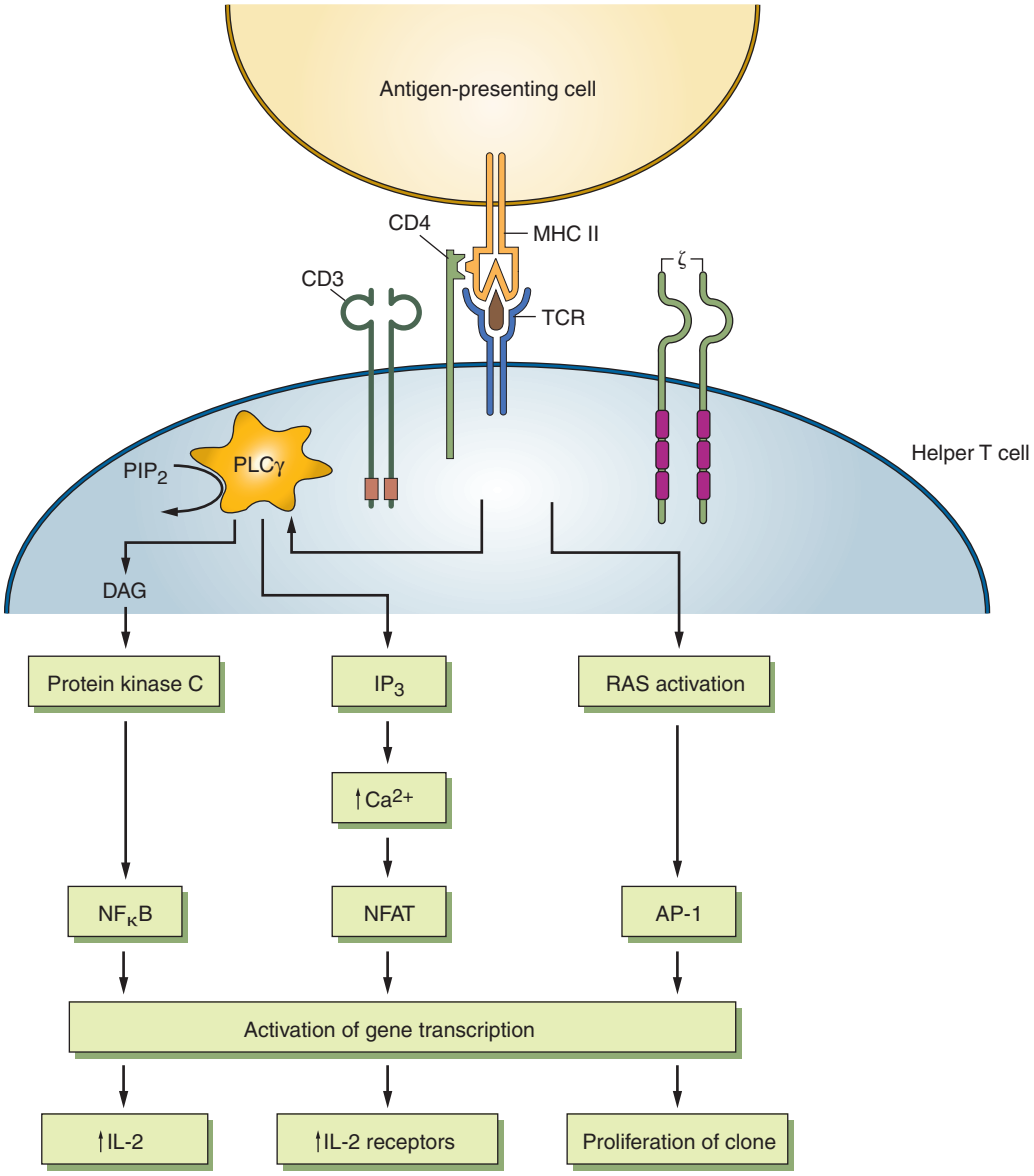


FIG 9.30 T helper cells can recognize and bind antigen in association with major histocompatibility complex (MHC) class II molecules. The T-cell receptor (TCR) on the T helper cell binds to the antigen, and the CD4 protein recognizes the MHC class II protein. Binding is highly specific because the TCR must match the antigen fragment precisely. Once binding is achieved, CD3 and ζ proteins associated with the TCR are activated to initiate intracellular enzyme cascades. Major signaling pathways in activated T cells are shown. These ultimately result in activity of transcription factors and changes in gene activity. *AP-1,* Activation protein-1; *DAG,* diacylglycerol; *IP3,* inositol 1,4,5-trisphosphate; *NFAT,* nuclear factor of activated T cells; *NFκB,* nuclear factor kappa B; *PIP2,* phosphatidylinositol 4,5-bisphosphate; *PLCγ,* phospholipase C-γ.

lymphocytes secrete inhibitory cytokines such as IL-10 and TGF-β to help control inflammation and immune reactions to prevent excessive tissue damage (Fig. 9.31).

Cytotoxic T Cells ($CD8^+$)

Cytotoxic T cells recognize antigen displayed in association with MHC class I protein. The CD8 protein is needed to facilitate binding to the MHC I, whereas the TCR specifically recognizes the presented antigen (Fig. 9.32). Binding of the TCR to its corresponding antigen triggers a number of responses in the cytotoxic T cell. This process is similar to that described for T helper cells and involves signal transduction through CD3 proteins. Antigen binding by cytotoxic T cells is not sufficient to activate them. Cytotoxic T cells also require costimulation by IL-2 cytokines. IL-2 is secreted primarily by activated T helper cells (T_H1). Thus cytotoxic T cells require cytokine "help" before they proliferate effectively. Cytokines are generally not enough to induce significant proliferation of target cells unless other coactivators are also presented by target cells (see Fig. 9.32). Once activated, cytotoxic T cells proliferate into memory cells as well as effector cells. Effector cells accomplish their cytotoxic functions in two ways: through perforins and through CD95.

Perforins are proteins manufactured in the cytotoxic T cell and stored in granules (vesicles) within the cytoplasm. A number of proteolytic enzymes (granzymes) are located in the granules along with the perforins. Binding to the target cell causes the granules to migrate to the contact site, where they are released onto the target cell membrane. The perforins assemble into pores, which then allow the granzymes to move into the target cell. Granzymes degrade DNA and trigger target cell death (apoptosis).

Perforins function in a similar manner to the complement membrane attack complex previously described. It is not entirely clear how the cytotoxic T cell manages to escape injury in this process. Presumably, the perforins and granzymes are focused on the target cell in some controlled manner.

The CD95 protein on cytotoxic T cells is called the *CD95 ligand* (CD95L) or the *Fas ligand (FasL)*. It can bind specifically to complementary CD95 proteins (Fas) found on the surface of target cells. Normal, healthy cells do not express CD95 and are not recognized by cytotoxic cells. Binding of the CD95L to CD95 triggers programmed cell death (apoptosis) of the target cell (see Fig. 9.32). This system is thought to be particularly important in culling senescent cells and self-reactive lymphocytes.

MECHANISMS OF HUMORAL IMMUNITY

B cells are responsible for antibody-mediated (humoral) immunity. B cells have two major subpopulations: memory cells and plasma cells. Memory B cells contain antigen receptors and function in a manner similar to memory T cells. In other words, memory of exposure to an antigen is stored in a clone of memory B cells. When exposed to the same type of antigen in the future, these memory B cells are able to respond rapidly with appropriate antibodies.

Some B cells differentiate into short-lived antibody-producing factories called *plasma cells*. All of the plasma cells in a clone secrete antibodies with identical antigen-binding specificity (monoclonal antibody). The secreted antibodies circulate in the blood and body fluids and bind specifically to the antigen that triggered their production. Once antigen is cleared, the population of plasma cells declines and the antibody concentration (titer) falls. However, some long-lived plasma cells migrate to the bone marrow where they continue to secrete a level of antibody sufficient to provide immediate protection upon the next exposure to the same antigen.

Antigen Recognition by B Cells

During their development in the bone marrow, B cells begin to express BCRs on their cell surfaces. The structure of the antigen-binding area on the BCR is randomly determined in a manner similar to that described for TCRs. Each BCR is coded for by two distinct types of genes: one for the variable region, which makes up the antigen-binding site; and one for the constant region, which is essentially the same for all antibodies of a given class (Fig. 9.33). The structure of the BCR bound to the B-cell

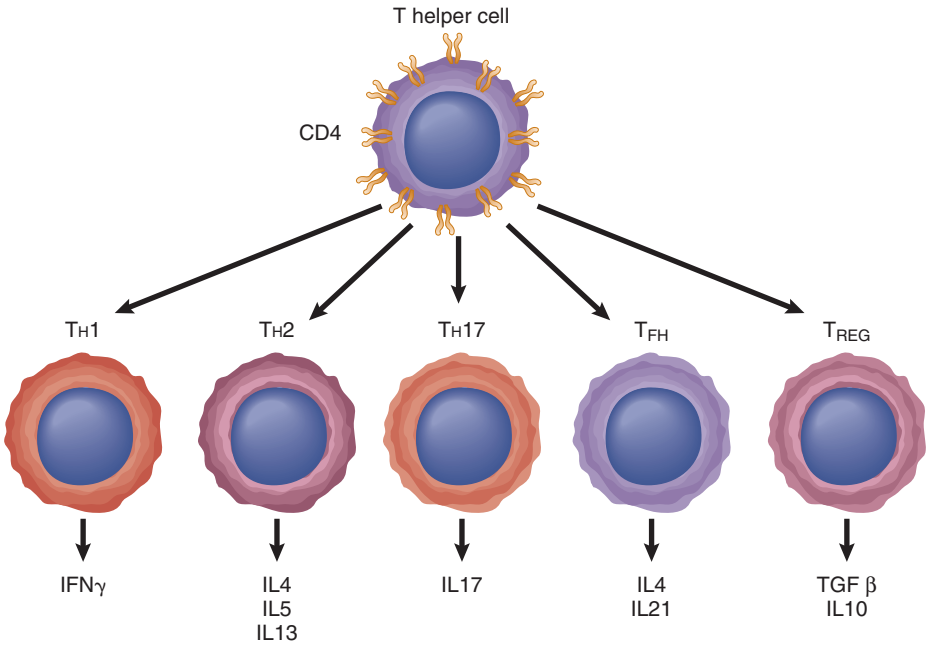


FIG 9.31 Subtypes of CD4 T cells have different functions, including different cytokine production and secretion.

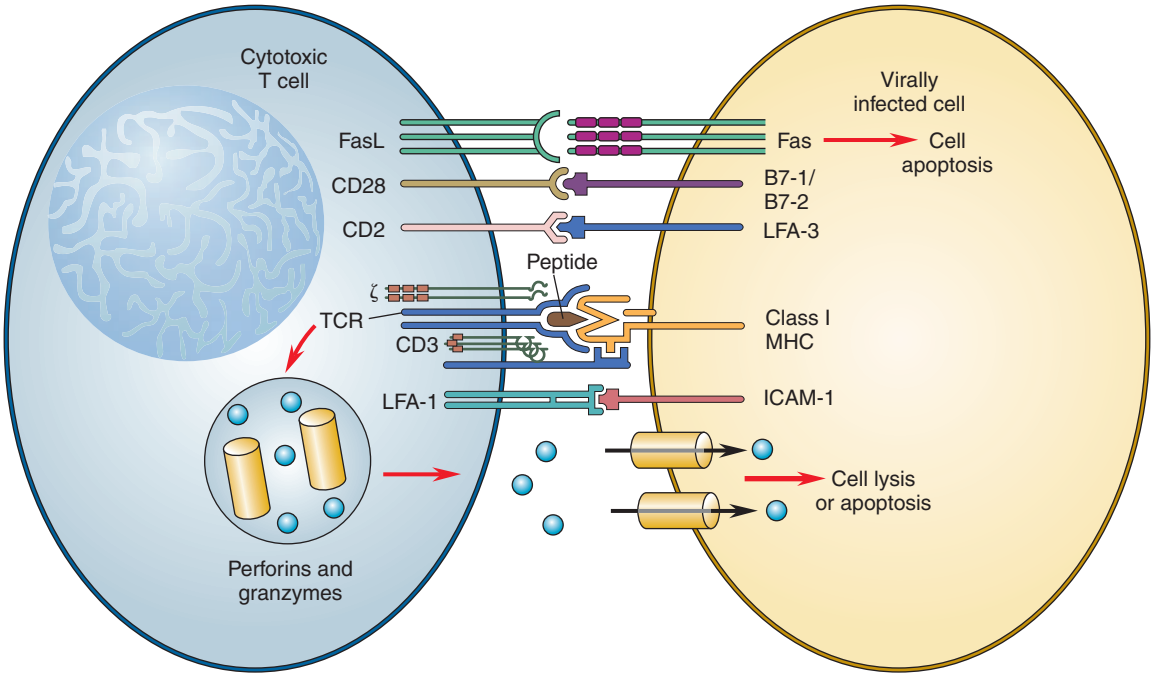


FIG 9.32 Cytotoxic T cells are able to recognize and bind antigen in association with major histocompatibility complex (MHC) class I molecules. The T-cell receptor on the cytotoxic T cell binds to the antigen, and the CD8 protein recognizes the MHC I protein. Binding is specific. Binding of a cytotoxic T cell to its target stimulates granules containing perforin and granzymes to migrate to the cell contact site. Perforins then assemble into pores on the target cell, through which the granzymes can enter the target cell cytoplasm. The granzymes interrupt the cellular DNA and trigger apoptosis. *FasL,* Fas ligand (CD95L); *ICAM,* intercellular adhesion molecule; *LFA,* leukocyte function–associated antigen; *TCR,* T-cell receptor.

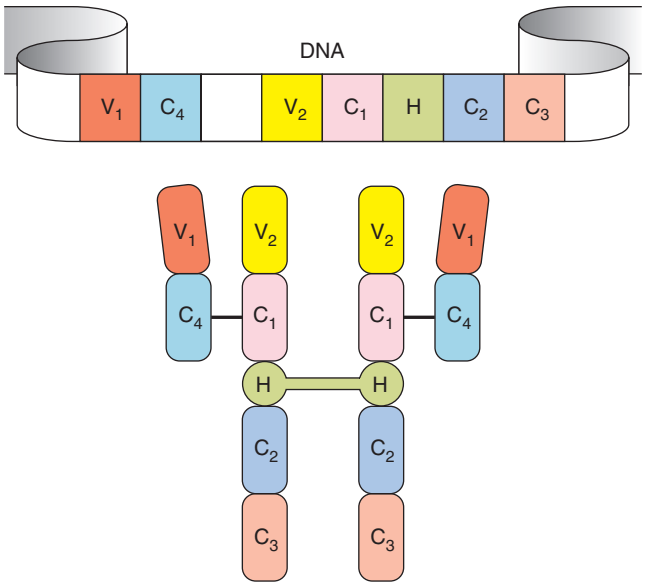


FIG 9.33 Two major classes of genes are responsible for coding for the variable (V) and constant (C) regions of an antibody. Variable genes code for the antibody region that binds to antigen. Constant genes form the stem of the antibody and are the same for any antibody of a given class.

surface is similar to the structure of IgM. Rearrangement, recombination, selective splicing, and random mutations of variable-region genes allow for great diversity of BCR-binding specificities. The potential number of different BCR structures is enormous. Some of these combinations are unsuitable for BCR assembly, and it has been estimated that a typical human B-cell population can recognize approximately 10^{11} different antigenic epitopes. As is the case with T cells, useful B cells—ones activated by antigen—will be preserved in the body, whereas B cells that encounter no antigen will not proliferate. Binding of an antigen to the B cell results in the cross-linking of two surface BCR proteins and initiates an intracellular cascade. These cascades activate transcription factors within the B cell (NFκB, NFAT, AP-1) that regulate genes involved in proliferation and differentiation of the B cell (Fig. 9.34).

The growth and activity of B cells that recognize protein antigens are regulated by helper T cells. Binding of antigen to the B cell's BCR is a necessary but insufficient stimulus to produce an effective B-cell clone in most cases. To be effectively activated, the B cell must engulf some of the antigen, process it, and present it to T helper cells. This activity will initiate cell-to-cell contact between the B cell and its complementary T-cell helper. A number of receptor interactions bind the T cells and B cells together, in addition to the main MHC II–TCR interaction, and include CD80–CD28 and CD154–CD40 (Fig. 9.35). These cell-to-cell binding interactions stimulate intracellular signaling pathways in the B cell (and T helper cell) that promote clonal expansion and differentiation. B cells also require certain cytokines to proliferate and begin antibody synthesis. B cells are quite dependent on T-cell help during the initial exposure to antigen (primary response), but less so on subsequent exposures. Some types of B cells have BCRs that bind nonprotein antigens, such as bacterial carbohydrates and lipids. Because T cells only recognize peptides, these types of B-cell responses are T-cell

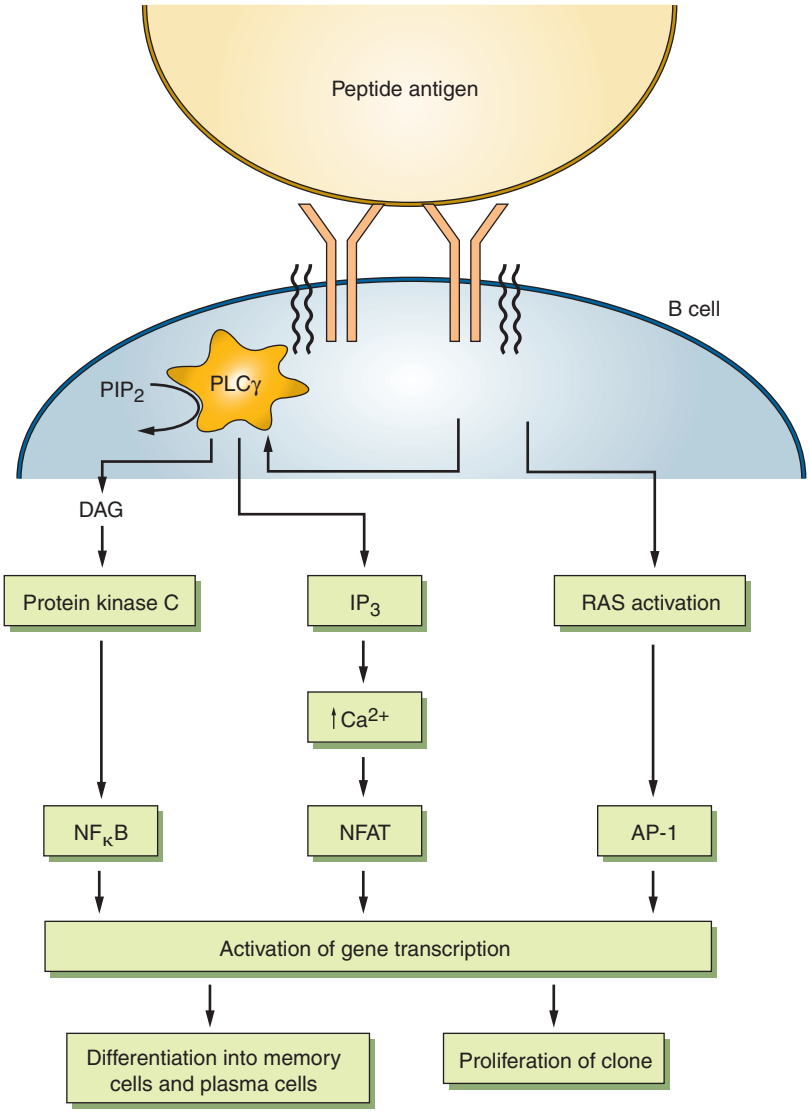


FIG 9.34 Major signaling pathways in B cells. Cross-linking of two surface B-cell receptors initiates intracellular pathways that subsequently activate several transcription factors leading to altered gene activity. *AP-1,* Activation protein-1; *DAG,* diacylglycerol; *IP3,* inositol 1,4,5-trisphosphate; *NFAT,* nuclear factor of activated T cells; *NFκB,* nuclear factor kappa B; *PIP2,* phosphatidylinositol 4,5-bisphosphate; *PLCγ,* phospholipase C-γ.

independent. Other costimulatory signals, such as the complement fragment C3d on the antigen, may provide the necessary costimulation to achieve a B-cell response and antibody production. It is doubtful that memory cells are formed in this process (Fig. 9.36).

ANTIBODY STRUCTURE

Each antibody (immunoglobulin) molecule contains two identical light polypeptide chains joined by disulfide bonds to two identical heavy polypeptide chains. The geometry of the relationship between the heavy (H) and light (L) chains forms a Y-like structure. The H chains form the stem of the Y, and the L chains are on the outside of the arms of the Y. The antigen-binding end of the antibody is often called *Fab* (antigen-binding fragment), whereas the stem is called *Fc* (constant fragment). It is the structure of the constant fragment that determines the antibody class.

Antibodies are differentiated into five classes: IgG, IgM, IgA, IgD, and IgE. The structure and properties of the immunoglobulin classes are listed in Table 9.4. IgG and IgE circulate as single molecules or monomers; IgA is a dimer (two antibodies joined together); and IgM consists of five antibody molecules joined together to form a pentamer. IgD is found mainly on the B-cell plasma membrane and does not circulate in significant quantity. IgD is thought to participate in signal transduction across the B-cell membrane along with the BCR. Different antibody classes serve different immune functions in the body.

IgG, the most common type of immunoglobulin, accounts for 75% to 80% of all immunoglobulins. It is found in nearly equal proportions in the intravascular and interstitial compartments and has a long half-life of about 3 weeks. IgG is the smallest of the immunoglobulins and can more easily escape the bloodstream to enter the interstitial fluid surrounding tissues.

IgM accounts for about 10% of circulating immunoglobulins and is predominantly found in the intravascular pool. Its large pentamer structure prevents it from migrating through the capillary wall. IgM has a half-life of 10 days. It is the first immunoglobulin to be produced on exposure to antigens or after immunization and is the major antibody

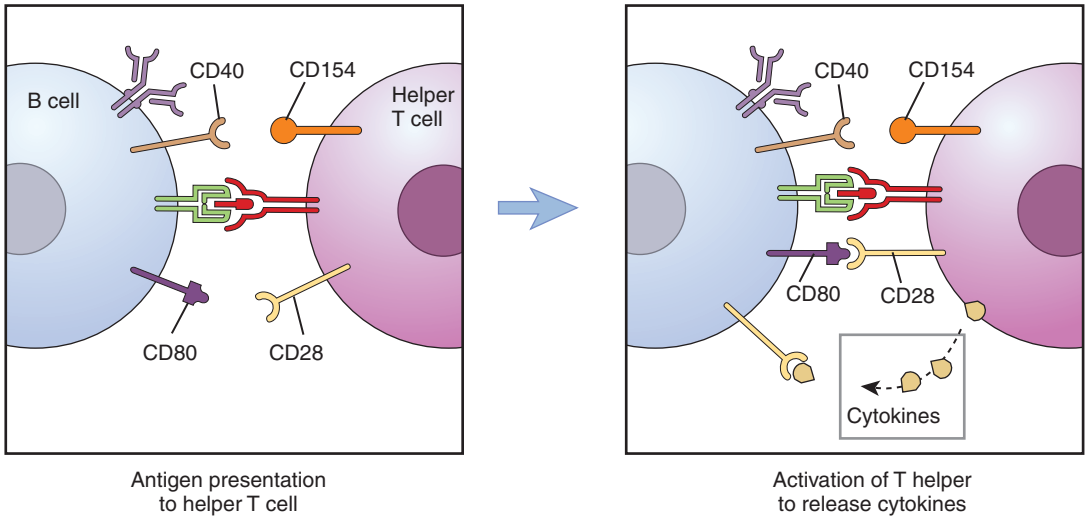


FIG 9.35 Activation of a B cell requires T helper cell "help." This help is given through a number of cell-to-cell interactions via receptors, as well as through the secretion of cytokines that stimulate B-cell growth and differentiation.

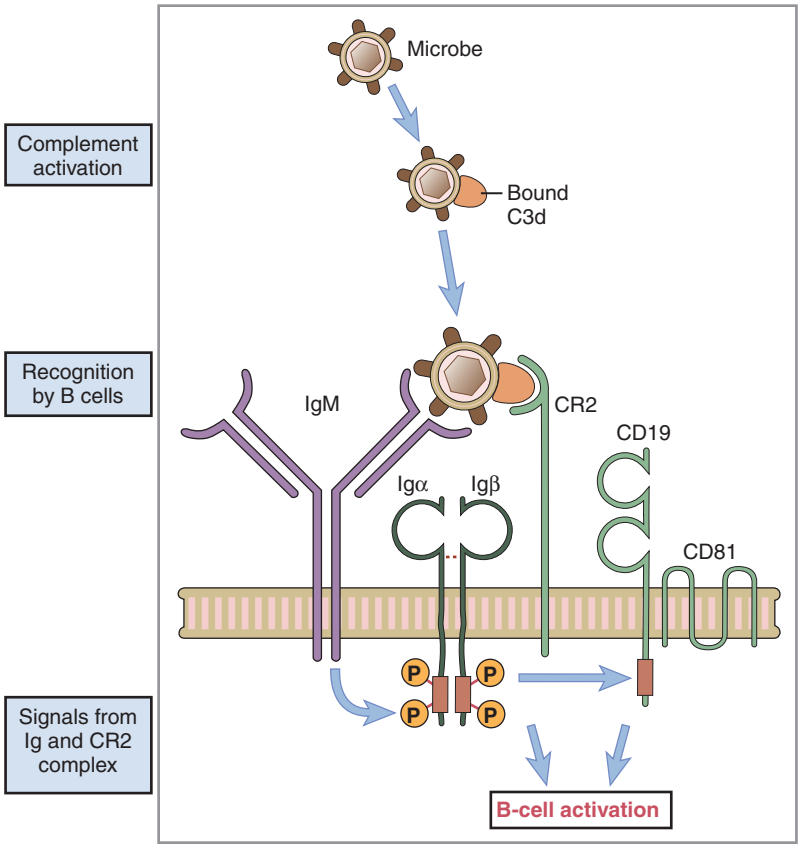


FIG 9.36 In response to nonprotein antigens (T-cell independent), B cells can be activated by complement opsonins on the microbial antigen. The complement–receptor (CR) interaction provides a costimulatory signal to the B-cell receptor–antigen signal. (Redrawn from Abbas AK et al: *Cellular and molecular immunology,* ed 8, Philadelphia, 2015, Saunders, p 159.)

found on B-cell surfaces. IgM is the antibody class that works best to activate complement, which is important for cytotoxic functions in the immune system. Only one molecule of IgM is needed to activate complement, whereas two molecules of IgG are needed to activate complement.

IgA is produced by plasma cells located in the tissue under the skin and mucous membranes. IgA is primarily found in saliva, tears, tracheobronchial secretions, colostrum, breast milk, and gastrointestinal and genitourinary secretions. Transport of IgA into secretions is facilitated by binding to a secretory component produced by epithelial cells. This

TABLE 9.4 Diagram and Properties of Immunoglobulin Classes

Property	IgG	IgM	IgA	IgD	IgE
Half-life (days)	23–25	5	6	3	2.5
Percent total immunoglobulin	80	6	13	0–1	0.002
Molecular weight (daltons)	146,000	900,000	160,000	184,000	200,000
Complement fixation	++	+++	–	–	–
Placental transfer	+++	–	–	–	–
Receptor for macrophage	+++	–	–	–	–
Reaction with staph protein A	+++	–	–	–	–
Passive cutaneous anaphylaxis	+++	–	–	–	+
Transported across epithelium	–	Occasionally ABO isoaggulutinins, rheumatoid factor	+	–	–
Prominent antibody activity	Anti-Rh against infections		Against infections	Binds to B cells in presence of IgM	Mast cell sensitization, cytophilic antibody skin sensitizing antibody
Cell-Binding Functions					
Mononuclear cells	+	–	–	–	?/+
Neutrophils	+	–	+	–	–
Mast cells/basophils	–	–	–	–	+++
T cells/B cells	+	+	+	+	+
Platelets	+	+	–	–	?

Data from Abbas AK: *Cellular and molecular immunology,* ed 8, Philadelphia, 2015, Saunders.

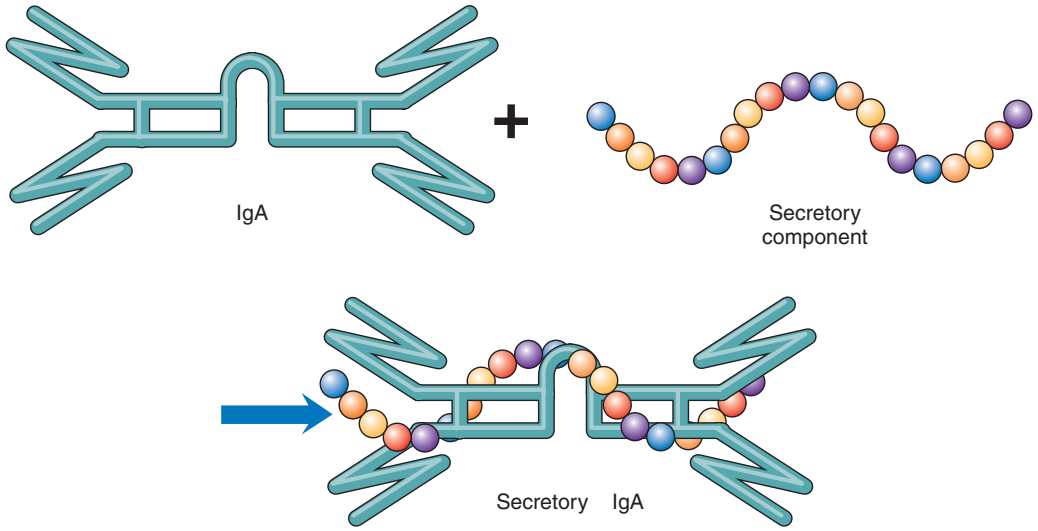


FIG 9.37 IgA is often combined with a protein called secretory component, which helps bind two IgA molecules together at their Fc ends.

complex is called *secretory IgA* (Fig. 9.37). The half-life of IgA is about 6 days.

IgD is found in trace amounts in the serum (1%) and is located primarily on the membranes of B cells along with IgM. IgD has a half-life of 3 days. IgD functions as a cellular antigen receptor acting to stimulate the B cell to multiply, differentiate, and secrete other specific immunoglobulins.

IgE is found bound by its Fc tail to receptors on the surface of basophils and mast cells (Fig. 9.38). Only trace amounts of IgE are identified in the serum. IgE has a half-life of 2 days. It has a role in immunity against helminthic parasites (worms) and is responsible for initiating inflammatory and allergic reactions (e.g., asthma, hay fever). IgE functions as a signaling molecule and causes mast cell degranulation when antigen is detected at the mast cell surface (see Chapter 10).

Class Switching and Affinity Maturation

During the course of an antibody response, the class of antibody manufactured by a particular B cell usually changes. The antigen-binding

site does not change significantly and remains specific for the particular antigen that initiated the response. To switch class, the B cell selects different constant region genes to splice to the antigen-binding fragment. Thus most B cells begin by using genes that code for IgM and IgD. Then the B cell switches to produce IgG, IgE, or IgA. The triggers that determine the class of antibody a particular B cell will produce are not completely understood. Some cytokines have a role in class switching.

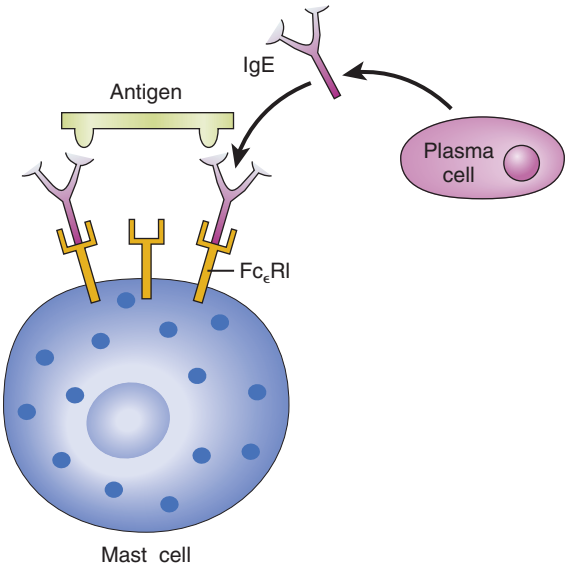


FIG 9.38 Mast cells bind IgE antibody with their Fc receptors (FcεRI) and display the IgE on the cell surface, where they are available to bind antigens.

For example, IFN-γ promotes IgG production, IL-4 promotes IgE production, and transforming growth factor-β (TGF-β) promotes IgA production (Fig. 9.39).

Knowledge about the normal progression of class switching may be helpful in determining whether an infectious process is acute or chronic. For example, a person newly infected with hepatitis B virus would be expected to have primarily IgM antihepatitis B antibodies, whereas in chronic or previous infection, B cells would switch class to produce mainly IgG. The relative concentrations of antihepatitis B IgM and IgG can help identify the time of onset of the infection.

Over the course of a B-cell antibody response, the affinity with which the antibodies bind to antigen often increases. This is thought to occur because of a process called *affinity maturation* during which B cells undergo a hypermutation response, producing random changes in the antigen-binding pocket of the BCR. Those that bind antigen most avidly are stimulated to proliferate to a greater extent. Thus the antibodies formed later in an immune response are more efficient in binding antigen at lower and lower concentrations. Affinity maturation occurs in specialized germinal centers in the lymph nodes.

Antibody Functions

Antibodies function in a number of ways to enhance the localization and removal of antigens from the body. These functions can generally be summarized as precipitation, agglutination, neutralization, opsonization, and complement activation. Precipitation and agglutination occur because each arm of the immunoglobulin Y structure can bind an antigenic epitope. This structure allows the antibodies and antigens to bind together into large insoluble complexes that precipitate out of body fluids (Fig. 9.40). Agglutination refers to the same process as applied to cellular antigens rather than soluble antigens. It is efficient for phagocytic cells to find the large complexes and clear them from the system.

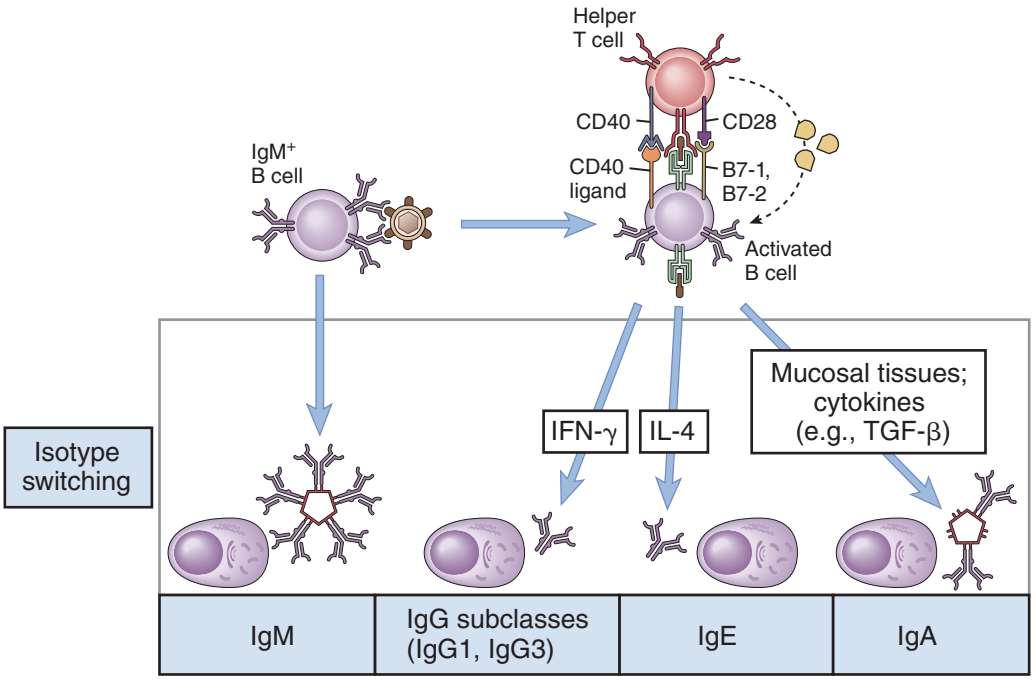


FIG 9.39 Activated B cells undergo class switching from IgM to IgG, IgE, or IgA. Class switching is influenced by the presence of specific cytokines. *IFN,* Interferon; *IL,* interleukin; *TGF,* transforming growth factor. (Redrawn from Abbas AK et al: *Cellular and molecular immunology,* ed 8, Philadelphia, 2015, Saunders, p 252.)

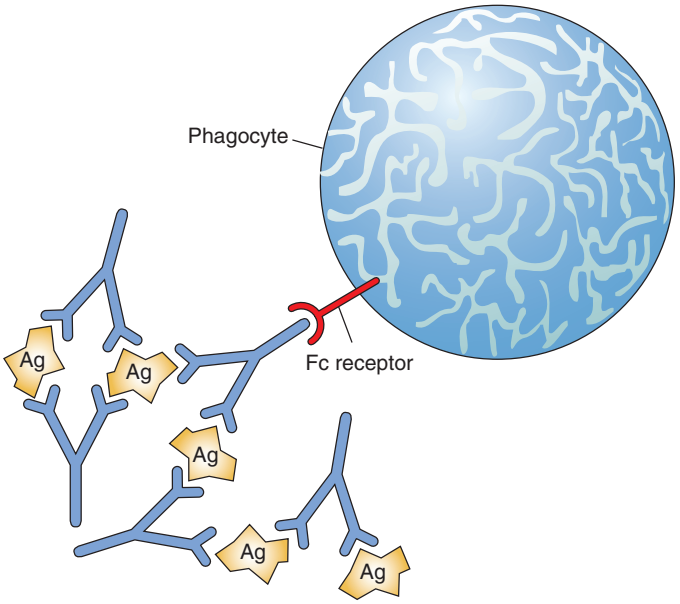


FIG 9.40 Large antigen (Ag)–antibody complexes tend to precipitate out of solution, which makes it easier for phagocytic cells to find and eliminate the antigens.

Antibodies can function as antitoxins by neutralizing bacterial toxin. This role is accomplished by binding the toxins before they can interact with cells or by covering the active portions of the toxin and inactivating it. Some antibodies are effective opsonins. They coat the foreign antigen and thereby make it more recognizable to phagocytic cells. Macrophages, neutrophils, eosinophils, and NK cells have receptors for the Fc ends of the antibodies, which help them bind to opsonized antigens. Antibodies thus make the innate phagocytic processes more efficient. Antibodies (IgG and IgM) bound to foreign antigens on cell membranes also can activate the complement cascade and trigger the release of chemotaxins and inflammatory mediators and the formation of the membrane attack complex.

PASSIVE AND ACTIVE IMMUNITY

Immunity is a state of resistance against infection from a particular pathogen. Immunity is provided primarily by adequate levels of circulating antibodies. Specific serum antibody concentrations can be measured by a blood test called an *antibody titer*. A sufficiently high antibody titer confers immunity by removing pathogens from the body before they cause signs and symptoms of illness. Immunity can be achieved passively or actively.

Passive Immunity

Passive immunity involves the transfer of plasma (sera) containing preformed antibodies against a specific antigen from a protected or immunized person to an unprotected or nonimmunized person. As a treatment, this is indicated in the following situations: (1) when B-cell immunodeficiency exists; (2) when highly susceptible persons are exposed to a disease without adequate time for active immunization; and (3) when antibody injection may alleviate or suppress the effects of an antigenic toxin.

Passive transfer of antibodies can occur in a variety of ways. In the fetus, certain maternal IgG antibodies can cross the placental barrier. Most of the time these antibodies are beneficial and assist the newborn in resisting pathogens. However, in some cases these antibodies can be damaging to the fetus, as occurs in hemolytic disease of the newborn. In this disorder, maternal antibodies bind to and lyse fetal red blood cells (see Chapters 10 and 13).

Antibody, complement, and macrophage function is deficient at birth. Newborns who are breast fed may have improved immune function. Newborns receive IgA antibodies through breast milk. The infant's immature gastrointestinal tract and low proteolytic enzyme activity do not destroy all protein, which allows some of the IgA antibodies to be absorbed. These antibodies assist the infant in defending against bacterial and viral infections during infancy. It has been hypothesized by some researchers that IgA antibodies in breast milk may modify the ways that proteins cross the infant's highly permeable intestinal mucosa and help prevent food allergies in later life (see Pediatric Consideration box).

Another method of passive immunity, called *serotherapy,* involves direct injection of antibodies into an unprotected person. The unprotected individual can receive a variety of substances, including immune globulin (human) such as IgG; specific immune globulins like hepatitis B immune globulin (human) or rabies immune globulin (human); plasma containing all human antibodies; or animal antibodies such as diphtheria antitoxin, tetanus antitoxin, botulism antitoxin, and antirabies serum.

Human immune globulin contains mostly IgG with traces of IgA and IgM. It is a sterile, concentrated protein solution that contains antibodies from the pooled plasma of many adults. It can be administered intramuscularly or intravenously, depending on the product. Human immune globulins may be used as prophylaxis against hepatitis B and as therapy for the following conditions: antibody deficiency disorders, pediatric acquired immunodeficiency syndrome, and hypogammaglobulinemia after bone marrow transplantation.

Animal antibodies are given in specific situations only when necessary because of significant allergic risks with animal sera. Patients who have specific animal allergies or a history of asthma, allergic rhinitis, or other allergies are highly susceptible to serum sickness, anaphylaxis, or acute febrile reactions. Serum sickness occurs when antibodies bind to foreign proteins in the injected sera, forming immune complexes that precipitate into capillaries and joints and cause inflammation. Animal antibodies may be given to ameliorate toxins or venoms, such as those associated with botulism, diphtheria, rabies, tetanus, and snake and spider bites.

Active Immunity

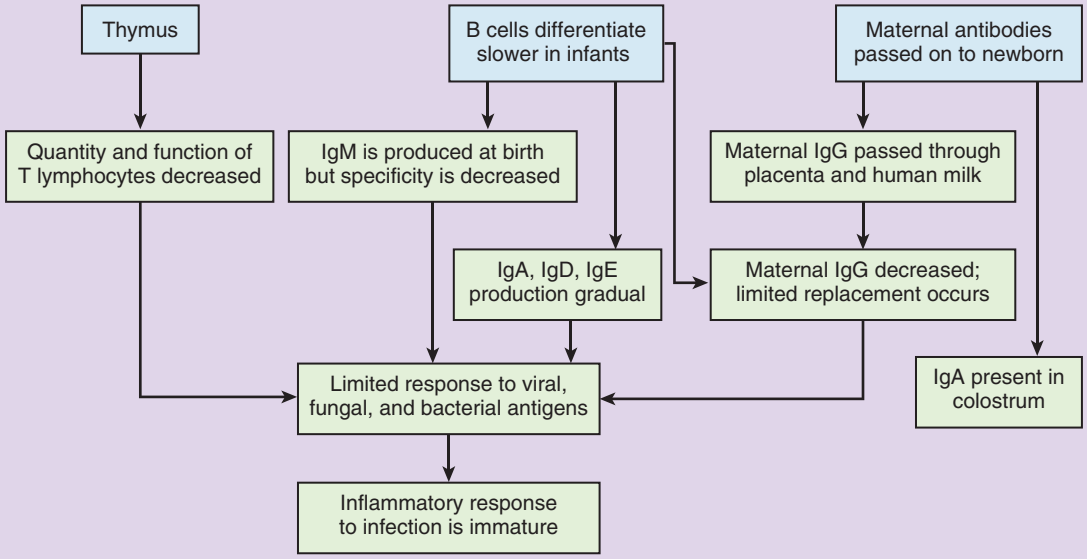
Active immunity confers a protected state attributable to the body's immune response as a result of active infection or immunization. The development of active immunity requires the maturation and maintenance of memory B cells. On second exposure to antigen, the antibody response is much greater and more rapid (Fig. 9.41). Exposure to antigen can be achieved through active infection or through immunization. The immune system must be exposed to the antigen at a sufficient dose for an adequate length of time to stimulate an immune response.

Immunization tricks the immune system into responding to a perceived infection. Vaccines contain altered microorganisms or toxins that retain their ability to stimulate the immune system (antigenic properties) but do not have pathogenic properties. Vaccines can contain live and attenuated (altered) or killed infectious agents.

Vaccines that contain live, altered viruses or bacteria cause active infection but little injury to the vaccinated individual. These vaccines mimic a natural immune response, activating B and T cells, and provide good humoral and cellular immunity with longer-lasting memory and often lifetime immunity. Examples of vaccines registered in the United States are listed in Box 9.1.

PEDIATRIC CONSIDERATIONS

Changes in the Immune System in Infants



The immune system is immature in an infant. Infants generally do not produce immunoglobulin (Ig) until the beginning of the second month of life, and then production is slow (Hockenberry and Wilson, 2011). IgM is produced at birth, but specificity is decreased, limiting the infant's ability to fight some antigens. IgM reaches adult levels by 9 to 12 months. IgA, IgD, and IgE begin gradual production in the second month of life and reach adult levels around early childhood.

Infants rely on maternal antibodies for production until their own immune system can mature. Maternal IgG is passed through the placenta and breast milk and for the first 3 months provides protection to the infant from antigens to which the mother has been exposed. IgG levels decline after 4 months and remain low until 6 months of age. Forty percent of infants achieve adult levels of IgG by 1 year of age, with the remaining reaching adult IgG levels by 4 years of age. IgA is present in the colostrum and believed to protect the infant's gastrointestinal tract until more IgA can be produced.

The thymus is large in infants and decreases in size over childhood. By the end of adolescence, the thymus shrivels and its function declines. The infant's immune response is immature, resulting in a limited response to viral, fungal, and bacterial antigens.

Reference

Hockenberry MJ, Wilson D: *Wong's nursing care of infants and children*, ed 8, St Louis, 2011, Mosby.

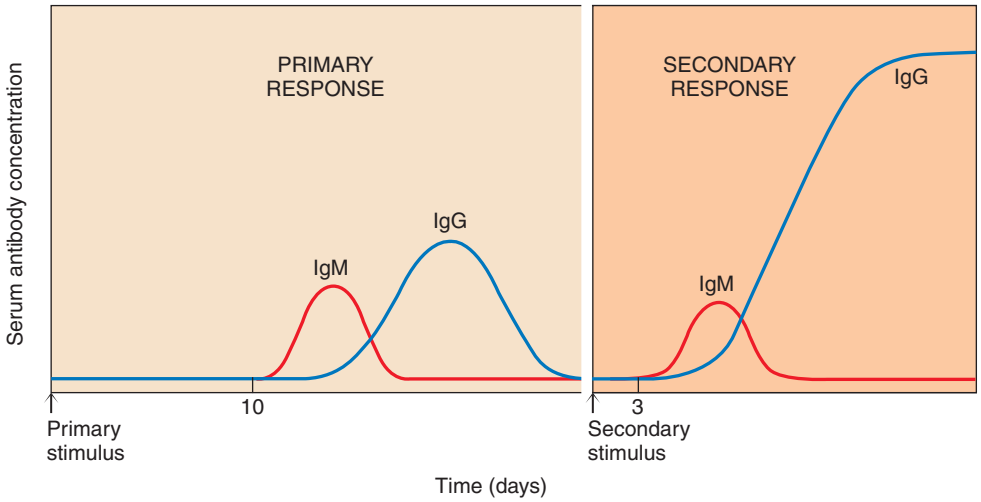


FIG 9.41 Time phases in the immune response. The primary response takes much longer to develop and declines rapidly. On second exposure, a much quicker and greater antibody response is achieved.

KEY POINTS

- Specific immunity refers to functions of B and T lymphocytes. Each lymphocyte recognizes and reacts to only one particular antigen. On initial exposure to an antigen, lymphocytes undergo clonal expansion; consequently, many lymphocytes are distributed throughout the body to recognize and react to that particular antigen. These cells are called *memory cells*. Subsequent exposure results in a much faster and larger lymphocyte response.
- T cells, which mature in the thymus, have two major subgroups: T helper cells and cytotoxic T cells. T helper cells perform a central role in specific immunity. Activation of T helper cells results in secretion of the cytokines necessary for clonal expansion of T and B lymphocytes. Cytotoxic T cells locate and lyse abnormal cells through the actions of perforins.
- T lymphocytes are able to bind antigens only when they are displayed on the surface of cells. Cytotoxic T cells ($CD8^+$) react to cells that have foreign MHC class I proteins on their surface. T helper cells ($CD4^+$) bind to cells that have MHC class II proteins on their surface. MHC class II proteins are found on antigen-presenting cells (B cells, dendritic cells, and macrophages). These cells engulf foreign antigens and combine the antigens with MHC class II proteins on their cell surface.
- B and T cell functions are interdependent. T cells cannot respond to soluble antigens. B cells can process free antigen and present it to T cells. On first exposure, B cells are minimally activated by antigen unless they are stimulated by cytokines and coreceptors from T cells.
- B lymphocytes mature in bone marrow and lymph tissue. B cells have receptors on their surfaces that can bind antigens. Each B cell binds only one particular antigen. With appropriate T helper cell "help," antigen binding causes the B cell to divide (clonal expansion). Some of the daughter cells become plasma cells, which actively produce and secrete antibodies. Other daughter cells (memory cells) resemble the original cell and are distributed in lymph throughout the body. On subsequent exposure to the antigen, antibody production is rapid.
- Antibodies are proteins that specifically bind a particular antigen. Antibodies have several functions, including precipitation, agglutination, neutralization, opsonization, and complement activation.
- The five major antibody classes are IgG, IgM, IgA, IgD, and IgE. Antibody class is determined by the structure of the Fc portion. IgG is the most prevalent antibody class (75%). IgM is the first kind to be produced on antigen exposure. IgA is found primarily in body secretions. IgD is present on the B-cell membrane and functions in signal transduction. IgE binds to basophil and mast cell membranes and mediates inflammation and allergy.
- Administration of preformed antibodies confers passive immunity. Passive immunity provides immediate but temporary protection. Active immunity occurs when individuals are exposed to antigen that stimulates their own lymphocytes to produce memory cells. Active immunity confers long-term protection but may take several weeks to develop.

INTEGRATED FUNCTION AND REGULATION OF THE IMMUNE SYSTEM

The innate and adaptive cells of the immune system work interdependently to protect the host from foreign antigens. Efficient interdependent function depends on a complex communication network that allows coordination of various immune components. One of the reasons that the immune system uses such a complex communication system is to ensure that normal healthy tissue is not injured. The destructive powers of the immune system must be tightly regulated to avoid undue tissue damage. These regulatory controls can be affected by aging and disease. The effects of aging on immune function are described in the box Geriatric Considerations: Changes in the Immune System. In the following sections, major events in the immune response to a new antigen are summarized and mechanisms of immune regulation are described.

BOX 9.1 Selected Vaccines Available for Immunization in the United States

Adenovirus Type 4 and Type 7 Vaccine, Live, Oral
Anthrax Vaccine Adsorbed
BCG Vaccine
Diphtheria and Tetanus Toxoids Adsorbed
Diphtheria and Tetanus Toxoids and Acellular Pertussis Vaccine Adsorbed
Diphtheria and Tetanus Toxoids and Acellular Pertussis Vaccine Adsorbed, Hepatitis B (recombinant) and Inactivated Poliovirus Vaccine Combined
Haemophilus b Conjugate Vaccine (plus various combinations including: Diphtheria, Meningococcal Protein Conjugate, Tetanus Toxoid Conjugate, and Hepatitis B)
Hepatitis A Vaccine, Inactivated
Hepatitis B Vaccine (Recombinant)
Human Papillomavirus Quadrivalent (Types 6, 11, 16, 18) Recombinant Vaccine
Human Papillomavirus Bivalent (Types 16, 18) and 9-Valent
Influenza Virus Vaccine (Various including Types A, H1N1, B, H5N1)
Japanese Encephalitis Virus Vaccine Inactivated
Measles Virus Vaccine, Live
Measles and Mumps Virus Vaccine, Live
Measles, Mumps, and Rubella Virus Vaccine, Live
Measles, Mumps, Rubella, and Varicella Virus Vaccine, Live
Meningococcal Polysaccharide Vaccine, Groups A, C, Y, and W-135 Combined
Pneumococcal Vaccine, Polyvalent
Pneumococcal 7-Valent Conjugate Vaccine
Pneumococcal 13-Valent Conjugate Vaccine Poliovirus Vaccine Inactivated
Rabies Vaccine
Rotavirus Vaccine, Live, Oral, Pentavalent
Smallpox (Vaccinia) Vaccine, Live
Tetanus and Diphtheria Toxoids Adsorbed for Adult Use
Tetanus Toxoid
Tetanus Toxoid, Reduced Diphtheria Toxoid, and Acellular Pertussis Vaccine Adsorbed
Typhoid Vaccine Live Oral Ty21a
Typhoid Vi Polysaccharide Vaccine
Varicella Virus Vaccine Live
Yellow Fever Vaccine
Zoster Vaccine, Live

Data from www.fda.gov/BiologicsBloodVaccines/Vaccines/ApprovedProducts/ucm093833.htm.

INTEGRATED RESPONSE TO MICROBIAL ANTIGEN

A new microbial antigen entering the body through the skin or mucous membranes will generally encounter tissue macrophages and dendritic cells stationed in strategic locations in the body as part of the mononuclear phagocyte system. Macrophages initiate activity of both innate and specific immune components (Fig. 9.42). First, activated macrophages release cytokines that initiate inflammation and chemotaxis. Some of these cytokines (TNF-α, IL-1) induce capillary endothelial cells to express selectins and integrin ligands that help circulating leukocytes adhere to the capillary wall (margination) and then move into the tissue locations of antigens (emigration). Neutrophils, macrophages, and NK cells are attracted to the infected area by chemotactic factors, some of which are released by macrophages, whereas others are products of the

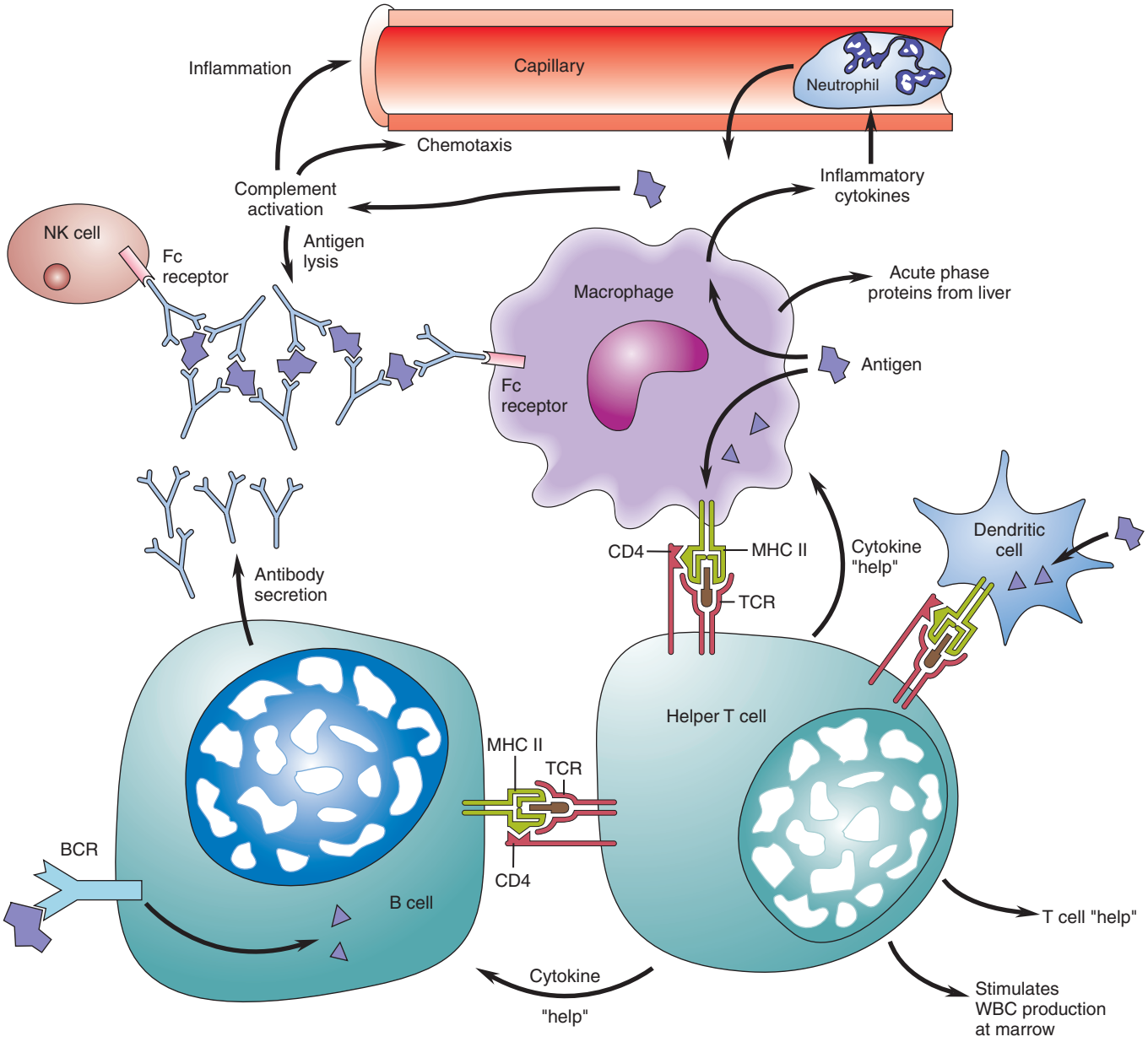


FIG 9.42 Diagram showing the integrated function of a number of immune components. Note that the macrophage is at the center of many immune functions, including chemotaxis and inflammation, presentation of antigen to T cells, and phagocytosis of antibody–antigen complexes. *BCR,* B-cell receptor; *MHC,* major histocompatibility complex; *NK,* natural killer; *TCR,* T-cell receptor; *WBC,* white blood cell.

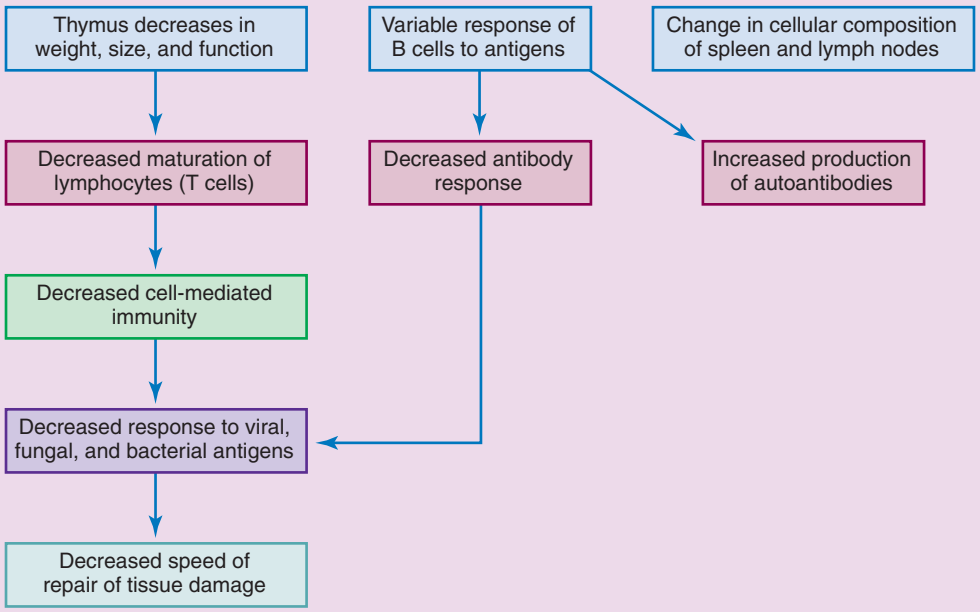
complement cascade and tissue injury. The complement cascade is activated by the alternative pathway on primary exposure to an antigen. Complement fragments C3a and C5a are potent inflammatory agents. Complement activation also results in the formation of membrane attack complexes that directly lyse cellular antigens.

Tissue injury associated with the infectious process also activates both the coagulation cascade, which forms a fibrin meshwork to help entrap and localize the agent, and the kinin system, which promotes vasodilation to increase blood flow to the area. A number of other vasodilatory chemicals may be released from mast cells in the area when they degranulate. These inflammatory chemicals lead to the classic manifestations of inflammation: warmth, swelling, redness, pain, and loss of function. Neutrophils, macrophages, and NK cells that emigrate to the tissue find their targets through innate receptors on their cell surfaces. Thus they are as effective on first exposure to an antigen as they are on subsequent exposures. These receptors bind to complement opsonins and molecules expressed on microbes such as lipopolysaccharide and mannose. NK cells release cytotoxins onto their targets; macrophages and neutrophils phagocytose and digest their targets.

Dendritic cells and macrophages ingest protein antigens to process and present them to T helper cells in association with MHC class II proteins. Dendritic cells move from the tissue and travel to the T-cell zones of lymph nodes. Activation of T helper cells results in the secretion of a variety of cytokines that boost the growth and activity of many immune cells, including macrophages, neutrophils, NK cells, cytotoxic and T helper cells, and B cells. Some of the cytokines produced by

GERIATRIC CONSIDERATIONS

Changes in the Immune System



In the elderly, immune system function is altered with a decreased ability to respond to antigenic stimulation. The elderly are able to respond to infections with previously produced "remembered" antibodies. However, they are less able to respond to new antigens. As a result of these changes, there is decreased speed of repair of tissue damage and increased vulnerability to disease. The cells of the immune system in elderly persons are not able to proliferate or reproduce as effectively as those in younger persons. Although the total number of T cells remains the same, T-cell function is decreased. T cells are less able to proliferate and have decreased cytotoxicity. Antibody production also decreases, especially antibodies such as IgG. There is also a rise in autoantibody production, which may influence the increase in autoimmune disease in the elderly. Thymus size decreases after puberty, causing a decline in thymic hormone production, decreased T-cell differentiation, and reduced T-cell–related B-cell differentiation. Usually thymic hormone secretions stop after age 60 years. However, the role of thymus involution in elderly immune system changes is currently uncertain.

activated T cells and macrophages stimulate stem cells in the bone marrow to produce more WBCs (neutrophilia). Other cytokines affect the brain (inducing fever, lethargy, and anorexia) and the liver (producing acute phase proteins).

Meanwhile, certain B cells that encounter the antigen in the lymph nodes will have the correct BCR to bind and internalize it. Internalized antigen is then processed and presented to T helper cells in association with B-cell MHC II proteins. Complementary T helper cells then bind the B cell (via MHC II–TCR–CD4 interactions) and provide help to the B cell through the secretion of cytokines and through coreceptor-mediated second-messenger signals. B cells thus activated proliferate into a clone of cells, with some becoming memory cells and others becoming plasma cells. Plasma cells synthesize and secrete antibodies that specifically bind the antigen. Significant antibody production takes 10 to 14 days to occur, and the infected individual may have signs and symptoms of illness during this time.

Antibodies enhance the function of innate phagocytic cells by collecting antigen into large complexes that are easier for nonspecific cells to locate and phagocytose. Activated T helper cells also secrete cytokines such as IL-2 and IFN-γ, which enhance the effectiveness of macrophages. After the antigen is cleared from the body, macrophages perform clean-up functions to remove inflammatory debris and dead neutrophils from the tissue. Macrophages also secrete enzymes and growth factors that stimulate tissue healing.

After the primary infection, B and T memory cells populate the body in much larger numbers and can mount an effective immune response quickly on second exposure. The individual then has immunity for the particular pathogen because the antigen will usually be cleared from the system before significant illness occurs.

INTEGRATED RESPONSE TO VIRAL ANTIGEN

If the infectious agent is a virus, the sequence of events is somewhat different. Virally infected cells initiate cytotoxic T-cell activity, which serves to kill the infected cells. NK cells are important for detecting and destroying virally infected cells that have down-regulated their MHC I proteins, making themselves invisible to cytotoxic T cells. Helper T-cell responses and B-cell production of antibody to the virus occur by the processes previously described. However, neutrophils are less important in the response to viral infection.

Virally infected cells initiate a number of mechanisms to disrupt viral replication and prevent spread of viruses to other cells. Viral components such as viral double-stranded RNA or viral DNA can be bound to pattern recognition receptors within the host cells. These internal receptors are of the Toll-like receptor family and when activated initiate the production of type I interferons (IFN-α and IFN-β). Secreted interferons bind to receptors on nearby cells and help block viral infection

and replication by inhibiting protein synthesis and activation of enzymes that degrade RNA. These measures are taken to rid the body of virus even though they may result in significant cell injury and death. Virally infected cells may even express "death receptors" on their surface to alert the immune system to trigger an attack on them.

REGULATION OF IMMUNE FUNCTION

The mechanisms that promote inflammation and enhance immune function are much better understood than those that negatively regulate these processes. However, the mechanisms for inactivating an immune response and keeping inflammation in check are just as important. The destructive powers of the immune system can cause severe tissue damage unless carefully controlled.

Inhibition of immune responses occurs in a number of different ways. The process of inducing tolerance to self antigens is of primary importance. Because both T and B lymphocytes produce antigen-binding receptors by a random process, generation of self-reactive lymphocytes cannot be prevented. As previously mentioned, B and T cells are subjected to a rigorous selection process as they mature in the bone marrow and thymus, respectively. Several theories have been proposed to explain how self-reactive cells are detected and eliminated. The clonal deletion theory suggests that cells in the thymus process and present self antigens to developing T cells. Those lymphocytes that avidly bind self antigens are triggered to initiate programmed cell death (apoptosis). There appears to be a critical time in fetal development when self antigens begin to be differentiated from foreign antigens. Before that time, antigens introduced into the fetus will be viewed as "self," and tolerance to them will develop. By the same token, self antigens that are not presented to T cells in the thymus may be viewed as foreign. This situation may occur with certain so-called *sequestered antigens,* as would be found in the interior of the eye or testes. If these antigens are later released by trauma, an immune response may be directed against them.

Clonal deletion may not rid the body of all self-reactive lymphocytes; therefore many safeguards are in place to prevent their activation. A complex process of antigen processing and presentation is required before T and B cells can be effectively activated. A certain "dose" of antigen must be present to achieve an effective response. Antigen in very high concentrations appears to cripple lymphocyte responsiveness and may initiate apoptosis. Self antigens may be present in such high quantities that reactive lymphocytes are killed. Because dendritic cells, macrophages, and B cells are important antigen-presenting cells, they can exert some influence on T-cell activation by controlling the dose of antigen presented. Certain cytokines are known to influence the production of MHC proteins and can therefore alter the amount of antigen to which T cells are exposed.

B-cell activation requires a number of costimulatory signals from different sources. This complexity helps ensure that B cells will be activated appropriately. These signals include antigen binding to the BCR, T helper binding to the B-cell MHC class II protein, expression of costimulatory ligands and receptors, and secretion of cytokines that promote B-cell growth and differentiation into memory cells and plasma cells. In addition, B cells are subject to negative feedback by circulating antibodies.

Another mechanism of immune suppression is accomplished through cells that secrete inhibitory chemicals. Some subtypes of CD4 T cells may perform regulatory functions. They can inhibit immune responses by secreting immunosuppressive cytokines such as IL-10 and TGF-β.

Control of the complement, kinin, and clotting systems is achieved by a number of inhibitory binding proteins. C1 inhibitor, a glycoprotein, inhibits both Hageman factor (factor XII) and activated portions of C1. Other portions of the complement system are regulated by other binding proteins (e.g., factor I, factor H, and S protein). S protein is of particular importance. It prevents the complement membrane attack complex from attaching to and lysing cell membranes.

Protease inhibitors synthesized by the liver, such as α_1-antitrypsin, help reduce excessive protein destruction by inhibiting destructive enzymes released from neutrophils.

The neuroendocrine system also has a role in immune regulation. Immune cells have receptors for glucocorticoid hormones and a number of neuropeptides, including enkephalins, endorphins, adrenocorticotropic hormone, oxytocin, somatostatin, and substance P. It is a well-known phenomenon that stress and depression can lead to reduced immune function. Some of these hormones are believed to be responsible for this effect. The immune system also affects the nervous system through secreted cytokines such as IL-1 and TNF-α, which induce sleep and malaise.

One of the most important mechanisms of terminating an immune reaction is the elimination of the inciting antigen. As the antigen is cleared, levels of many of the cytokines and costimulators are reduced so that "survival signals" are no longer given to lymphocyte populations and they undergo apoptosis.

Despite these complex and effective regulatory mechanisms, immune and inflammatory disorders are extremely common. Chapter 10 describes the pathophysiology of the common overreactions and underreactions of the immune system.

KEY POINTS

- Specific and innate immune cells work together to protect the body from foreign antigens. Macrophages and dendritic cells play a central role because they are commonly the first immune cells to encounter the antigen. Macrophages secrete cytokines that stimulate WBC production and help WBCs locate the area. Tissue reactions activate the clotting cascade and kinin system, which help localize the antigen and promote movement of fluid and immune cells into the tissue.
- Macrophages and dendritic cells are antigen-presenting cells that engulf and display antigen on their cell surface in association with MHC class II proteins. T helper cells are specifically activated by these antigen-presenting cells. T helper cells secrete cytokines that promote the production of WBCs in the marrow, initiate proliferation of mature B and T cells, and stimulate the phagocytic potential of macrophages and neutrophils.
- B-cell proliferation and antibody secretion usually require T cell help. B cells internalize and present antigen to T cells, which then stimulate B-cell proliferation. B cells secrete antibodies that help phagocytic cells localize and destroy antigens.
- The immune response to primary exposure is slow and often insufficient to prevent illness. Memory cells that develop during primary exposure can mount a more effective response on subsequent exposure and usually prevent manifestations of illness.
- T and B lymphocytes must be tolerant to self. T lymphocytes capable of reacting with self tissue are thought to be destroyed or permanently inactivated during development in the thymus. One theory suggests that lymphocytes must come in contact with all self antigens during development and those that do not specifically bind self antigens are allowed to survive.
- B cells are subject to careful regulation by T helper cells and by negative feedback from high concentrations of circulating antigen–antibody complexes.
- Mechanisms to inhibit and control the immune response include activation of regulatory T-cell cytokines, complement inhibitors, circulating antiproteases, and antioxidants as well as degradation of inflammatory mediators.

SUMMARY

Cells and tissues throughout the body participate in defense against foreign antigens. Some components of the immune system are able to react to a large number of foreign invaders upon first exposure. These innate components are essential for protecting the body while the specific immune defenses are being activated. Innate defenses include physical and biochemical barriers of the skin and mucous membranes, cells of the mononuclear phagocyte system, neutrophils, NK cells, and a large number of chemical mediators such as complement, clotting factors, kinin, and cytokines. Immunity to specific antigens is provided by B and T lymphocytes. T helper cells are important regulators of the immune system because they secrete cytokines that enhance T-cell, B-cell, and macrophage function.

The forces of inflammation and immunity must be carefully controlled to prevent excessive tissue damage. Extensive measures are used to rid the body of self-reactive lymphocytes and to control reactions once a foreign antigen has been cleared. A well-functioning immune system not only successfully protects against foreign invaders and learns from the process so that it is even more effective on subsequent exposure, but also leaves healthy normal tissue unharmed.

RESOURCES

Immunology

Abbas AK, Lichtman AH, Pillai S: *Cellular and molecular immunology*, ed 8, Philadelphia, 2015, Saunders.

Alberts B, et al: *Molecular biology of the cell*, ed 6, New York, 2015, Garland Science, pp 1297–1342.

Brostoff J, Roth D, Roitt I, editors: *Immunology*, ed 8, St Louis, 2012, Mosby, pp 17–50.

Dieffenbach CW, Tramont EC: Innate (general or nonspecific) host defense mechanisms. In Bennett JE, Dolin R, Blaser MJ, editors: *Mandell, Douglas, and Bennett's principles and practice of infections diseases*, ed 8, Philadelphia, 2015, Saunders, pp 26–33.

Hall JE: Resistance of the body to infection: II. Immunity and allergy. In *Guyton and Hall textbook of medical physiology*, ed 13, Philadelphia, 2016, Saunders.

Hanson LA, Korotkova M, Telemo E: Breast-feeding, infant formulas, and the immune system. *Ann Allergy Asthma Immunol* 90(6 Suppl 3):59–63, 2003.

Kasahara E, Inoue M: Cross-talk between HPA-axis-increased glucocorticoids and mitochondrial stress determines immune responses and clinical manifestations of patients with sepsis. *Redox Rep* 20(1):1–10, 2015.

Kumar V, Abbas A, Fausto N, Aster J: Inflammation and repair. In Kumar V, Abbas A, Fausto N, Aster J, editors: *Robbins and Cotran pathologic basis of disease*, ed 9, Philadelphia, 2015, Saunders, pp 69–111.

Lydyard PM: Porakishvili N: Cells, tissues and organs of the immune system. In Male D, Brostoff J, Roth D, Roitt I, editors: *Immunology*, ed 7, St Louis, 2006, Mosby, pp 15–45.

McNab R, Mayer-Barber K, Sher A, et al: Type I interferons in infections disease. *Nat Rev Immunol* 15(2):87–103, 2015.

Patton KT, Thibodeau GA: *Anatomy & physiology*, ed 8, St Louis, 2013, Mosby.

10

Alterations in Immune Function

Faith Young Peterson

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- What are the potential mechanisms whereby erroneous reaction of the immune system with self tissue leads to autoimmune diseases?
- How do type I, II, III, and IV hypersensitivity reactions differ according to the immune cell types involved and the mechanism of tissue injury?
- What are the common features of autoimmune disorders and certain types of hypersensitivity disorders?
- How are hypersensitivity disorders detected, prevented, and treated?
- How do the etiologic processes of primary and secondary immune deficiency disorders differ?
- What are the clinical features of the common immunodeficiency disorders?

CHAPTER OUTLINE

EXCESSIVE IMMUNE RESPONSES, 195

Autoimmunity, 195
- Genetic Factors, 196
- Environmental Triggers, 196
- Pharmacotherapies, 197

Hypersensitivity, 198
- Type I Hypersensitivity, 199
  - *Etiology, 199*
  - *Pathogenesis, 199*
  - *Clinical Manifestations, 200*
  - *Treatment, 200*
  - *Prevention, 201*
  - *Pharmacotherapeutic Prevention, 201*
- Type IIa Hypersensitivity, 201
  - *Etiology and Pathogenesis, 201*
  - *Transfusion Reaction, 202*
  - *Hemolytic Disease of the Newborn, 203*
  - *Myasthenia Gravis, 204*
  - *Hyperacute Graft Rejection, 204*
- Type IIb Hypersensitivity, 204
  - *Graves Disease, 204*
- Type III Hypersensitivity, 204
  - *Etiology, 204*
  - *Pathogenesis, 205*
  - *Tissue Deposition, 206*
  - *Immune Complex Glomerulonephritis, 206*
  - *Systemic Lupus Erythematosus, 206*
- Type IV Hypersensitivity, 207
  - *Type IVa – Granulomatous Hypersensitivity, 207*
  - *Type IVa – Tuberculin-Type Hypersensitivity, 207*
  - *Type IVa – Allergic Contact Dermatitis, 207*
  - *Type IVb – Persistent Asthma, 209*
  - *Type IVc – Stevens–Johnson Syndrome and Toxic Epidermal Necrolysis, 209*
  - *Type IVd – Pustular Psoriasis, 209*

DEFICIENT IMMUNE RESPONSES, 210

Primary Immunodeficiency Disorders, 210
- B-Cell and T-Cell Combined Disorders, 210
  - *Severe Combined Immunodeficiency Disorders, 210*
  - *Wiskott–Aldrich Syndrome, 211*
- T-Cell Disorders, 211
  - *22q11.2 Deletion Syndrome (DiGeorge Syndrome), 211*
  - *Chronic Mucocutaneous Candidiasis Disease, 212*
- B-Cell Disorders, 212
  - *IgA Deficiency, 212*
  - *X-Linked Agammaglobulinemia, 212*
  - *Transient Hypogammaglobulinemia, 212*
  - *Common Variable Immunodeficiency Disease, 212*

Secondary Immunodeficiency Disorders, 212

The purpose of the immune system is to defend the body against invasion or infection by foreign substances, called *antigens,* and to patrol for and destroy cells that are abnormal or damaged. Normally, the immune system works efficiently to accomplish these purposes, but in some situations inappropriate immune responses lead to disease.

These disorders can be divided into two general categories: (1) excessive immune responses, and (2) deficient immune responses. The category of excessive immune responses includes disorders in which the immune system is overfunctioning or hyperfunctioning, such as autoimmunity and hypersensitivity disorders. Autoimmune diseases are complex, multifactoral chronic diseases that occur in response to polygenetic as well as environmental factors. The estimated overall prevalence ranges from 3% to 10% of the general population.

The category of deficient immune responses includes disorders in which the immune system fails to respond to provide protection and is ineffective because of disease-causing genotypes or secondary/acquired dysfunction. Examples of deficient immune responses are severe combined immunodeficiency (SCID) syndrome, DiGeorge syndrome, and selective immunoglobulin A (IgA) deficiency. Human immunodeficiency virus/acquired immunodeficiency syndrome (HIV/AIDS) is a primary acquired immunodeficiency disorder that is discussed in Chapter 12. The secondary immunodeficiencies associated with white blood cell malignancies are included and discussed in Chapter 11.

EXCESSIVE IMMUNE RESPONSES

Excessive immune response disorders result from a functional increase in the activity of the immune system involving multiple interacting immune cells. Autoimmunity and hypersensitivity are types of excessive immune response disorders that are often related, and both may be present in patients. It may be helpful to think of autoimmunity as a way of describing the *etiologic process,* or cause, of abnormal excessive immune responses toward self tissues. Hypersensitivity disorders describe *mechanisms* of injury, or how the injury occurs, which may or may not involve autoimmunity. Autoimmunity is a general term that is used when the immune system attacks its own tissues. Most autoimmune reactions toward self tissues are mediated through type II (cytotoxic) and III (immune complex) hypersensitivity mechanisms. For this reason, many autoimmune diseases also are considered hypersensitivity reactions. For example, myasthenia gravis is both an autoimmune disease and a type II hypersensitivity reaction. Immune complex glomerulonephritis is both an autoimmune disease and a type III hypersensitivity reaction. When hypersensitivity reactions occur in response to foreign antigens, such as bee venom, they are not autoimmune.

The causes of immune system overreactions involve a complex interplay between genetic factors, including major histocompatibility complex (MHC) genes, and environmental factors thought to be important in the development of autoimmune disorders. Autoimmune disorders range from organ-specific diseases to organ-nonspecific diseases. Examples of organ-specific diseases include multiple sclerosis, where the target organ is the central nervous system, or in type 1 diabetes, where the target organ is pancreatic beta-islet cells. Examples of organ-nonspecific diseases include systemic lupus erythematosus (SLE) or Sjögren syndrome where multiple organs are involved.

AUTOIMMUNITY

Autoimmunity occurs when the immune system recognizes a person's own cells ("self") as foreign and mounts an immune response that injures self tissues. It is a failure of self-tolerance. The adaptive immune response is the responsibility of antigen-specific T cells and B cells that learn to identify "self" in the thymus and bone marrow. Self-tolerance is a complex interaction that includes central tolerance in the thymus and bone marrow and peripheral tolerance involving T regulatory cells, peripheral anergy, and homeostasis produced by cytokines, chemokines, and their receptors. The thymus gland is principally responsible for eliminating or suppressing aggressive or intolerant self-reactive (autoreactive) lymphocytes (see Chapter 9). However, a small number of T cells escape thymic control and move into the peripheral circulation. The immune system has other peripheral "checkpoints" to detect, limit, and control these "self" or "auto" reactive T cells that have escaped the thymus gland. Peripheral tolerance involves T-cell intrinsic mechanisms that lead to clonal deletion, or anergy, as well as extrinsic control by T suppressor cells, primarily $CD4^+CD25^+FOXp3^+$Tregs. T suppressor cells regulate potentially harmful autoreactive T and B cells. Autoimmune diseases result when self-tolerance is lost and reactions between self antigens and the immune system occur, causing dysregulation of proinflammatory and antiinflammatory mediators, cytokines, $CD4^+$ and $CD8^+$ T cells, B cells, and ubiquitin-editing enzyme A20 in dendritic cells.

Multiple theories have been proposed to explain how various immune system components and environmental triggers might interact to produce autoimmunity. However, no single theory can fully explain the loss of self-tolerance that occurs in autoimmune diseases. A number of genetic, cellular, and environmental factors interacting together contribute to the development of autoimmunity. The triggers for autoimmunity are not known exactly because autoimmune diseases are mediated by a variety of mechanisms, biochemical and cellular events, and responses to those events. Newer theories emphasize the relationship between pathogens and both the innate immune system and the adaptive immune system.

The theory of antigenic or molecular mimicry emphasizes the similarities between certain molecular segments of foreign antigens called *epitopes* and the person's own cells. For example, all cells, whether self or foreign, are composed of proteins, carbohydrates, nucleic acids, and lipids. Certain viruses and bacteria evolve to look like "self" and use "molecular mimicry" to slip past the immune system defenses. Self cells with the same or similar molecular segments as these foreign epitopes can "fit" lymphocyte receptors. Therefore these self antigens or autoantigens can be attacked as foreign under certain circumstances when the normal cell has been altered, such as by a bacterial or viral infection that stimulates the immune response. The persistent presence of these autoantigens then acts as a constant source of stimulation to the immune system. This theory is suggested as a cause of rheumatic heart disease attributable to cross-reactions between *Streptococcal* antigenic protein M and human protein lysoganglioside leading to the development of cardiac-reactive T cells and autoimmunity. Another example is the interaction between Epstein–Barr virus (EBV) peptide with a "self" protein (PPPGMRPP peptide of Sm) that occurs in patients with SLE. In both cases, autoantigens have been identified following exposure to bacterial or viral infection.

Another theory proposes that activation and release of preexisting autoreactive immune cells (or sequestered antigens) trigger the autoimmune response. This theory of hidden/cryptic antigens suggests that certain self antigens are isolated from the immune system within an organ during the neonatal period. They are not in contact with antigen-processing cells during the embryonic period when self-tolerance usually occurs. These hidden self antigens or sequestered proteins, normally sheltered from immune recognition, occur in sites such as the cornea of the eye, the testicles, brain, or other areas not drained by lymphatics. If and when these sites are damaged later in life, the hidden or sequestered proteins are exposed to the immune system, which does not recognize them as self. Therefore the damaged cells are attacked. In this theory, bacterial or viral infections cause the tissue damage that leads to exposure

of the "hidden" self antigens. However, this theory is now in question because it states that microbes and tissue damage are needed to cause autoimmunity. For example, people undergoing heart surgery and transplantation all produce autoantibodies to heart antigens, but they do not develop myocarditis, an autoimmune cardioinflammatory disease.

A similar theory is called the *adjuvant* or *bystander effect*. This theory states that an infection, which stimulates the immune system, activates receptors on immune cells causing the release of proinflammatory cytokines. These cytokines then activate preexisting autoreactive T and B cells that previously escaped thymic detection and deletion. It is unclear if this is a nonspecific (any immune cell) reaction or an innate immune-specific reaction, such as against a specific toll-like receptor.

A number of T -cell theories of autoimmunity have been proposed, including thymus gland defects, decreased suppressor T-cell function, and altered T helper cells. The epitope spread theory states that recurrent infection or damage to a particular organ causes the immune system to develop multiple autoantibodies. Theories attributing autoimmunity to thymus gland defects state that maturation and differentiation of T cells are affected either by decreased hormone secretion or by failure of the thymus to expose T cells to all self products. The thymus gland is responsible for exposing developing T cells to self products produced in the thymus or carried to the thymus gland. If some self products are not exposed to the developing T cells, the product will not be recognized as self and will subsequently be attacked. However, recent evidence demonstrates that not all T cells maturate in the thymus. Thus not all T cells may experience or "learn" to recognize "self." This lack of exposure to self products is thought to be a major factor in the development of generalized autoimmune diseases.

The theory attributing autoimmunity to decreased or lost regulatory/suppressor T-cell (Treg) activity states that decreased numbers of Treg cells fail to repress immunoglobulin activity. Some data suggest that Treg cells lose their forkhead box p3(FOXP3) protein and are transformed into effector T cells. It is also unclear if this disruption in the number of, or activity of, T cells may also be the result of genetic mutations. For example, a mutation of the transcription factor (FOXP3) causes impaired development and function of $CD4^+$ T cells, which can lead to autoimmune inflammation. T helper cells 22 (Th22 cells) are a newly identified subset of T cells that are important mediators of chronic inflammation and autoimmunity by inducing the production of inflammatory cytokines (ll17a, ll17f, ll22, ll26) by $CCR6^+$ cells. Additionally, the development of autoimmunity may be secondary to an imbalance or alteration of autoreactive effector T cells and their ability to be controlled by Tregs. Decreases in peripheral Tregs (pTreg) and/or thymus Tregs (tTreg) can lead to autoimmunity, as both are needed to prevent it in certain inflammatory conditions. Current thinking suggests that both are needed to prevent autoimmunity in peripheral tissues of healthy people.

A number of B-cell theories of autoimmunity also have been proposed. The theory attributing autoimmunity to escape of B-cell tolerance proposes that certain B cells lose their responsiveness to suppressor T-cell messages. The B-cell activation theories, which are well supported clinically and experimentally, suggest that extrinsic factors or intrinsic genetic B-cell defects cause autoantibody production and an increase in the number and activity of B cells. A number of extrinsic factors, including viruses, bacteria, antibiotics, proteolytic enzymes, and lipopolysaccharides, have been found to be B-cell–activating factors that could trigger autoantibody production.

Research has linked mast cells to autoimmunity as well as to hypersensitivity. Mast cells reside in most mucous membranes waiting for foreign proteins or bacteria to invade. Their cell membranes are studded with bacteria-sensing proteins, called *toll-like receptors,* which cause release of up to 10,000 different chemicals in response to activation. Mast cells release the cytokine interleukin-1 (IL-1), attracting and enlisting inflammation-inducing cells to joints and leaking fluid into joints in the autoimmune disease of rheumatoid arthritis. They are also thought to be involved in the development of irritable bowel syndrome and other functional gastrointestinal disorders.

Genetic Factors

Genetic factors play a role in the development of autoimmune disorders. Gender, which is genetically determined, also influences the expression of autoimmune disorders. The exact mechanisms of gender and genetic influence on autoimmune expression have not been established, but the relationship is significant. Females are at significantly higher risk for developing autoimmunity compared with males. Different cytokine profiles can be associated with autoimmunity. Those with genetically low levels of tumor necrosis factor-α (TNF-α) and high levels of IL-10 may be more tolerant than those with normal levels.

The role of genetics is also supported by the observation that certain human MHC genes located on chromosome 6p21 (also called *human leukocyte antigen [HLA] genes*) encode molecules involved in antigen presentation and are critical in distinguishing self from nonself. The HLA genes are frequently associated with autoimmune disorders (Table 10.1). The MHC gene region demonstrates a high level of polymorphism. They are associated with the highest genetic risk in many autoimmune diseases.

One of the strongest correlations of MHC molecules with autoimmune disease is the linkage between the HLA-B27 phenotype and ankylosing spondylitis. In this case 95% of all people with ankylosing spondylitis have a positive B27 phenotype. However, not everyone with a positive B27 phenotype develops ankylosing spondylitis, both because of differences in the way antigen is presented to the immune system and because of environmental factors. Other diseases are associated with different MHC phenotypes, but the correlation between risk for disease and presence of the disease marker is much lower. For example, Addison disease is associated with the HLA-DR3 phenotype, but it has only a 6% risk correlation. Juvenile rheumatoid arthritis is strongly associated with HLA-DR5.

There may also be disease-causing genotypes or genetic factors affecting expression of immune factors. For example, the lymphoid protein tyrosine phosphatase nonreceptor type 22 (PTNP22) gene is found in patients with many autoimmune disorders, including type 1 diabetes, rheumatoid arthritis, SLE, Grave disease, and Crohn disease. TNF-α is involved in acute and chronic inflammation, autoimmunity, and malignancies. Of special interest is the *FOXP3* gene, which is expressed by $CD4^+$ regulatory T cells. *FOXP3* deficiency is associated with both primary immune deficiency disorders and autoimmune disorders owing to its effect on regulatory T-cell immune function.

Environmental Triggers

Chronic or multiple viral or bacterial infections may trigger the development of autoimmune disease in susceptible persons. For example, acute rheumatic fever occurs in people who are genetically susceptible to *Streptococcus pyogenes* through molecular mimicry. Viruses can activate B cells, decrease the function of T cells, contribute to the development of antigenic mimicry, or insert viral components on cell surfaces and trigger immune reactions. For example, viruses (EBV, cytomegalovirus [CMV]), bacteria (gram positive and negative such as *Streptococcus, Campylobacter jejuni,* and *Helicobacter pylori*), parasites (trypanosomes and *Ascaridia galli*), and fungi *(Saccharomyces cerevisiae)* have been cited as potential triggers of autoimmune disease.

A variety of noninfectious environmental factors have been associated with autoimmunity such as foods, pharmaceutical agents, smoke, ultraviolet light, vitamin D, hormones, vaccines, and heavy metals. A

TABLE 10.1 Autoimmune Diseases, Major Histocompatibility Genes, Auto-Antibodies, and Target Organ

Disease	Age at Onset and Gender Ratio (F/M)	HLA (MHC) Antigen	Auto-Antibodies	Target Organ
Multiple Sclerosis	20–40 2/1	HLA Class I: HLA-A*02:01 HLA Class II: DRB1 …	Antibodies against myelin protein	Central nervous system (CNS)
Rheumatoid arthritis	44–55 2/1	HLA Class II: DR4 DR1 HLA Class III: TNF	Anti-CCP, RF-IgG, ACPA, anti-Carp	Synovium of joints
Graves disease	50–60 5/1	HLA Class I: HLA-B8 HLA Class II: DR3 and DR4	Anti-TSHR auto-antibodies	Thyroid
Type 1 diabetes mellitus	6–13 1/1	HLA Class I: HLA-A and B HLA Class II: DQ2 and DQ8	IAA, GADA, IA-2A, ZnT8A	Pancreas β-islets
Systemic lupus erythematosus	30–50 9/1	HLA Class II: DR3 DR2 DR8 DR6 HLA Class III: TNF C2, C4, ……	Antinuclear antibody, antidsDNA antibody, anti-Sm, antiphospholipid antibody	Multiple organs: heart, joints, skin, lungs, blood vessels, liver, kidneys, CNS
Primary biliary cirrhosis	50–60 10/1	HLA class II: DRB1*08, DRBI*11, DRBI*13 protective	AMA	Small and medium-sized intrahepatic bile ducts
Sjögren syndrome	40–50 9/1	HLA Class II: DRB1 ……	Anti-Ro/SSA, anti-La/SSB, ANA	Mainly salivary and lacrimal glands; other organs: lungs, liver, kidneys, CNS
Crohn disease	15–30 and 60–80 1/1.2	HLA Class II: DR7, DRB3, DR2, DR3	Anti-TG2, antigliadin	Gastrointestinal tract
Addison disease	15–45 0.8–2.4/1	HLA Class II: DR3/DQ2 DR4.4/DQ8 DQB1	ACA	Adrenal glands

Data from Wang L et al: Human autoimmune diseases: a comprehensive update, *J Intern Med* 278(4):369–395, October 2015.

good example is the relationship between gluten ingestion and the development of celiac disease. Another example is the exposure to ultraviolet light that leads to oxidative damage to DNA and RNA, induces cell death, and promotes the release of cytokines TLR7 and TLR9. This can induce the development of SLE, rheumatoid arthritis, or multiple sclerosis in susceptible individuals.

Environmental stress and occupational stress can affect the immune system because of their relationship to neuroendocrine system hormones, leading to inflammation or lymphokine release that activates T cells. Neuroendocrine and immune system interaction during life stress, such as shift work or workplace stress, promotes the synthesis and overproduction of proinflammatory cytokines. In genetically susceptible persons, this increase in systemic and local proinflammatory cytokines may affect the system's balance enough to trigger autoimmune disease. There are also other linkages between the neuroendocrine and immune systems.

Although the etiology of autoimmunity continues to be investigated, the mechanisms whereby autoantibodies injure tissues are better understood. The autoantibodies produced by autoimmune disorders affect tissue by the mechanisms described for type II and type III hypersensitivity reactions found later in this chapter.

Pharmacotherapies

The ideal therapeutic approach to treat autoimmune disease would be a medication that inhibits abnormal immune responses without limiting the positive and protective functions of the immune system or causing any organ toxicity. Unfortunately, no medication with these specific properties yet exists.

Immunosuppressive therapy is a common treatment for autoimmune disease. Because autoimmunity is expressed in different ways, the immunosuppressive treatment for each type of autoimmune disease is individualized, depending on disease expression. Immunosuppressive

therapy has become an increasingly important treatment choice. These medications are essential for inhibiting excessive or aberrant immune responses, but must be taken continuously to achieve the goal of inhibiting the immune system response to self antigens.

Immunosuppressive agents include corticosteroids, TNF inhibitors, immunomodulators, and cytotoxins. Corticosteroids (cortisone, prednisone, or budesonide) decrease the number of lymphocytes and decrease antibody formation, as well as alter the functional activities of lymphocytes. They also have many other activities as a result of their glucocorticoid function. Corticosteroids tend to be used in the treatment of many autoimmune diseases and are the oldest of the immunosuppressive drugs. The adverse effects that occur during corticosteroid use often limit their extended use over time. The common side effects include hypercorticism with changes in fat distribution and buffalo hump formation, suppression of the hypothalamic–pituitary–adrenal (HPA) axis, congestive heart failure, hypertension, emotional changes, thinning of skin, petechiae, diabetes mellitus, menstrual irregularities, electrolyte imbalances, liver and pancreatic dysfunction, exophthalmos, glaucoma, osteoporosis, loss of muscle mass, and muscle weakness. Because corticosteroids affect the HPA system, gradual withdrawal by tapering the dose over time is necessary when discontinuing long-term therapy.

Cytotoxins, such as methotrexate, are used to manage autoimmune disorders because of their ability to inhibit folic acid metabolism and kill actively proliferating lymphocytes after they are transformed from their resting G_0 state. They also selectively down-regulate B-cell function. The key to the use of cytotoxins is to effectively apply their killing activity without damaging the rest of the body. Cyclosporine (Sandimmune) is a more selective immunosuppressant that reversibly suppresses T helper cells in the G_0 or G_1 phase of the cell cycle without killing them. As a result, it inhibits the development of killer or cytotoxic T cells without decreasing the numbers of cells. It also impairs the ability of T cells to respond effectively to foreign antigens. It is used to suppress reactions during tissue or organ transplantation. Frequent side effects include predisposition to infection, pulmonary fibrosis, edema, hypertension, headache, hirsutism, elevated triglyceride levels, gastrointestinal effects, nephropathy, emotional changes, skin pigmentation, gynecomastia, leukopenia, anemia, and hepatotoxicity.

TNF inhibitors or immunomodulators, such as etanercept (Enbrel), adalimumab (Humira), golimumab (Simponi), or infliximab (Remicade), are used as disease modifiers that bind to and block the activity of TNF-α and TNF-β. They may also modulate TNF-mediated responses, such as leukocyte migration and expression of adhesion molecules. The most common side effects of TNF inhibitors include predisposition to infections, headache, gastrointestinal changes, injection-site skin reactions or infusion reactions, respiratory tract infections, edema, dizziness, dyspepsia, weakness, malignancies, and neurologic disorders.

Purine analogs such as azathioprine inhibit the synthesis of DNA and are used in multiple autoimmune diseases. Side effects include increased risk of malignancy, bone marrow suppression, nausea, and vomiting. There are also new medications, such as tocilizumab, an IL-6 inhibitor; abatacept, a CTLA4-immunoglobulin; alemtuzumab, an anti-CD52 medication; and rituximab, an anti-CD20 antibody, that target specific immune system components or cells. Other new therapies are being developed that target Treg defects in order to restore their function. These new medications are expected to be safer than older, nonspecific therapies. Novel approaches are also being developed such as transplantation of pancreatic B cells for type 1 diabetes along with thymus tissue to ensure graft tolerance without long-term immunosuppression.

Therapeutic plasmapheresis is another type of therapy occasionally used in the management of autoimmune diseases. Plasmapheresis is analogous to dialysis and involves the selective filtering or removal of plasma or a plasma cell type as well as protein-bound toxic substances. The patient's whole blood is filtered, blood cells and platelets are returned, and the plasma component containing the autoantibodies is removed and replaced with 5% albumin or another colloid solution. According to the American Society for Apheresis, this type of therapy has been effective in the management of diseases such as myasthenia gravis, thrombocytopenia purpura, multiple sclerosis, and Rh-negative hemolytic disease of the newborn. Plasmapheresis is generally well tolerated; however, there are both major and minor risks involved in the process. Some of the risks include insertion of large intravenous (IV) catheters to perform the procedure, decrease in serum ionized calcium concentration, shifts of fluid levels, and risks of infection and bleeding from loss of coagulation factors.

KEY POINTS

- Autoimmune disorders occur when the immune system erroneously reacts with "self" tissues. These disorders are thought to be polygenic and multifactorial; however, the exact etiologic process is unknown.
- The antigenic mimicry theory involves the alteration of viruses or bacteria to look like "self" and the precipitation of immune reactions.
- The theory involving release of sequestered antigens suggests that self antigens that do not come in direct contact with lymphocytes during fetal development may cause autoimmune reactions if they are subsequently released from sequestration.
- Abnormal production of subclasses of T lymphocytes, particularly suppressor T cells, has been proposed as a reason for the development of autoimmunity, as well as the development of abnormal B cells that do not respond to suppressor T-cell signals.
- Genetic factors such as female gender and major histocompatibility complex (MHC) genes are associated with certain autoimmune disorders.
- Autoantibodies injure body tissues through the mechanisms described for type II and type III hypersensitivity reactions.

HYPERSENSITIVITY

Hypersensitivity is a normal immune response that is inappropriately triggered, excessive, or produces undesirable effects on the body. The basic mechanism that triggers hypersensitivity is a specific antigen–antibody reaction or a specific antigen–lymphocyte interaction. Four classes or types of hypersensitivity are differentiated: types I, II, III, and IV. Each type is characterized by a specific cellular or antibody response. Types I, II and III are mediated by antibodies. Hypersensitivity type I is an immediate or IgE-mediated reaction, such as anaphylaxis. Hypersensitivity type IIa is cytotoxic or IgG/IgM mediated, and type IIb is antibody-mediated cell stimulating. Hypersensitivity type III reaction is IgG/IgM immune complex mediated such as immune complex glomerulonephritis and SLE. Type IV hypersensitivity is T cell mediated and has four major categories. Type IVa is mediated by $CD4^+$ Th1 cells with activation of macrophages. Type IVb is $CD4^+$ Th2 lymphocyte mediated with eosinophilic involvement, such as persistent asthma. Type IVc is mediated by cytotoxic $CD8^+$ T lymphocytes with involvement of perforin-granzyme B in apoptosis, as seen in Stevens–Johnson syndrome (SJS). Type IVd is T-lymphocyte–driven neutrophilic inflammation.

Hypersensitivity reactions are specific to a particular antigen and usually do not occur on first exposure to the antigen. Although the diseases or syndromes associated with each type differ in their clinical signs and symptoms, the underlying pathophysiologic process is similar within each type. The four major types of hypersensitivity are contrasted in Table 10.2. The complex interactions between immune system inflammatory mediators, cytokines, T cells, B cells, and mast cells

TABLE 10.2 The Four Types of Hypersensitivity

Characteristic	Type I: Atopic, Anaphylactic	Type II: Cytotoxic, Cytolytic	Type III: Immune Complex (Arthus Reaction)	Type IV: Delayed Hypersensitivity
Mediated by:	IgE	IgM or IgG	IgG and complement	$CD4^+$ Th1 or Th2 cells, $CD8^+$ cells
Complement activation	No	Yes	Yes	No
Immune response	Ag plus IgE, leading to mast cell degranulation	Surface Ag and Ab, leading to killer cell cytotoxic action or complement-mediated lysis	Ag–Ab complex deposited in tissues; complement activated and PMNs attracted	Ag-sensitized T cells release cytokines, leading to inflammatory reaction and recruitment of macrophages, eosinophils, or neutrophils, which release cytokines
Peak action	15–30 minutes	15–30 minutes or can occur over time	6 hours or more	24–48 hours or more
Serum transferability	Yes	Yes	Yes	No
Cell transferability	No	No	No	Yes (T cells)
Genetic mechanisms	Familial High IgE level HLA-linked Ir genes General hyperresponsiveness	HLA linked in some cases	Familial (autoimmune) HLA specificities	Unknown
Causes of reaction	T-cell deficiency Abnormal mediator feedback Environmental factors and Ag exposure	Exposure to Ag or foreign tissue, cells, or graft	Persistent infection—microbe Ag Extrinsic environmental Ag Autoimmunity—self Ag	Intradermal Ag Epidermal Ag Dermal Ag
Manifestation (examples)	Bee sting reaction, allergic reactions, rhinitis	ABO transfusions, hemolytic disease of newborn, myasthenia gravis Graves disease	Immune-complex glomerulonephritis, SLE, farmer's lung arthritis, vasculitis	Allergic contact dermatitis, Stevens–Johnson syndrome, toxic epidermal necrolysis, pustular psoriasis

Ab, Antibody; *Ag,* antigen; *DTH,* delayed-type hypersensitivity; *Ig,* immunoglobulin; *HLA,* human leukocyte antigen; *PMN,* polymorphonuclear leukocyte; *SLE,* systemic lupus erythematosus.

characterize hypersensitivity reactions. Some diseases have multiple types of immunologic hypersensitivity.

Type I Hypersensitivity

Etiology

Genetic mechanisms influence type I hypersensitivity with strong genetic or hereditary linkage regarding the IgE response to antigens (allergens). This genetic component involves both the ability to respond to an allergen and the general ability to produce an IgE antibody response. Genome-wide association studies have identified hundreds of genetic variants and have contributed to a greater understanding of type I hypersensitivity/allergic diseases. However, the specific causal genetic mechanisms are still unclear. A positive family history remains a strong predictor of allergic disease.

Pathogenesis

Type I hypersensitivity is also known as *immediate hypersensitivity,* because the reaction is immediate. It is a sensitization reaction characterized by signs and symptoms of an allergic reaction that usually occurs 15 to 30 minutes after exposure to an antigen (allergen).

At the cellular level, immunoglobulin E (IgE) is the principal antibody mediating this reaction. IgE is produced by specialized plasma B cells and circulates in very small amounts in the blood. When an individual is exposed to an allergen, selected plasma B cells produce allergen-specific IgE. It usually takes repeated exposures to the allergen to cause significant levels of IgE to be present in the blood. Environmental pollutants may play a role by increasing mucosal permeability and enhancing antigen (allergen) entry into the body. This increased entry would subsequently increase IgE responsiveness.

Mast cells and basophils are the principal effector cells, although many other cells with histamine and other inflammatory mediators can be involved in the reaction. These may include neutrophils, eosinophils, lymphocytes, macrophages, epithelial cells, and endothelial cells. Mast cells are found throughout the body in all loose connective tissue. They are covered with IgE receptors—up to 500,000 on their cell surfaces—and they are filled with vesicles or granules containing potent vasoactive, proinflammatory chemical mediators (especially histamine) that produce inflammation when they are released. The IgE receptors on mast cells bind the Fc portion of an IgE antibody. The IgE antigen-binding sites are then displayed on the mast cell surface, where they can bind to antigens that pass by the mast cell (Fig. 10.1). This process makes the mast cells responsive to particular antigens.

The initial incident during a type I hypersensitivity response is the cross-linking of two IgE receptors to one antigen on the mast cell located at the site of the allergen's entry into the body (see Fig. 10.1). Cross-linking of IgE and the antigen causes an increase in intracellular calcium (Ca^{2+}) concentration that results in immediate, massive, local mast cell degranulation of preformed and newly formed proinflammatory mediators. The release of mediators causes an inflammatory response.

Mast cells, basophils, and other effector cells release many chemicals that act to trigger evolving immune responses, promote continuing sensitivity, and promote the transition of allergic disease into chronic illness. Some of the mediators are preformed and stored in vesicles, such as histamine, heparin, proteolytic enzymes, cytokines, and

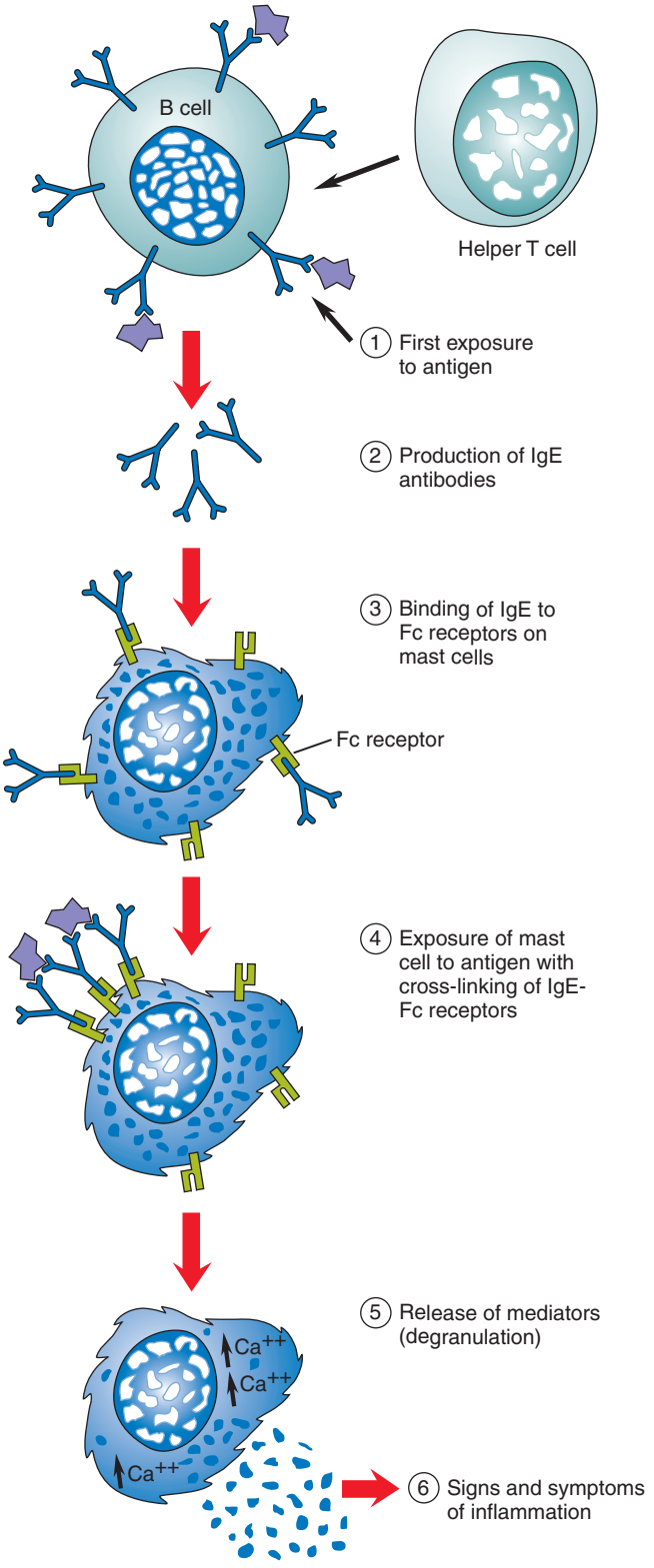


FIG 10.1 Type I hypersensitivity reaction.

chemotactic factors. Other mediators are formed during the degranulation process. Examples of newly formed mediators include superoxide, prostaglandins, thromboxanes, leukotrienes, bradykinin, and interleukins (see Chapter 9).

One of the most important mediators of type I hypersensitivity is histamine. Histamine binds to H_1 (histamine 1), H_2, H_3, and H_4 receptors, which are located on many types of cells. Mast cells have receptors for H_1, H_2, and H_3, with H_1 receptors being the most active. Basophils express predominantly H_2 receptors, whereas neutrophils and eosinophils have both H_1 and H_2 receptors. Recent evidence shows that H_1 and H_2 receptors are present on monocytes and macrophages, with an increase in H_1 receptors when monocytes differentiate into macrophages. Histamine binding to H_1 receptors triggers increased vascular permeability, vasodilation (flushing), urticaria formation (hives), smooth muscle constriction (bronchoconstriction), increased mucus secretion and pruritus (increased itching), and increased gut permeability. The activation of H_2 receptors has opposing effects to H_1 receptors in some tissues and causes smooth muscle relaxation in the lower airways, augments gastric acid secretion from parietal cells, and in high concentrations has an inhibitory effect on inflammatory cells, decreasing both degranulation and neutrophil chemotaxis. The H_3 receptors are located in the brain, in the spinal cord, and on sensory neurons such as postganglionic cholinergic nerves in lung bronchi. The H_4 receptors are found on immune system cells such as dendritic cells, eosinophils, T cells, monocytes, macrophages, and natural killer cells, as well as in hematopoietic cells such as the spleen, thymus, bone marrow, and blood leukocytes. H_4 receptors are highly attracted to histamine and are also involved in chemotaxis and inflammatory responses.

The proteolytic enzymes kininogenase and tryptase activate the kinin pathway, and C3 activates the complement cascade via the alternative pathway. Heparin decreases clot formation. The chemotactic factors recruit or activate other inflammatory and immune cells. Leukotrienes cause smooth muscle contraction and increase vascular permeability. Increased intestinal expression of IL-9 by mucosal mast cells is found in patients with food-mediated hypersensitivity.

Clinical Manifestations

Manifestations of an immediate hypersensitivity reaction vary in severity and intensity. For many people, type I hypersensitivity reactions are annoying, such as hives (urticaria), seasonal allergic rhinitis, eczema, or mild bronchoconstriction. In other people, the symptoms are more problematic, including tightening of the throat, localized edema, wheezing, and tachycardia, such as is associated with localized angioedema reactions or severe airway reactions. In a very small number of highly allergic people, the type I hypersensitivity reaction can be expressed as a life-threatening allergic reaction known as *anaphylaxis.* Anaphylaxis is often caused by bee stings, seafood, or peanut ingestion. Common allergens that can trigger type I hypersensitivity reactions are listed in Box 10.1.

Treatment

Treatment for type I hypersensitivity primarily involves pharmacologic management with antihistamines, β-adrenergics, corticosteroids, anticholinergics, and anti–immunoglobulin E therapy (IgE blocker therapy). Antihistamines such as diphenhydramine (Benadryl) are used to block the effect of histamine. This action decreases vascular permeability and bronchoconstriction. β-Adrenergic sympathomimetics are used to decrease bronchoconstriction and bronchospasm. Epinephrine is an adrenergic agent (α, β_1, and β_2) given subcutaneously or intravenously during acute allergic reactions, especially after medication, food, or bee sting reactions. Most patients with severe allergies to food or insect bites are given prescriptions for epinephrine in the form of EpiPen with an autoinjector. Corticosteroids are used to decrease the inflammatory response. Anticholinergics are used to block the parasympathetic system and thus allow greater sympathetic activity. This action indirectly causes bronchodilation. Anti–immunoglobulin E therapy (omalizumab) may be used for persons with severe persistent asthma. Omalizumab (Xolair) is a subcutaneously injected, monoclonal anti-IgE antibody

BOX 10.1 Possible Causes of Human Allergic Reactions or Anaphylaxis

Medications
Penicillin, penicillin analogs, and other antibiotics
Radiographic contrast media
Aspirin, indomethacin, and other nonsteroidal antiinflammatory drugs (NSAIDs)
Anesthetic agents

Biological Agents
Allergenic extracts and antisera
Serum proteins, including γ-globulin and monoclonal antibodies
Insulin and other hormones
Vaccines
Enzymes such as penicillinase
Blood products
Latex
Pollen from grass, trees, flowers, and weeds
Mold spores
Pet dander
Dust mites
Cockroach saliva, feces, or shedding body parts

Insect Venom
Hymenoptera (stinging) insects
Polistes wasps
Honeybees
Fire ants
Hornets and yellow jackets

Foods
Peanuts
Tree nuts
Seafood, especially shellfish
Eggs
Fruit, especially citrus and strawberries
Milk
Tartrazine, yellow dye no. 5
Wheat
Soy
Sesame seeds, mustard seeds, or other seeds

Inorganic Chemicals
Nickel
Aluminum
Zinc

Irritants
Cigarette smoke
Perfumes
Diesel exhaust
High levels of air pollution

that binds to the IgE molecule, thus inhibiting the binding of IgE to mast cells and basophils and promoting down-regulation of IgE receptors. It is used to improve asthma control in patients with moderate-to-severe persistent asthma not controlled with inhaled corticosteroids. It is used in children 12 years or older and in adults.

Prevention

Some protective, proactive actions taken during pregnancy are thought to decrease the likelihood that type I hypersensitivity will develop in children from families with a history of allergies. These actions include avoiding foods to which the mother is allergic, limiting excesses of one type of food during the last trimester of pregnancy, avoiding whole eggs during the last month before delivery and while breast feeding, and limiting cow's milk to two glasses per day. Other actions that may be helpful during the child's infancy include avoiding exposure to environmental pollution, breast feeding for a minimum of 6 months, supplementing the child's diet with non–cow's milk products such as soy milk, giving solid foods only after the infant is 6 months old, keeping the infant's room as free of dust and molds as possible, and keeping pets (dogs, cats, birds) out of the home.

Pharmacotherapeutic Prevention

Another avenue for prevention of type I hypersensitivity reactions involves the use of desensitization therapy (immunotherapy). Desensitization is more successful in patients with hay fever than in those with other types of allergies. It involves both environmental control of external allergens and titrated pharmacologic exposure to allergens. Pharmacologic immunotherapy is particularly effective when exposure to the allergen can't be avoided (i.e., pollens, molds, insect venoms, and dust mites). Environmental control involves a systematic plan to decrease exposure to house dust, molds, and animal dander. Pets are kept out of the house. The person must avoid food allergens, wool carpets, goose down or feather pillows, dried plants, and exposure to other animal and vegetable products. The person is urged to use air conditioning and electronic air filters.

Pharmacologic immunotherapy or desensitization involves giving a person an antigen (allergen) in gradually increasing dosages weekly or biweekly over a course of months or years in order to induce tolerance. The antigen (allergen) can be given sublingually or by subcutaneous injection. After the initial series, periodic maintenance or booster therapy is used to maintain tolerance. Gradually, the dose is increased until the person can tolerate the allergen without a type I hypersensitivity reaction. The goal of this therapy is a change in immunoglobulins so that there is an increase in IgG- and IgA-blocking antibodies, no increase in IgE during allergy season, decreased basophil reactivity, and decreased lymphocyte reactivity to allergens.

Type IIa Hypersensitivity

Etiology and Pathogenesis

Type IIa hypersensitivity, also known as *tissue-specific,* or *cytotoxic,* or *IgG/IgM-mediated hypersensitivity,* is characterized by antibodies that attack antigens on the surface of specific cells or tissues. Often the reaction is immediate (15 to 30 minutes after exposure to the antigen). However, it can occur over time, such as in myasthenia gravis. The mechanisms that encompass type II tissue-specific hypersensitivity all occur after the binding of antibody to tissue-specific antigens. The reaction is mediated by the complement system and a variety of effector cells, including tissue macrophages, platelets, natural killer cells, neutrophils, and eosinophils. IgG and IgM are the principal antibodies. Examples of this type of hypersensitivity reaction include ABO transfusion reactions, hemolytic disease of the newborn, myasthenia gravis, hyperacute graft rejection, and autoimmune hemolytic anemia. Transfusion reactions, hemolytic disease of the newborn, and graft rejection are examples of isoimmunity (alloimmunity), a condition in which the immune system reacts against antigens on tissues from other members of the same species.

The initial mechanism during a type II hypersensitivity response is exposure to antigens on the surface of foreign cells. The Fab portion of IgG or IgM antibodies binds to antigens on the target foreign cell to form an antigen–antibody complex (Fig. 10.2). (Refer also to Chapter 9 for a discussion of IgG and IgM antibodies.) The Fc region of the

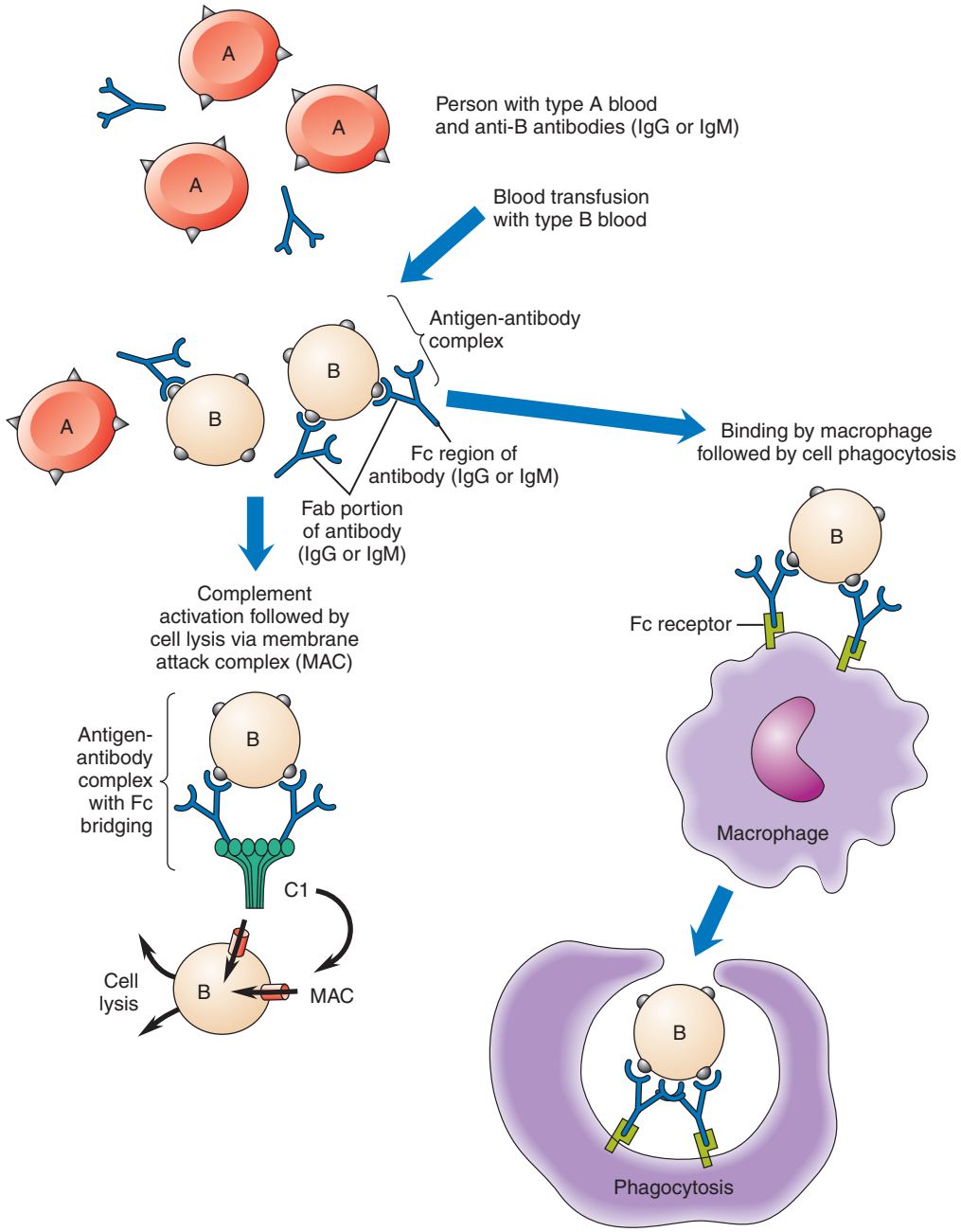


FIG 10.2 Type II hypersensitivity reactions.

IgG or IgM antibodies protrudes away from the cell membrane surface. The Fc region then acts as a bridge between the antigen and complement or the effector cells. This antigen–antibody binding with Fc bridging is the key and leads to lysis of the cell by one of several mechanisms. One mechanism is complement-mediated lysis. Complement-mediated lysis occurs through the classical pathway for activation of complement. The classical pathway of complement generates the activated complement component C3b via splitting of C4 and C2 by C1 (see Chapter 9). The activated complement component C3b is bound to the target cell by the Fc region of IgG or IgM. C3b increases opsonization, which in turn increases the capacity of the system to allow lysis by other effector cells or by complement itself. Lysis of the foreign cell by complement occurs via the C5–C9 membrane attack complex (MAC), which disrupts the plasma membrane of the cell.

Transfusion Reaction

An example of this type of mechanism is an acute hemolytic blood transfusion reaction that occurs when a person receives blood from someone with a different blood group type (Table 10.3). In this case the recipient of the blood transfusion has antibodies to the donor's red blood cell (RBC) antigens. For example, if a person with type A blood with type A antigens and anti-B antibodies incorrectly receives type B blood with B antigens and anti-A antibodies, the anti-B antibodies will attach to the surface of the infused type B RBCs (see Fig. 10.2) and the anti-A antibodies in the infusion will attach to the surface of the circulating type A RBCs. This event will stimulate the destruction of large numbers of RBCs. The resulting signs and symptoms of this major blood group reaction include fever, chills, flushing, tachycardia, hypotension, low back pain, pleuritic chest pain, nausea, vomiting, restlessness,

TABLE 10.3 Major Blood Groups

Blood Group:

Phenotype	Antigens	Possible Genotype	Antibodies in Serum	Chromosome
A	A	AA or AO	Anti-B	9q34.2
B	B	BB or BO	Anti-A	
AB	A and B	AB	None	
O	H	OO	Anti-A, Anti-B	

Other Clinically Significant Major Blood Group Systems:

Phenotype	Number of Antigens	Most Common Major Antigens	Clinically Significant Antibodies
Rhesus (Rh):	50	D, C, E, c, e	
Rh+		DD; Dd	
Rh–		dd	Anti-D; Anti-c; Anti-C; Anti-E; Anti-e
Kell or	25	K, k	Anti-K; Anti-Ku;
Kell – Cellano:			Anti-k; Anti-Kp^a
			Anti-Kp^b; Anti-Js^a; Anti-Js^b
MNS:	46	M, N, S, s, U	Anti-S; Anti-s; Anti-U; Anti-M; Anti-N
Duffy:	6	Fy^a; Fy^b; Fy3; Fy4; Fy5, Fy6	Anti-Fy3; Anti-Fy^a; Anti-Fy^b
Kidd:	3	Jk1 (Jk^a); Jk2 (Jk^b) and Jk3	Anti-Jk^a; Anti-Jk^b

Derived from: Dean L. Blood Groups and Red Cell Antigens [Internet]. Bethesda, MD – National Center for Biotechnology Information (US), 2005.

Frequency of Blood Types in the U.S. Population:

Blood Type	Caucasian	African American	Hispanic	Asian
O+	37%	47%	53%	39%
O–	8%	4%	4%	1%
A+	33%	24%	29%	27%
A–	7%	2%	2%	0.5%
B+	9%	18%	9%	25%
B–	2%	1%	1%	0.4%
AB+	3%	4%	2%	7%
AB–	1%	0.3%	0.2%	0.1%

Source: American Red Cross Blood Types Chart.

anxiety, oliguria, and headache. The reaction may progress to shock and death.

Transfusion reactions are not always immediate. They may be delayed from a few days to 2 weeks, especially in persons requiring periodic transfusions such as in sickle cell anemia or thalassemia. In many cases, antibodies occur as a result of blood component exposure from multiple transfusions. Delayed reactions can also occur as a result of transplantation. In delayed transfusion reactions, often the mechanism of action is related to differences in erythrocyte (RBC) antigens between blood donors rather than differences in the major ABO blood groups. Many hundreds of RBC antigens have been identified; fortunately, most of them are rare. In patients with sickle cell anemia, delayed hemolytic transfusion reactions can occur due to differences in RBC antigens between blood donors with resulting alloimmunization. The rate of alloimmunization with the development of delayed transfusion reaction in sickle cell patients is higher than in other patient populations. The resulting type II hypersensitivity reaction causes the destruction of both the donor's and the recipient's RBCs along with the symptoms of acute hemolysis, and severe vasoocclusive crisis.

Hemolytic Disease of the Newborn

A second mechanism for antigen–antibody binding in type II hypersensitivity reactions is direct destruction by Fc-bearing effector cells, such as macrophages. The macrophage can link to exposed Fc antibody regions. Once this bridging occurs, the foreign cell is phagocytized and destroyed by lysosomes within the effector cell. This can be mediated with or without complement involvement. An example of this type of mechanism is hemolytic disease of the newborn (erythroblastosis fetalis).

This condition occurs during pregnancy when an Rh-negative mother is sensitized to the Rh-positive red cell group antigens of her fetus because of exposure during her current or a previous pregnancy. The mother's IgG Rh-positive antibodies cross the placental barrier and attack the red blood cells of the fetus. The mother's exposure occurs when mixing of fetal and maternal blood takes place. This can occur during an episode of antepartal bleeding or trauma to the placenta, during birth, or during miscarriage of an Rh-positive child. Of these situations, the most likely time for mixing of fetal and maternal blood is at the time of delivery. After this exposure, Rh-positive antibodies gradually develop in the mother and can affect her subsequent children. It takes as little as 1 cm^3 of fetal blood exposure for antibodies to Rh-positive RBCs to develop in the mother. Usually, the first Rh-positive child is not affected unless placental tearing or leakage into the mother's circulation occurs during pregnancy.

Antibody screens are routinely performed during pregnancy to determine a mother's Rh status, including indirect Coombs test and identification of specific antibodies. If the mother is Rh-negative, RhoGAM is administered at 28 weeks, after any prenatal bleeding, and at delivery for prevention of Rh-positive antibodies. RhoGAM contains antibodies against Rh antigens on fetal blood cells and is given to the mother to destroy fetal cells that may be present in her circulation before her immune system becomes activated and begins to produce anti-Rh antibodies. RhoGAM is not effective if the mother already has

a positive antibody titer for fetal Rh antigens. The routine administration of RhoGAM to women at risk is clinically effective with rare findings of hemolytic disease of the newborn.

If the mother is Rh negative, already has fetal Rh antibodies (i.e., is sensitized), and is carrying an Rh-positive fetus, she will not exhibit any significant physiologic effects except some discomfort from potential polyhydramnios. The mother may experience psychological distress. Most of the symptoms of Rh sensitization occur in the fetus, including signs of anemia, hypoxia, decreased fetal activity, ascites, congestive heart failure, and an elevated baseline heart rate of 180 beats/minute or greater with late decelerations. As a result, there is frequent fetal surveillance by 26 weeks' gestation, including assessment of fetal heart rate, biweekly performance of nonstress tests, execution of serial amniocenteses to measure levels of bilirubin, and/or direct evaluation of fetal hemoglobin and hematocrit levels by means of percutaneous umbilical blood sampling. If fetal anemia becomes severe any time after 18 weeks' gestation, an intrauterine blood transfusion may be necessary to try to prolong the time the fetus is able to stay in utero. After 34 weeks, the fetus is often taken by cesarean section.

Myasthenia Gravis

Myasthenia gravis is an autoimmune disease of the neuromuscular junction. Antibodies are directed primarily against the nicotinic acetylcholine receptor (AChR) on the postsynaptic membrane at the neuromuscular junction. However, antibodies against other proteins can occur, including muscle-specific kinase, or low-density lipoprotein receptor–related protein 4 on muscle membrane surfaces, especially the motor end-plate, or agrin in the postsynaptic membrane at the neuromuscular junction (Fig. 10.3). With antigen–antibody formation at the receptor site, complement is activated and disrupts the muscle cell membrane. B cells are the principal mediator cells. The thymus gland is believed to be the site of anti–AChR antibody development. It also sustains the autoimmune, hypersensitivity reaction. The loss of acetylcholine stimulation at the motor end-plate causes the extreme muscular weakness associated with myasthenia gravis. The major symptoms of myasthenia gravis include ptosis, diplopia, and muscle weakness after exercise that improves with rest.

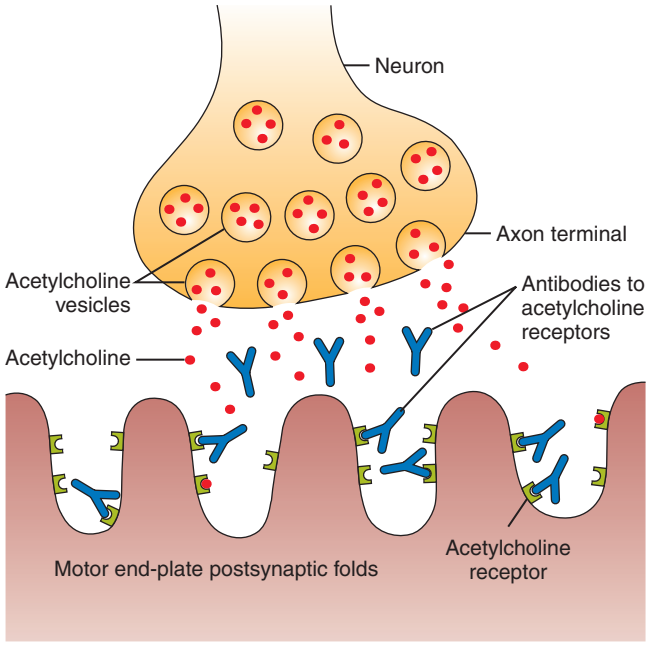


FIG 10.3 Type II hypersensitivity reaction in a person with myasthenia gravis. Having limited receptors available for acetylcholine impairs neuromuscular transmission.

Hyperacute Graft Rejection

Another type II hypersensitivity mechanism for antigen–antibody binding that involves both effector cells and complement is the hyperacute graft rejection that affects transplanted tissues. It occurs when the transplanted donor tissue has an antigen to which the recipient has preformed antibodies. For example, when tissue from a blood type A or B donor is transplanted into a blood type O recipient, the recipient has anti-A and anti-B antibodies. These antibodies will immediately attack the foreign transplanted tissue.

Onset begins immediately after revascularization of the transplanted tissue. At this time, the blood supply from the patient is established in the newly transplanted organ. The patient's antibodies attack the foreign protein antigens and form an antigen–antibody complex. Effector cell infiltration and complement-mediated lysis of donor tissues, inflammation, vascular thrombosis, and hemorrhage occur. The reaction happens so quickly that within 48 hours after transplantation the graft tissue is no longer functioning.

To prevent hyperacute graft rejection, tissue and blood typing of donors and recipients of transplanted tissue is extensive. Lists of potential recipients are matched to donors through organ donation laboratories both regionally and nationally. Only rarely has hyperacute graft rejection occurred because of an error in tissue or blood typing.

Type IIb Hypersensitivity

Graves Disease

Graves disease is an example of a type IIb hypersensitivity reaction that occurs as a result of an immune attack on the thyroid gland. The mechanism for antigen–antibody binding in this type IIb hypersensitivity reaction is mediated by lymphocytes, which infiltrate the thyroid parenchyma. The underlying mechanism(s) that triggers this process is still being investigated, but infection, diet, iodine, stress, and smoking may be contributing factors. In Graves disease, circulating IgG autoantibodies (called *anti-TSHR antibodies*) attack the thyroid-stimulating hormone receptor. This causes the cells to malfunction, stimulating the receptor and producing excess thyroxine (T_4) and triiodothyronine (T_3). The major symptoms of hyperthyroidism include tachycardia, fatigue, weight loss, tremor, heat intolerance, diarrhea, and emotional/mental changes.

Type III Hypersensitivity

Etiology

Type III hypersensitivity results from failure of the immune and phagocytic systems to effectively remove antigen–antibody immune complexes and is not tissue specific. It is also known as an *immune complex reaction.* Type III hypersensitivity is characterized by antigen–antibody complex deposition into tissues, with consequent activation of complement and a subsequent self-sustaining inflammatory reaction. It is not an immediate reaction; it occurs over a period of several hours or longer (i.e., glomerulonephritis) and is often ongoing.

Three possible scenarios can precipitate type III hypersensitivity. First, a recent history of infection or persistent low-grade infection by a microbial or viral agent can stimulate a weak antibody response. The continuing nature of persistent infection or untreated/undertreated infection provides a source of circulating antigen. The antigen and antibody response leads to chronic immune complex production. These immune complexes are not successfully removed from the blood and are deposited in many sites, including blood vessels, glomeruli, and joints. Second, an extrinsic environmental antigen from molds, plants, or animals can be inhaled into the lung, where it is exposed to antibody in the body fluid. This inhalation of an antigen causes antigen–antibody

complex formation in alveoli with immune complex deposition in the alveolar walls. Third, an autoimmune process can develop in which autoantibodies attack self antigens. In this case the body forms both parts of the immune complex. Autoantibodies to either circulating or tissue-fixed self antigens may be produced. Because the self antigens persist over time, chronic immune complex production and deposition in tissues take place.

The mechanism of injury in type III hypersensitivity reactions is from activation of complement and other proinflammatory mediators in response to the antigen–antibody complex deposition. The antibody–antigen complex deposition does not cause the injury directly. The tissue injury is caused by an inflammatory reaction to the antibody–antigen complex. Therefore it is not a tissue-specific reaction. The onset of this reaction occurs up to 6 or more hours after exposure to the antigen. IgG and IgM are the principal antibodies. The principal effector cells are neutrophils and mast cells. The principal mediator of the reaction is complement. Examples of type III hypersensitivity reaction include SLE, immune complex glomerulonephritis, serum sickness, and drug-induced vasculitis. Diseases associated with type III hypersensitivity are listed in Table 10.4.

Pathogenesis

Type III hypersensitivity reactions tend to be ongoing with variations in symptoms based on the fluctuation of antibody-to-antigen ratios, the amount of complement available to mediate the inflammation, and the dynamic nature of the antibody–antigen reaction. It is sometimes difficult to differentiate between type II and type III hypersensitivity reactions. The key differences between a type II and a type III reaction are the location of antigens and the mechanism of injury. As previously described, type II reactions occur in response to tissue-specific antigen located on cell surfaces and involve direct cell death or malfunction from the antigen–antibody reaction. Type III hypersensitivity reactions involve antigens forming antigen–antibody complexes that precipitate out of the blood or body fluid and are deposited into tissues.

Type III hypersensitivity reactions involve a sequential process that begins with interaction between a circulating soluble antigen and soluble antibody or between an insoluble antigen and a soluble antibody. Depending on the concentration of antigen and antibody, multiple cross-linking of antigen and antibody occurs and immune complexes are formed. Most immune complexes are removed effectively before they can cause injury. In type III hypersensitivity, the immune complexes are not removed, which causes an inflammatory process leading to tissue injury.

When both antigen and antibody are small or intermediate in size and soluble, the immune complex precipitates out of the body fluid and is deposited into tissues. When only the antibody is soluble, the antibody reacts with fixed antigen in the tissues. Then the antibody within the complex links with the complement system by its Fc receptors (Fig. 10.4).

Activation of the classic complement cascade causes release of C3a and C5a, as well as the membrane attack complex. C3a stimulates the release of histamine from mast cells, indirectly increasing vascular permeability and vasodilation. Bronchial smooth muscle contraction occurs, resulting in bronchial constriction, wheezing, and coughing. C3a also causes the endothelial cells to become rounder, thereby increasing vascular permeability. The increased vascular permeability leads to edema formation, which provides more space for the movement of cellular inflammatory components. It also dilutes and limits the duration of action of mediators. C5a is an even more powerful component than C3a.

TABLE 10.4 Diseases Associated With Type III Hypersensitivity

Disease	Antigen
Immune complex glomerulonephritis	GBM, exogenous antigens, drugs
SLE	Double-stranded DNA, DNA-histone complex, Sm, RNP, Ro:SSA, La:SSB, centromere
SLE-associated glomerulonephritis	Double-stranded DNA, DNA-histone complex, Sm, RNP, Ro:SSA, La:SSB
Acute allergic alveolitis	Various puffball spores from moldy dwellings
Farmer's lung disease	Thermophilic *Actinomycetes* from contaminated hay or grains
Chemical worker's lung	Isocyanates
Still disease—postinfectious arthritis	RANA, or none identified
Rheumatoid arthritis	RANA
Serum sickness	Lymphocytes or thymocytes from heterologous serum
Henoch–Schönlein purpura	Upper respiratory tract viruses, drugs (antibiotics and thiazides), foods (milk, fish, eggs, rice, nuts, beans), and immunizations
Drug-induced vasculitis	Drugs (antibiotics and thiazides)
Polyarteritis nodosa	Antineutrophil cytoplasmic
Wegener granulomatosus	Antineutrophil cytoplasmic
Goodpasture syndrome	GBM

GBM, Glomerular basement membrane; *La,* Lane; *RANA,* rheumatoid arthritis nuclear antigen; *RNP,* ribonucleoprotein; *Ro,* Robert; *SLE,* systemic lupus erythematosus; *Sm,* Smith; *SS,* single stranded.

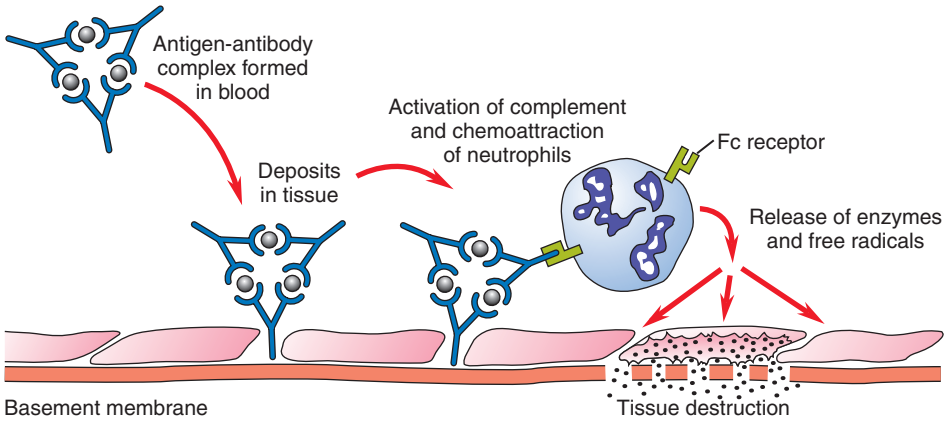


FIG 10.4 Type III hypersensitivity reaction.

It causes a powerful release of proinflammatory mediators with actions that are identical with those of C3a. It is also a powerful chemotactic agent for neutrophils and causes a respiratory burst within neutrophils in which oxygen consumption is increased to 50 times normal along with increased glucose uptake and procoagulant activity.

As a result of the activation of complement, neutrophils, macrophages, and mast cells are attracted to the area and are activated. These cells begin lysis and destruction of tissue via the release of cytokines and the inflammatory response. The inflammation causes tissue destruction, scarring, and further reaction of the immune system against the damaged tissue (see Fig. 10.4). The persistence of the inflammation is due to the ongoing release of autoantigen particles in the damaged tissue that stimulate autoreactive B cells, leading to the formation and deposition of more immune complexes.

Tissue Deposition

Antigen–antibody complex deposition in tissues is affected by a number of factors, including size and clearance rate. Smaller immune complexes are able to circulate for longer periods, which may increase the immune response. However, small complexes also can be removed more easily because they can pass through the glomerular basement membrane. The very large complexes can be phagocytized more easily because they are easily marked or fixed by complement and bound to red blood cells. The large complexes can then be transported to the liver, where they are phagocytized by the reticuloendothelial system—particularly the Kupffer cells in the liver—and easily removed from the system. However, large complexes can become stuck in the kidney where they are unable to cross the glomerular basement membrane.

Increased vascular permeability as a result of histamine or other vasoactive mediator release is also hypothesized to be an important factor in tissue deposition. Researchers have found that small immune complexes can be deposited in tissues treated with vasoactive mediators. Sites of increased turbulence and blood pressure tend to have increased immune complex deposition. These sites include the glomerular capillaries, joint linings, ciliary body, pulmonary alveolar membranes, and vascular endothelial linings, especially around curves or bifurcations.

Intermediate-sized immune complexes tend to be deposited because they do not fix complement well, do not bind with RBCs well after fixation, and are not readily removed by the mononuclear phagocyte system. Large numbers of any size immune complex can be deposited if they are so numerous that the phagocytic cells are overwhelmed. Deposition of immune complexes also depends on their immunoglobulin class and the affinity between the antigen and antibody. Finally, deposition may be affected by the type of antigen or by the relationship between the immune complex and sites with increased collagen. Because DNA and collagen have a strong affinity, an increased quantity of DNA–anti-DNA immune complexes may be deposited in collagen membranes, such as in the kidney. The electrical charge of the immune complex may affect where it is deposited. For example, a positively charged immune complex may be attracted by a negatively charged basement membrane.

Immune Complex Glomerulonephritis

Etiology. Immune complex glomerulonephritis (an inflammatory renal disorder) is an example of a type III hypersensitivity reaction typically occurring 10 to 14 days after infection with a *Streptococcus* or staphylococcal bacterial strain. It involves the interaction of soluble bacterial antigens with soluble antibodies and is a frequent cause of glomerulonephritis. The circulating immune complex is then deposited in the glomerular capillary wall and mesangium.

Clinical manifestations and treatment. This deposition causes glomerular inflammation, activation of mesangial cells, and ultimately damage to the glomerular basement membrane with resultant proteinuria, hematuria, hypertension, oliguria, and red cell casts in the urine (see Chapter 27). In some types of glomerulonephritis, the patient may have nephrotic syndrome and acute renal failure that may progress to chronic renal failure.

Treatment of glomerulonephritis involves the use of corticosteroids and other medications to decrease inflammation. Antihistamines and antiserotonins have been tried in attempts to decrease vasoactive mediators and vascular permeability. Anticoagulants and antiplatelet medications such as aspirin, as well as plasmapheresis, are currently being studied. In plasmapheresis, plasma is removed from the blood and fresh frozen plasma or albumin is used to replace the withdrawn plasma.

Systemic Lupus Erythematosus

Etiology. SLE is an example of a type III hypersensitivity reaction caused by autoantibody production. SLE tends to occur more frequently in women than in men (ratio of 7:1) and with an incidence of 20 cases per 100,000 persons. It is primarily characterized by the development of antibodies against nuclear antigens such as DNA, deoxyribonucleohistone, and RNA. Production of autoantibodies to RBCs, neutrophils, platelets, lymphocytes, and other organs or tissues may also occur. The antinuclear antibodies (ANAs) and anti-DNA autoantibodies attach to components of the nucleus to form immune complexes that are deposited on collagen-rich tissues, including the glomerular basement membrane and the dermal-epidermal junction. The exact mechanism causing cell damage and the release of nuclear components and subsequent development of ANA and anti-DNA antibodies is not known. It is proposed that the release of nuclear material from the cell occurs during cell death or secondary to pyroptosis, a proinflammatory process that leads to the release of intact nuclei. Once formed, the autoantibodies can react with DNA and nuclear components from damaged cells anywhere in the body. The resulting inflammatory response causes increased cell damage and further antigen–antibody immune complex formation, thus leading to a cyclic process. The immune complex deposition and resulting inflammatory response cause the signs and symptoms of SLE.

Clinical manifestations. SLE can affect any organ system and, as such, it can present with a variety of signs and symptoms. Because of the variable presentation, diagnostic-specific criteria based on the presence of specific signs, symptoms, and laboratory findings have been developed. Kidney involvement is common and may lead to nephritis and glomerulonephritis. Skin symptoms are wide ranging and include malar "butterfly" rash, erythematous rash on exposed skin, purpura, alopecia, mucosal ulcerations, subcutaneous nodules, and splinter hemorrhages. The malar butterfly rash occurs across the nose onto the cheeks in approximately 26% of patients with acute SLE. The butterfly rash can be flat or raised and always spares the nasolabial folds. The rash that occurs on sun-exposed skin can be superficial or indurated and is nonpruritic. It can occur on the face, chest, shoulders, extensor surfaces of the arms, and backs of the hands. The color ranges from red to reddish purple. Most patients have symptoms of arthritis or polyarthralgia. Other symptoms include pleurisy, pericarditis, restrictive pulmonary disease, retinal changes, thrombocytopenia, anemia, and gastrointestinal ulceration. Central nervous system involvement includes neuritis, seizures, depression, or psychosis. A positive ANA test is usually present with positive anti-DNA (antibody to native DNA) and anti-Sm (antibody to Smith nuclear antigen) (see Chapter 52 for more information).

Treatment. Treatment of SLE depends on the organ system affected and includes administration of nonsteroidal antiinflammatory drugs (NSAIDs) such as aspirin, systemic corticosteroids, antimalarials such as hydroxychloroquine (Plaquenil), alkylating agents such as cyclophosphamide, and immunosuppressives such as methotrexate, abatacept or belimumab (Benlysta). NSAIDS are used for arthritis symptoms, generalized aching, and fever. Antimalarials are a common treatment for SLE

because of their effectiveness in treating arthritis and rash symptoms. The major side effect of antimalarials is ocular toxicity, which can be prevented by limiting therapy to 10 years or less, not exceeding 6.5 mg/kg/day dosages, and undergoing yearly ophthalmic evaluations. Other side effects include myopathy, pigmentation changes, gastrointestinal effects, liver enzyme changes, dizziness, and emotional changes.

Corticosteroids decrease inflammation and provide immunosuppression, which can decrease symptoms and add to the patient's quality of life. Corticosteroids are used for kidney and central nervous system involvement in SLE. If the patient does not respond to corticosteroids, immunosuppressives or cytotoxic agents can be used. Methotrexate is a commonly used immunosuppressive (see earlier). Benlysta is a monoclonal antibody that inhibits B cells and is used only when there is severe, aggressive disease. It is given intravenously every 4 weeks and is very expensive. It is associated with risks for infection, malignancy, and depression. Antibiotics are commonly prescribed because of frequent infections secondary to immune system compromise. Sunscreens with maximum sun protection factor values are necessary to prevent photosensitivity reactions. Patients are also encouraged to avoid direct sunlight from 10 AM until at least 3 PM.

Type IV Hypersensitivity

Type IV hypersensitivity, also known as a *delayed hypersensitivity reaction,* is characterized by tissue damage resulting from a delayed cellular reaction to an antigen. This hypersensitivity category has been reclassified into four subclasses that are divided by the type of principal mediator and cytokines involved. The principal mediators are lymphocytes, including $CD4^+$ T helper (Th1) cells, Th2 cells or antigen-sensitized cytotoxic $CD8^+$ T cells (Tc), macrophages, mast cells (in the early phases), and neutrophils. They mediate the reaction by releasing a cascade of cytokines. This reaction is slow in onset and evolves gradually—beginning 24 hours after exposure and lasting up to 2 to 4 weeks after exposure. Clinical manifestations may linger for an even longer period.

Mast cell degranulation occurs early in the evolution of a delayed hypersensitivity reaction, followed by lymphocyte and macrophage invasion. The mast cells are gatekeepers that regulate leukocyte migration in the microvasculature. Unlike that occurring in type I hypersensitivity reactions, the mast cell degranulation is more limited and localized. The reaction is also limited by the action of suppressor T cells, which inhibit other T-cell actions. The reasons for mast cell activation in this type of hypersensitivity reaction are not well understood. However, the combined action of mast cell and T-cell mediators recruits other T cells and macrophages to the site.

The four subclasses of type IV hypersensitivity are IVa) mediated by $CD4^+$ Th1 cells and macrophages/monocytes with cytokines interferon (IFN) gamma, IL-1, IL-2, IL-3, and IL-31; IVb) mediated by $CD4^+$ Th2 cells and eosinophils with cytokines IL-5, IL-4, and IL-13; IVc) mediated by cytotoxic $CD8^+$ T cells with perforin, granzyme B, and Fas ligand; and IVd) mediated by $CD4^+$ T cells, $CD8^+$ T cells, and neutrophils with cytokines IL-6, IL-8, IL-17A, IL-17F, IL23, IL-36, and granulocyte-macrophage colony-stimulating factor.

Many types of delayed hypersensitivity reactions are recognized, including granulomatous hypersensitivity, persistent asthma, allergic contact dermatitis, tuberculin-type hypersensitivity, SJS, toxic epidermal necrolysis (TEN), and pustular psoriasis (PP).

Type IVa — Granulomatous Hypersensitivity

Granulomatous hypersensitivity reaction is a primary defense against intracellular infections and represents a chronic type IV hypersensitivity reaction. It is a protective defense reaction that eventually causes tissue destruction because of persistence of the antigen. In this type of hypersensitivity, antigen is not destroyed within the macrophages, either because of failure of lysosome–phagosome fusion, as in tuberculosis and leprosy, or because of the resistance of various materials to internal lysozymes, as in retained suture material or talc. In an effort to protect the host, lymphocytes and macrophages actually cause the tissue damage by releasing cytokines and stimulating an inflammatory response.

Antigen is engulfed and ingested by macrophages attempting to destroy the antigen, but these actions are unsuccessful in type IV hypersensitivity. The macrophages form a core of inflammatory cells that include lymphocytes, tissue histiocytes, eosinophils, plasma cells, giant cells, and epithelioid cells. This collection of inflammatory cells develops into a ball-like mass called a *granuloma.* The predominant cell in the granuloma is the macrophage. Epithelioid cells originate from macrophages and are large, flat cells with a large amount of endoplasmic reticulum. When epithelioid cells fuse, they form multinucleated giant cells. This core is surrounded by lymphocytes. Gradually, fibroblastic activity and increased collagen synthesis cause the granuloma to become fibrotic with scar formation. Often, central necrosis occurs within the granuloma and is called *caseous* or *cheesy necrosis.* Patients with granulomatous disease have a variety of symptoms. Granulomatous disease examples and the pathogens associated with them are listed in Table 10.5.

Testing for granulomatous diseases also involves skin testing. For example, the tuberculin skin test (i.e., Mantoux or purified protein derivative [PPD]) identifies infection but is not diagnostic of active disease. The test involves *intradermal* injection of 0.1 mL of PPD into the volar surface of the forearm. There should be a clear bleb at the site of the test. The skin test is then read in 48 to 72 hours. A negative PPD skin test is one in which there is either no reaction or only redness at the site. A positive result is one in which there is equal to or greater than 10 to 15 mm of induration (or tissue firmness) at the site. In patients with HIV a positive result is equal to or greater than 5 mm of induration. However, if the HIV patient's $CD4^+$ cell count is low (<200 cells/μL), or if the patient is febrile or elderly, there may be no reaction to any skin testing. This is called *anergy* and causes a false negative in many cases.

Type IVa – Tuberculin-Type Hypersensitivity

Tuberculin-type hypersensitivity is a type of allergic contact dermatitis and occurs when someone who has been previously infected by tuberculosis is exposed to tuberculin antigen in a tuberculin test. It is a dermal phenomenon that peaks in 48 to 72 hours. The person experiences erythema, induration, and inflammation at the site of the intradermal injection. Because the amount injected is so small, the reaction disappears when the antigen has degraded. However, people with severe reactions may experience tissue necrosis at the site.

Type IVa – Allergic Contact Dermatitis

Allergic contact dermatitis is the most familiar kind of type IV hypersensitivity. It is an immune or inflammatory response to a wide variety

TABLE 10.5 Granulomatous Disease Associated With Type IV Hypersensitivity

Disease	Bacterium
Tuberculosis	*Mycobacterium tuberculosis*
Leprosy	*Mycobacterium leprae*
Histoplasmosis	*Histoplasma capsulatum*
Coccidioidomycosis	*Coccidioides immitis*
Brucellosis	*Brucella abortus*
	Brucella suis (less common)
	Brucella melitensis (less common)
Tularemia	*Francisella (Pasteurella) tularensis*

of plant oils, chemicals, ointments, clothing, cosmetics, dyes, and adhesives. It is an epidermal phenomenon with two phases: an induction or sensitization phase and an elicitation phase. As a delayed reaction, it peaks in 48 to 72 hours and is usually resolved within 4 weeks, provided that continued exposure to the antigen is avoided.

During the induction phase, a skin-penetrating antigen is applied to or touches the epidermis or a mucosal membrane. The antigen is very small, low in molecular weight, and in an incomplete form. This incomplete, lipid-soluble antigen is called a *hapten*. The hapten must first penetrate the epidermal barrier (stratum corneum), where it links with a normal body protein, called a *carrier*. Only after the hapten combines with the carrier is it a complete antigen—often called a *hapten conjugate*. The means of contact or the antigenic compound itself may help in the sensitization of the epidermis during the induction phase.

The complete antigen is then processed by Langerhans cells or other dermal dendritic cells located in the suprabasal epidermis. The dermal dendritic cells or Langerhans cells are the antigen-presenting immune cells of the skin and mucous membranes. These cells then move to the local lymph channel, where they migrate to the regional lymph node. Within the lymph node, the dermal dendritic cells display the now-processed complete antigen to $CD4^+$ T cells in the paracortex. Then the antigen-sensitized $CD4^+$ T cells release a cytokine cascade of vasoactive substances, which initiate an inflammatory response and attract other effector cells. The primary cytokines include IL-2, IL-3, IL-31, IFN, TNF, and macrophage-stimulating factors (Fig. 10.5). After about 72 hours, the reaction begins to decrease because of degradation of the antigen and production of prostaglandin E, which inhibits IL-1 and IL-2 production.

As a result of the antigen exposure, T cells bearing receptors for the specific antigen are formed and enter the bloodstream. Any future exposure or contact with the antigen can trigger a cellular reaction, which takes place at each subsequent exposure site. This is called the

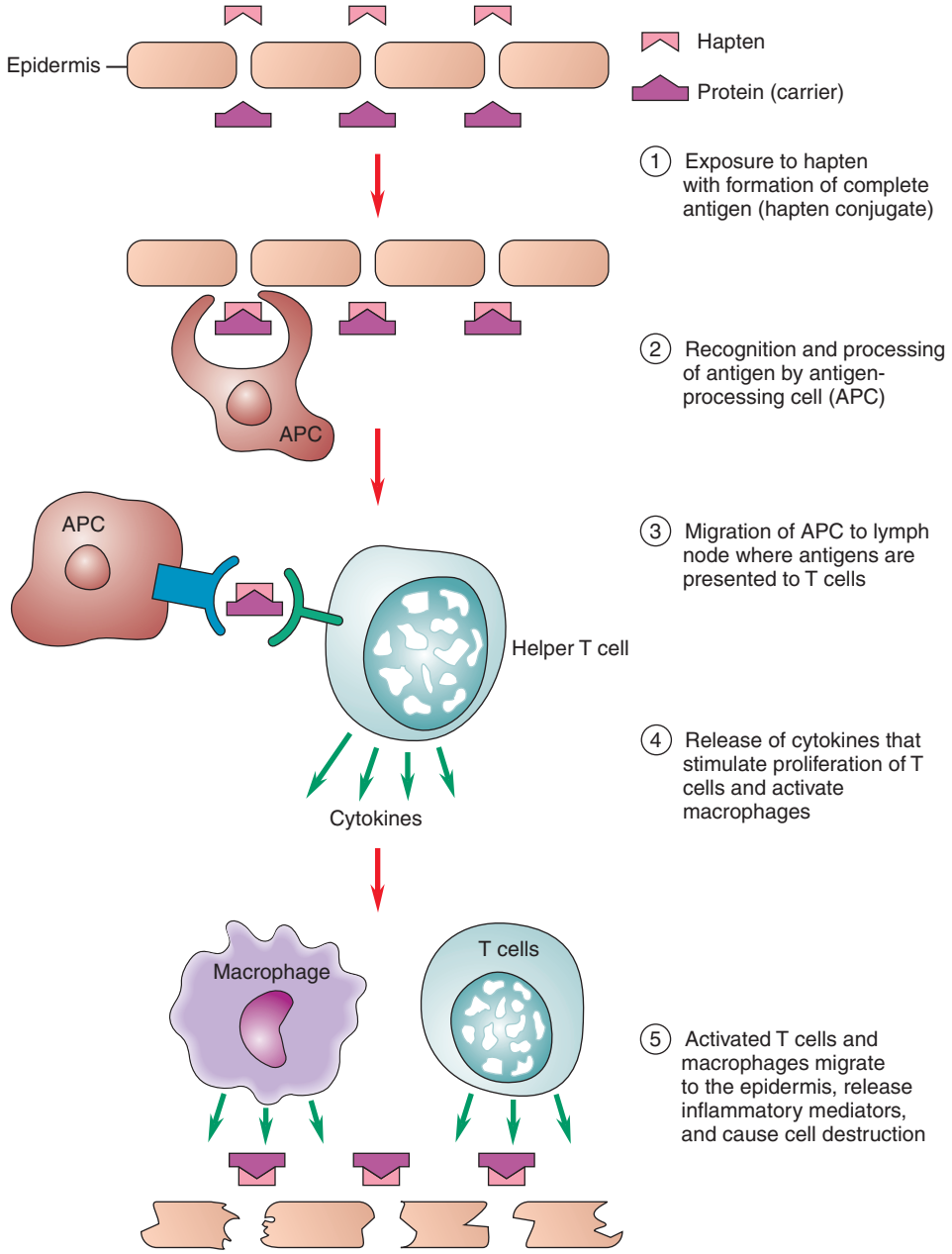


FIG 10.5 Type IV hypersensitivity reaction.

elicitation phase. Langerhans cells with the antigen interact with specific T cells, resulting in cytokine activation and release, as well as the proliferation of other inflammatory mediators and $CD4^+$ T cells.

Cross-reactivity with the related substances also occurs. For example, a person with allergic contact dermatitis to nickel will react when exposed to a variety of nickel alloys, including the metal in earrings, zippers, snaps, and belt buckles. Skin symptoms resulting from contact dermatitis include redness (erythema), edema, pruritus, and blisters. People with sensitivities may also experience respiratory symptoms if exposed to aerosolized hapten. This situation could occur when a person is downwind from burning poison ivy or burning tires. Delayed hypersensitivity reaction (type IV) caused by exposure to the compounds in latex gloves is an occupational risk in surgeons and nurses who work in operating rooms.

Patch testing for diagnosis of allergic contact dermatitis is indicated when symptoms persist despite avoidance of the suspected agent and with appropriate use of topical therapy. There are several types of patch tests, including open and closed testing used by dermatologists and allergists. In open patch testing, the suspected allergen is applied twice daily to the upper arm and left uncovered for 2 days. In closed patch tests, the suspected allergen is applied to the skin away from the original site of eruption and is covered with an adhesive bandage. The bandage is then removed in 2 days. The results are graded according to response. A 1+ positive patch test response involves erythema of the entire area of exposure. Patchy pustular responses are not positive and are irritant reactions. A 2+ patch test response involves erythema and vesicles. A 3+ patch test response involves erythema, vesicles, and bullae.

Type IVb – Persistent Asthma

Persistent asthma is a form of type IVb hypersensitivity mediated by $CD4^+$ T helper 2 (Th2) cells in response to a variety of stimuli. The $CD4^+$ Th2 cells, as well as other T cells such as Th17, Th 22, and Th9, produce and release inflammatory cytokines and initiate eosinophilic involvement. This production of inflammatory cytokines includes IL-4, IL-5, IL-9, and IL-13, among others, as well as increased levels of IgE and eosinophil production. The interaction of these inflammatory vasoactive substances leads to the development of severe airway inflammation and asthma symptoms. The variability of individual asthma symptoms is influenced by multiple genetic, racial, and environmental factors that also influence the reactivity of the individual's immune system. (See Chapter 22 for further discussion of asthma.)

Type IVc – Stevens–Johnson Syndrome and Toxic Epidermal Necrolysis

SJS and TEN are type IVc hypersensitivity reactions in response to medications or upper respiratory tract infection caused by herpes simplex virus or *Mycoplasma pneumoniae.* Medications or their metabolites can interact with HLA class I alleles such as HLA-B*57:01 causing $CD8^+$ T-cell activation with the production of cytokines. This type of immune system–mediated reaction is an adverse drug reaction that usually occurs approximately 4 to 14 days after starting a new medication, but may not occur for up to 6 weeks. In these cases, the immune-mediated reaction is a delayed reaction to a medication, in contrast to type I hypersensitivity reactions, which have an onset within 1 hour. In other patients, upper respiratory tract infectious agents interact with $CD8^+$ T cells causing activation and release of cytokines. Because it is a delayed reaction, symptoms may not occur for 3 to 4 days and may last for 6 weeks.

When the onset of SJS/TEN occurs after ingestion of a medication, the medication or its metabolic byproduct binds directly to immune receptors on cytotoxic $CD8^+$ T cells. Activation of the $CD8^+$ T cells causes the release of secretory granulysin, perforin, and tumor suppressor protein p53, as well as the release of retinoic acid. The result of this activation causes extensive keratinocyte apoptosis, leading to the symptoms of severe blistering, skin peeling, and multiorgan damage, particularly in the liver. A similar cascade of immune system–activated products occurs when a type IVc hypersensitivity occurs to respiratory infectious agents.

The medications most commonly implicated in SJS/TEN include anticonvulsants (phenytoin, phenobarbital), antibiotics (sulfonamides, aminopenicillins, antifungals), antimalarials, allopurinol, and certain NSAIDs. The treatment of SJS/TEN includes fluid replacement, nutritional support, and extensive wound care, as well as the use of medications such as antihistamines, corticosteroids, or immunoglobulin therapy and plasmapheresis.

Type IVd – Pustular Psoriasis

PP is a type IVd hypersensitivity reaction. It is a rare and severe form of psoriasis distinguished by the presence of neutrophilic collections intradermally. In type IVd hypersensitivity, $CD4^+$ T cells or $CD8^+$ T cells activate neutrophils as well as a cascade of proinflammatory cytokines, including IL-6, IL-22, IL-23, IL-17A, and IL-17F. The skin lesions are a complex interaction between dendritic cells, keratinocytes, and T cells. Symptoms of PP include repeated attacks of diffuse, erythematous, pustular rash associated with high-grade fever, general malaise, and leukocytosis. The rash is often located in flexural areas. Treatment includes use of topical or systemic corticosteroids and phototherapy, as well as medications including acitretin, cyclosporine, methotrexate, and TNF-blocking medications such as infliximab (Remicade).

KEY POINTS

- Type I hypersensitivity is an immediate allergic or anaphylactic type of reaction mediated primarily by sensitized mast cells. The reaction is initiated when IgE antibodies located on the mast cell membrane are bound by antigen, with subsequent cross-linking of IgE receptors. Mast cell degranulation releases chemicals that mediate the signs and symptoms of anaphylaxis. Released histamine, kinin, prostaglandins, interleukins, and leukotrienes cause increased vascular permeability, vasodilation, hypotension, urticaria, and bronchoconstriction. Examples of type I reactions include acute drug reactions, bee sting reactions, and allergic contact dermatitis.
- Type II hypersensitivity occurs when antibodies are formed against antigens on cell surfaces, usually resulting in lysis of target cells. Cell lysis may be mediated by activated complement fragments (MAC) or by phagocytic cells that are attracted to target cells by the attached antibodies. Examples include transfusion reactions, erythroblastosis fetalis, myasthenia gravis, and hyperacute graft rejection.
- Type III hypersensitivity reactions occur when antigen–antibody complexes are deposited in tissues and result in the activation of complement and subsequent tissue inflammation and destruction. The antigen–antibody complexes activate the complement cascade and subsequently attract phagocytic cells to the tissue. History of persistent low-grade infections, inhalation of antigens into alveoli, and autoimmune production of antibodies may result in chronic production of antigen–antibody complexes. Examples include immune glomerulonephritis and systemic lupus erythematosus (SLE).
- Type IV hypersensitivity reactions are T-cell mediated and do not require antibody production, in contrast to type I, II, and III reactions. Sensitized T cells react with altered or foreign cells and initiate inflammation. Granulomatous hypersensitivity, tuberculin reactions, persistent asthma, Stevens-Johnson syndrome (SJS), and pustular psoriasis (PP) are examples.

DEFICIENT IMMUNE RESPONSES

Deficient immune responses result from a functional decrease in one or more components of the immune system. These disorders can affect lymphocytes, antibodies, phagocytes, and complement proteins. Two types of immune deficiency are differentiated: primary and secondary. Primary disorders are immune deficiencies not attributable to other causes; these may be congenital or acquired. Examples of primary immunodeficiency disorders include SCID syndrome, DiGeorge syndrome, selective IgA deficiency, and HIV/AIDS. Persons with a primary immunodeficiency can be predisposed to multiple deficiencies, as in agammaglobulinemia or SCID, or can have a single phenotype deficiency predisposing to a specific problem, such as in patients with STAT-1 dependent chronic mucocutaneous candidiasis. Secondary immunodeficiency disorders are a consequence of non–immune system disorders or treatments that secondarily affect immune function. Examples of secondary disorders include those associated with hyperlipidemia or malnutrition, medical treatments such as cancer chemotherapy, or biopsychosocial stress such as postsurgical immune suppression.

PRIMARY IMMUNODEFICIENCY DISORDERS

Primary immunodeficiency disorders (PIDs) include congenital phenotypes that result from abnormal development or maturation of immune cells, as well as acquired primary disorders of immune cells such as HIV/AIDS. HIV/AIDS is described in Chapter 12, and only the congenital forms of primary immunodeficiency are included here. Primary congenital immunodeficiency disorders are genetic disorders that are often sex linked and occur in 1 in 58,000 infants. The most common primary disorders are listed in Table 10.6. There are also strong genetic epidemiologic data suggesting that many people have some inborn errors of immunity that may vary in clinical significance, expression, and severity. This novel observation is changing how we view immunodeficiency. The first clinical indicators of immunodeficiency disorders are the signs and symptoms of infection.

B-Cell and T-Cell Combined Disorders

Severe Combined Immunodeficiency Disorders

Etiology and pathogenesis. SCIDs are inherited PIDs that arise from a variety of genetic defects. There is wide phenotypic and immunologic variability in SCIDs with differences in their severity and clinical features. Nearly all SCIDs are characterized by the absence or dysfunction of T cells, which affects both cellular and humoral adaptive immunity. However, different forms of SCID may also have B-cell and natural killer (NK)–cell absence or dysfunction. For example, depending on the genetic defect, B cells may be present (T–B+) or absent (T–B–). Although B cells may be present, the B cells are unable to produce functional antibodies because of the absence of T-cell help.

T–B+ NK– X-linked SCID is thought to be the most common type of SCID. It is caused by mutations in the *IL2RG* gene that is responsible for encoding the common γ chain, which is needed by multiple cytokine

TABLE 10.6 Primary Congenital Immunodeficiency Disorders

Disorder	Functional Deficiency	Error
X-linked agammaglobulinemia	B cells; antibody	*btk* gene mutation on long arm of X chromosome (Xq21.3-Xq22)
Common variable (acquired) hypogammaglobulinemia	Antibody; B cells may be decreased	Unknown: possible *CARMA1* variant and *BOB1* variant
Selective IgA deficiency	IgA antibody	Unknown: possible alterations in transmembrane activator, calcium modulator, and cycophilin ligand interactor gene
Chronic mucocutaneous candidiasis	B cells; secretory IgA antibody	Autosomal-recessive deficiency in IL-17RA; autosomal-dominant deficiency of IL-17F
X-linked hyper-IgM syndrome	Low levels of IgG and IgA; normal to elevated IgM antibody	Defect in CD40 ligand on T cells
Selective deficiency of IgG subclasses	IgG antibody subclass	Unknown
Transient hypogammaglobulinemia of infancy	Low IgG and IgA antibodies	Unknown
X-linked lymphoproliferative disease or Duncan X-linked lymphoproliferative disease	Anti–Epstein–Barr virus–linked antigen antibody	Mutations in *SH2D1A* and *XIAP (BIRC4)*
22a11.2 deletion syndrome	Primarily T cells	Chromosome 22q11.2 deletion
T–B+ NK– SCID — Autosomal recessive	T cells, antibody	Defects of JAK3
X-Linked recessive, T–B+ NK– SCID	T cells, antibody	Defects of *IL2RG* gene
T–B–NK+ SCID (Omenn syndrome)	T cells, B cells, antibody	Mutations of *RAG1* or *RAG2*
MHC class II expression deficiency (bare lymphocyte syndrome)	Lack of antigen Class II gene expression, T cells, antibody	Mutations in genes coding: class II transactivator (CIITA), RFX5 (regulatory factor X5), RFXAP (RFX-associated protein), or RFXANK
Wiskott–Aldrich syndrome (immunodeficiency with eczema and thrombocytopenia)	Antibody, T cells, platelets	Mutation of *WASp* gene; defect of short arm of X chromosome at Xp11.23
Chédiak–Higashi syndrome	Natural killer cells, phagocytic cells, granulocytes, platelets	Autosomal-recessive disorder; mutations in *LYST* (aka *CHS1*) gene on chromosome 1
Chronic granulomatous disease of childhood	Phagocytic cells (neutrophils)	Defect in superoxide generation; Mutations in genes *CYBA, NCF-1, NCF-2, CYBB;* most are X-linked recessive

btk, Bruton tyrosine kinase; *JAK,* Janus kinase; *MHC,* major histocompatibility complex; *SCID,* severe combined immunodeficiency; *WASp,* WAS protein.

receptors, including IL-2, IL-4, IL-7, IL-9, IL-15, and IL-21. This defect results in the loss of all T-cell and NK-cell development. B cells are present but do not function normally. Another type of T−B+ NK− SCID is caused by a mutation in the gene encoding janus kinase 3 (JAK3 deficiency), which is important for differentiation of lymphoid cells. The result of this defect also causes low T cells and NK cells with nonfunctional B cells. The IL-7R*a* gene–deficiency SCID is noted for low T cells but normal numbers of nonfunctional B cells and normal NK cells (T−B+ NK+ SCID).

T−B−NK+ SCID patients have a defect caused by mutations of recombination activating genes 1 or 2 *(RAG1* or *RAG2),* both of which are involved in the process of antigen receptor gene assembly. This absence leads to abnormalities in T-cell and B-cell receptor expression and development. NK cell development is not affected because they do not express immunoglobulin-like receptors. Omenn syndrome is an example of this type of defect characterized by SCID, eczema-like rash, hepatosplenomegaly, and alopecia.

Adenosine deaminase deficiency SCID (ADA SCID) patients have a genetic mutation on chromosome 20q12.2-13.11. This T−B−NK− ADA SCID results in a deficiency of ADA, an intracellular enzyme of purine metabolism, whose function is to remove toxic metabolites formed in lymphocytes. As a result of this deficiency, there is an accumulation of toxic metabolites (particularly deoxyadenosine (dAdo) and deoxyadenosine triphosphate (dATP)) which are lethal to T cells, B cells, and NK cells, as well as their precursors. Patients with this autosomal-recessive form of SCID have a complete absence of lymphocytes. They have severe recurrent infections and skeletal abnormalities.

Clinical manifestations and treatment. Untreated infants with SCIDs are severely ill with high early mortality — usually at less than 1 year. They are usually ill within 3 months of age and present with failure to thrive, persistent oral thrush, persistent diarrhea, severe *Candida* diaper dermatitis, recurrent severe infections, lymphadenopathy, and hepatosplenomegaly. Depending on the form of SCID, the infant may or may not have lymphopenia. They are prone to sepsis and opportunistic infections because they are unable to fight infections even if their B cells are present. They are susceptible to multiple pathogens such as *Candida albicans* or *Pneumocystis jiroveci,* viruses such as CMV or herpes virus, and common childhood diseases such as varicella and measles. Because of their severity, early recognition of infants with SCIDs is important and has led to the development of SCID screening of newborns. Persons with a positive family history can be tested antenatally if the genetic mutation is known, or screening can occur at birth. Newborn screening for SCIDs is currently conducted population-wide in 23 states, the District of Columbia, and the Navajo Nation.

Infants suspected of having SCID should be placed in protective isolation, and all staff involved in the infants' care should be vigilant about following proper handwashing procedures and avoiding exposure to infections. The practice of avoiding other children is recommended when they are able to be discharged home. Breast feeding is encouraged if the mother is CMV negative and without active infection. Nasogastric or parenteral nutrition may be needed for adequate nutrition. Curative therapy involves hematopoietic stem cell transplantation with either umbilical cord blood or bone marrow preferably from HLA-haploidentical siblings. Other therapies include chemotherapy conditioning regimens, long-term immunoglobulin replacement therapy every 2 to 3 weeks, enzyme replacement therapy, and gene therapy.

Wiskott–Aldrich Syndrome

Etiology and pathogenesis. Wiskott–Aldrich syndrome is an X-linked immunodeficiency disorder occurring only in males. It is caused by a mutation of the WAS protein *(WASp)* gene and has been mapped to the short arm of the X chromosome at Xpll.23. The *WASp* gene is involved in cytoplasmic signaling, immune cell function, and regulation of the actin cytoskeleton in hematopoietic cells. The *WASp* deficiency upsets B-cell homeostasis with a decrease in mature peripheral B-cell subsets, which causes a dysfunction in B cells and abnormal serum immunoglobulins. Usually there is decreased IgM level, elevated IgE and IgA levels, and normal-to-low IgG level. This antibody variability is the result of increased antibody catabolism. Other laboratory findings include low-to-absent isohemagglutinins and low platelet count secondary to increased platelet destruction by the spleen due to abnormalities in platelet size and shape. Defects in peripheral self-tolerance are due to dysfunctions in regulatory and effector T cells. Affected infants have particular difficulty mounting immune responses to polysaccharide-encapsulated bacterial antigens.

Clinical manifestations and treatment. Wiskott–Aldrich syndrome is clinically characterized by the presence of microthrombocytopenia, eczema, and increased susceptibility to developing infection and malignancies. Affected children can develop bleeding manifested as bloody diarrhea, or cerebral hemorrhage and infections due to polysaccharide-encapsulated bacteria (*S. pneumoniae* and *H. influenza*) including pneumonia, meningitis, otitis media, and sepsis. They can also develop systemic autoimmunity, which increases the severity of their disease. The average age of these infants at death is 3.5 years without treatment.

Infants with Wiskott–Aldrich syndrome are treated with antibody replacement therapy and antibiotic therapy. Splenectomy, bone marrow transplantation, stem cell transplantation, and gene therapy are options used to manage this disorder in affected children.

T-Cell Disorders

22q11.2 Deletion Syndrome (DiGeorge Syndrome)

Etiology and pathogenesis. 22q11.2 deletion syndrome, also known as *DiGeorge syndrome,* is an autosomal-dominant developmental T-cell disorder due to a hypoplastic or aplastic thymus gland. The development of this syndrome is caused by a chromosomal 22q11.2 deletion (del 22q11) that results in defective fetal development. The abnormal chromosome is usually inherited from the mother. It is the most common microdeletion syndrome with an estimated prevalence of 1 in 4000 live births. Male and female children are equally affected by this condition. The immunodeficiency seen in this syndrome is due to total or partial loss of thymus gland function where T-cell maturation occurs. As a result, T cells are deficient to varying degrees. B cells are usually normal in number and function. Often the immune system impairment is minimal with normal-to-decreased numbers of T cells. In those patients with severe impairment due to absent T cells secondary to an aplastic thymus gland, B-cell function is abnormal as well. These patients have severe immunodeficiency that resembles SCID. There may also be associated genetic modifiers that vary in clinical presentation.

Clinical manifestations and treatment. Because 22q11.2 deletion syndrome is a congenital disorder of fetal organ development, it is often associated with other congenital problems, such as cardiac and great vessel anomalies, hypoparathyroidism with hypocalcemia, hypothyroidism, esophageal atresia or reflux, urogenital anomalies, growth restriction, developmental delay, psychiatric or behavioral problems, microcephaly, and dysmorphic facial features, including mandibular hypoplasia, short forehead, and protuberant or low-set ears. If the individual has minimal loss of thymus function, recurrent infections may not be a significant problem. However, recurrent infections occur in many to most patients. This increases the risk of fetal loss or infant death. For these children, thymic transplantation or bone marrow transplantation has been helpful in reestablishing T-cell populations. Specific treatments and therapies are needed for the management of these patients, particularly those with cardiac defects and hypocalcemia with seizures.

Chronic Mucocutaneous Candidiasis Disease

Etiology and pathogenesis. Chronic mucocutaneous candidiasis disease (CMCD) is a T-cell disorder caused by an autosomal-dominant mutation affecting STAT 1 GOF with impaired-to-absent IL-17. CMCD results from the deficiency of the cytokine IL-17 or IL-17F, leading to selective susceptibility to fungal infections and *Staphylococcus aureus.* There is variability in impairment from complete to partial with varying cytokine activity. As a result of these deficiencies, CMCS is characterized by a selective deficiency of cell-mediated immunity against *Candida albicans* and *S. aureus.* In this case T cells do not produce the correct cytokines needed for the cell-mediated immunity to *C. albicans.* This causes persistent or recurrent isolated skin, nail, and mucous membrane infections. B-cell and T-cell functions are usually normal, except for the inability of the T cells to respond to *Candida* infections. Occasionally IgA or other antibody levels may be affected.

Clinical manifestations and treatment. The goal of treatment is to reduce the severity of skin and mucous membrane infections and to decrease the disfigurement from infection and scarring. Treatment involves antifungal therapy or antibiotic therapy. Gene therapies are being investigated.

B-Cell Disorders

IgA Deficiency

Etiology and pathogenesis. The most common B-cell PID is selective IgA deficiency (SIgAD). The incidence of this disorder varies geographically, affecting 1 in 600 persons in Caucasian populations and 1 in 2600 among Asian populations. It is a B-cell disorder characterized by either apoptosis of plasma cells or failure of IgA-bearing lymphocytes to become plasma cells. The result is the lack of IgA antibodies in the serum. Genetically, it can be an autosomal-recessive or autosomal-dominant disease. The B-cell level is normal, but there is a lack of B-cell response to cytokines (IL-4, IL-6, IL-7, IL-10). Serum levels of IgM and IgG, as well as T-cell numbers, are normal.

Clinical manifestations and treatment. People with this disorder may exhibit no symptoms. However, when symptoms occur, they are prone to respiratory, gastrointestinal, and genitourinary tract infections. They tend to have many autoantibodies (including anti-IgA antibodies) with a high incidence of allergic, vascular, endocrine, and collagen autoimmune diseases. They often react to cow's milk, and inflammatory bowel conditions such as celiac disease can develop. Treatment includes prevention of infection and management of infection with appropriate antibiotics. SIgAD patients are at risk of developing anti-IgA antibodies when exposed to blood or blood products. As a result, exogenous IgA replacement is contraindicated. When blood is needed, washed and packed RBCs with volume expanders are used to avoid serum antibody exposure.

X-Linked Agammaglobulinemia

Etiology and pathogenesis. X-linked agammaglobulinemia (XLA) (also known as *Bruton X-linked agammaglobulinemia*) is a B-cell genetic disorder caused by a lack of normal B-cell development in the bone marrow. The disorder is linked to a mutation of the gene *btk* (Bruton tyrosine kinase) located on the long arm of the X chromosome at position Xq21.3 to Xq22. It occurs in about 1:200,000 male births. In 1952 it was the first genetic immunodeficiency disorder identified. Females are carriers. This mutation occurs in a cytoplasmic signal-transducing molecule encoded by the *btk* gene, which results in decreased or absent B cells. The low B-cell numbers cause decreased or absent serum concentrations of IgG, IgA, or IgM immunoglobulins. The number of T cells is normal. The thymus functions normally. Plasmocytes are absent, and reticuloendothelial tissue and lymphoid organs (e.g., tonsils, spleen, Peyer patches, lymph nodes) are poorly developed. The disease is characterized by recurrent bacterial infection and profound hypogammaglobulinemia resulting from decreased numbers of circulating B cells. Male infants affected by this disorder are typically diagnosed after the infant reaches 4 months of age because of passive maternal IgG protection.

Clinical manifestations and treatment. Patients exhibit failure to thrive and frequent serious infections. The infections are most often attributable to *Haemophilus influenzae* and *S. pneumoniae,* which cause pneumonia, otitis media, meningitis, sinusitis, and septicemia. The recurrent infections can lead to tissue destruction and injury. Many children die before the age of 6. If the child survives to adulthood, life expectancy is decreased. Chronic lung disease, large joint arthritis, and the development of B-cell malignancies occur in adults with XLA. Treatment of this disorder includes implementation of antibiotic therapy and prophylactic antibiotics, as well as monthly administration of immunoglobulin infusions. However, passive immunotherapy is not always effective. Currently, curative hematopoietic stem cell–based gene therapy is advocated to correct B-cell and myeloid deficiencies. Children who have been diagnosed with XLA should not be immunized with live virus vaccine due to the risk of the development of severe infection.

Transient Hypogammaglobulinemia

Transient hypogammaglobulinemia of infancy is a self-limiting condition in which the infant is slow to acquire normal immunoglobulin levels. The infant experiences a lengthened period of low IgG, IgM, and IgA levels after birth. Normal IgG production begins only after 2 months of life and increases slowly as maternal immunoglobulins are metabolized. Affected infants usually demonstrate normal immunoglobulin levels and immune system function by between 18 and 30 months. During the period of low antibody levels, they are more susceptible to infections, particularly respiratory tract infections.

Common Variable Immunodeficiency Disease

Common variable immunodeficiency disease (CVID) is a B-cell disorder characterized by low titers of immunoglobulins, particularly IgG and IgA, as well as abnormal specific antibody responses. The primary defect is the inability of B cells to differentiate into plasma cells. In most cases, there are multiple mutations in several genes that encode the production of antibody subclasses and cytokines. There are usually normal numbers of B cells. The onset of CVID disease is variable from early childhood to the second or third decade of life. Males and females are equally affected with an incidence of 1 in 30,000 people.

Clinical manifestations and treatment. CVID is characterized by recurrent and severe infections, particularly respiratory tract disorders, including sinusitis, otitis, laryngitis, and pneumonia. The most frequent pathogens involved are encapsulated bacteria such as *S. pneumoniae* and *H. influenzae.* Infections of the gastrointestinal tract also occur from *Giardia* and *Salmonella.* There is an increased incidence of associated autoimmune disorders, including autoimmune thrombocytopenia purpura, arthritis, or thyroiditis; malignancies such as lymphoma, gastric, or colon cancer; granulomas in multiple locations; and gastrointestinal diseases, including colitis. Treatment involves IV or subcutaneous administration of immunoglobulin every 2 to 4 weeks, use of corticosteroids, and prompt treatment with antibiotics as needed. Occasionally stem cell transplantation is used in treatment.

SECONDARY IMMUNODEFICIENCY DISORDERS

A number of physical, psychosocial, nutritional, environmental, and pharmacologic factors can singly or in combination lead to the

development of secondary immunodeficiency disorders. Many of these linkages are discussed in Chapter 8.

The direct and indirect linkages between the brain and the endocrine and immune systems are well known. As a result, excessive or defective neuroendocrine responses can lead to disease. For example, an excessive neuroendocrine response to stress with increased secretion of corticosteroids boosts a person's susceptibility to infectious agents and tumors but enhances resistance to autoimmune disease. On the other hand, a defective neuroendocrine response to stress with low corticosteroid levels enhances autoimmune disease and inhibits infections and tumors. Individuals experiencing physical and psychosocial stress, decreased social support, depression, and bereavement show decreased immune system functioning.

Morbidity and surgery also affect the function of the immune system. After surgery, T-cell and B-cell numbers decrease. This temporary deficiency can last up to 1 month and is most likely a result of the stress of surgery. Some types of surgery, such as splenic surgery, actually reduce the effectiveness of the immune system. Removal of the spleen reduces serum IgM and the antibody response to encapsulated bacteria (e.g., *S. pneumoniae, H. influenzae, S. aureus*). Disease states such as diabetes mellitus, drug- or alcohol-induced cirrhosis, severe burns, severe trauma, sickle cell anemia, malignancies, and severe infections are associated with secondary immune deficiencies. For example, when blood glucose level rises in patients with diabetes mellitus, white blood cell response to infection declines.

Pregnancy requires immune adaptation in order to be successful. During pregnancy, many factors are released from the placenta influencing the immune system, including B-cell and T-cell function and cytokine balance. The placenta functions to regulate and coordinate all of the components of the complex maternal and fetal immune interface such as the secretion of cytokines: proinflammatory and antiinflammatory.[34] Pregnancy's proinflammatory state requires careful coordination, and failure to adapt can lead to spontaneous abortion or preeclampsia. For example, exosomes (syncytiotrophoblast microvesicles) are shed from the placenta into the maternal circulation. These exosomes have both physiologic and immunologic functions.

A number of pharmaceuticals affect the functioning of the immune system. Cytotoxins and other cancer pharmacotherapeutic drugs cause a state of generalized immunosuppression, including growth retardation, susceptibility to infection, impaired wound healing, and hypertension. For example, methotrexate is a phase-specific cytotoxin in which cells are killed only if they are in the S, or DNA-synthetic, phase. Cyclophosphamide is toxic to cells in any mitotic phase, although it is better at killing active cells. Anesthetics (e.g., halothane, cyclopropane, nitrous oxide, ether), alcohol, antibiotics, antithyroids, anticonvulsants, antihistamines, and steroids decrease cellular or humoral immunity by various methods. For example, chronic nitrous oxide toxicity leads to cell-mediated immune deficits. Therapeutic radiation (x-rays) also affects the immune system by destroying rapidly proliferating cells. When T-cell and B-cell clones are needed, irradiation eliminates these cells, thus blunting or reducing the effectiveness of the body's response.

A number of studies have linked immune system competency and nutritional status. Malnutritional states can lead to protein depletion, as well as carbohydrate, lipid, vitamin, and mineral deficiencies. Protein and calorie depletion causes T-cell reductions and dysfunction. Antibodies are composed of proteins, levels of which are also low in a state of depletion. Low levels of zinc, an enzyme cofactor needed for lymphocyte function, as well as low levels of folic acid and vitamins B_6, A, D, and E, can result in T-cell and B-cell dysfunction. Obesity, low physical activity, and hyperlipidemia can cause chronic systemic inflammation due to local immune response in visceral adipose tissue. Obesity has been found to cause up-regulation of a wide variety of proinflammatory cytokines, including IL-5, IL-10, IL-12, and IFN-γ. Nutritional balance is best for the development of all stem cells, the recognition and processing of antigens, and the attainment of optimal immune cell function.

In the elderly, immune system function is altered with decreased response to antigenic stimulation. This decrease is due to increased CD8+ T cells that lack expression of the CD28 molecule. This subset (CD28-CD8+) does not respond well to antigenic stimulation. Immune system cells in the elderly are not able to proliferate or reproduce as effectively as in younger persons, making them less able to respond to "new" antigenic stimuli. Although the total number of T cells remains the same, T-cell function is decreased. T cells are less capable of proliferating and have decreased cytotoxicity. Antibody production also decreases. A rise is seen in autoantibody production, which may influence the increase in autoimmune disease in the elderly.

KEY POINTS

- Primary deficiencies in immune function may be from congenital, genetic, or acquired defects that directly affect immune cell function.
- Secondary deficiencies are conditions that impair immune function as a result of other nonimmune system disorders, such as poor nutrition, pregnancy, stress, or drugs, that secondarily suppress immune function.
- Primary genetic immunodeficiency disorders are fairly common, with most causing moderate immune impairment that may not be diagnosed. Severe congenital immunodeficiency disorders are less common, but clinically significant. Impairment in T cells and B cells results in severe combined immunodeficiency (SCID). Functional B and T lymphocytes are lacking, and infants with SCID easily succumb to sepsis and opportunistic infections. Other types of primary immunodeficiency disorders (PIDs) affect a particular cell type: DiGeorge syndrome (now called *22q11.2 deletion syndrome*) occurs with T-cell agenesis related to a lack of thymus function; chronic mucocutaneous candidiasis is caused by abnormal T cells that cannot respond to *Candida;* and selective immunoglobulin A (IgA) deficiency is caused by B-cell abnormality.
- Problems in neuroendocrine and immune system interactions are a cause of secondary immunodeficiencies. Excessive neuroendocrine response to stress with increased corticosteroid production increases susceptibility to infection.
- Medications such as cytotoxins and other cancer pharmacotherapeutic drugs cause generalized secondary immunosuppression. However, other medications, such as anesthetics, alcohol, antibiotics, and steroids, also affect the immune response and can lead to secondary immunosuppression.
- Malnutrition, a major cause of immune system dysfunction, leads to lymphocyte dysfunction and altered stem cell development.

SUMMARY

Human beings live in internal and external environments teeming with antigens capable of producing immunologic responses. Contact with an antigen, or antigens, usually leads to induction of a normal protective immune response. However, some individuals experience disease and tissue damage caused by either excessive or deficient immune responses.

Excessive immune reactions are common and involve a complex interplay between antigen and components of the immune system. Autoimmune disease is a type of excessive immune reaction in which the immune system reacts against the body's own cells, such as in rheumatoid arthritis or Graves disease. The tissues that are affected in

excessive immune reactions depend on the type of antigen, the type of antigen exposure, and the degree of immune responsiveness. Hypersensitivity disorders are differentiated by the cell type involved and the time course of the reaction. For example, type I hypersensitivity is a rapid response caused by a host of lethal chemicals generated by but involving only one antibody, IgE, and one effector cell, the mast cell. It is an antigen–antibody reaction causing the release of potent chemicals that can lead to extreme, even life-threatening, anaphylaxis reactions in susceptible individuals.

Deficient immune reactions are caused by genetic or embryonic defects. The lack of or dysfunction of T cells, B cells, and antibodies can lead to lethal or recurrent infections or other diseases that severely limit the patient's ability to interact with the environment. The degree of immune system dysfunction or deficiency varies greatly between individuals due to the complex interaction between the various components of the immune system. For example, the most common type of deficient immune response is SIgA. As a result of variable IgA production by B cells, the range of symptoms varies from none to recurrent infections with autoimmune disease.

Research into the complex interaction between genetic factors, environmental factors, and the immune system is ongoing. We are challenged to expand our understanding and knowledge of immune components, responses, and their physiologic or pathologic effects on individuals, families, and society as a whole.

RESOURCES

Autoimmune and Hypersensitivity Disorders

Aitken SL, Tichy EM: Rh(O)D immune globulin products for prevention of alloimmunization during pregnancy. *Am J Health Syst Pharm* 72(4):267–276, 2015.

Antonelli A, et al: Autoimmune thyroid disorders. *Autoimmun Rev* 14(2): 174–180, 2015.

Berrih-Aknin S, Le Panse R: Myasthenia gravis: a comprehensive review of immune dysregulation and etiological mechanisms. *J Autoimmun* 52:90–100, 2014.

Bomback AS: Anti-complement therapy for glomerular diseases. *Adv Chronic Kidney Dis* 21(2):152–158, 2014.

Chen CY, et al: Induction of interleukin-9-producing mucosal mast cells promotes susceptibility to IgE-mediated experimental food allergy. *Immunity* 43(4):788–802, 2015.

Dong YH, Fu DG: Autoimmune thyroid disease: mechanism, genetics and current knowledge. *Eur Rev Med Pharmacol Sci* 18:3611–3618, 2014.

Gardner K, et al: How we treat delayed haemolytic transfusion reactions in patients with sickle cell disease. *Br J Haematol* 170(6):745–756, 2015.

Gilhus EN, Verschuuren JJ: Myasthenia gravis: subgroup classification and therapeutic strategies. *Lancet Neurol* 14(10):1023–1036, 2015.

Magna M, Pisetsky DS: The role of cell death in the pathogenesis of SLE: is pyroptosis the missing link? *Scand J Immunol* 82(3):218–224, 2015.

Marson A: Genetic basis of autoimmunity. *J Clin Invest* 125(6):2234–2241, 2015.

Mawson AR, Eriator I, Karre S: Stevens-Johnson syndrome and toxic epidermal necrolysis (SJS/TEN): could retinoids play a causative role? *Med Sci Monit* 21:133–143, 2015.

Noutsios GT, Floros J: Childhood asthma: causes, risks, and protective factors; a role of innate immunity. *Swiss Med Wkly* 144:w14036, 2014.

Oettgen HC, Burton OT: IgE and mast cells: the endogenous adjuvant. *Adv Immunol* 127:203–256, 2015.

Ortiz RA, Barnes KC: Genetics of allergic diseases. *Immunol Allergy Clin North Am* 35(1):19–44, 2015.

Pacheco-Tena C, Gonzalez-Chavez SA: The danger model approach to the pathogenesis of the rheumatic diseases. *J Immunol Res* 506089, 2015. May 2015.

Pavlos R, et al: T cell-mediated hypersensitivity reactions to drugs. *Annu Rev Med* 66:439–454, 2015.

Root-Bernstein R, Fairweather D: Complexities in the relationship between infection and autoimmunity. *Curr Allergy Asthma Rep* 14(1):407–416, 2014.

Saggini A, Chimenti S, Chiricozzi A: IL-6 as a druggable target in psoriasis: focus on pustular variants. *J Immunol Res* 964069, 2014. [epub July 2014].

Uzzaman A, Cho SH: Chapter 28: classification of hypersensitivity reactions. *Allergy Asthma Proc* 33(Suppl 1):S96–S99, 2012.

Wang L, et al: Human autoimmune diseases: a comprehensive update. *J Intern Med* 278(4):369–395, 2015.

Zhang C, et al: Immunoglobulin A nephropathy: current progress and future directions. *Transl Res* 166(2):134–144, 2015.

Deficient Immune Responses

Abolhassani H, et al: Autoimmunity in patients with selective IgA deficiency. *J Investig Allergol Clin Immunol* 25(2):112–119, 2015.

Burnside RD: 22q11.21 deletion syndromes: a review of proximal, central and distal deletions and their associated features. *Cytogenet Genome Res* 146:89–99, 2015.

Burton GJ, Jauniaux E: What is the placenta? *Am J Obstet Gynecol* 213(4 Suppl):S6–S8, 2015.

Carmina Castiello M: Wiskott-Aldrich syndrome protein deficiency perturbs the homeostasis of B-cell compartment in humans. *J Autoimmun* 50(100):42–50, 2014.

Du HQ, et al: Effects of Wiskott-Aldrich syndrome protein deficiency on IL-10-producing regulatory B cells in humans and mice. *Scand J Immunol* 81(6):483–493, 2015.

Ifrim DC, et al: Defective trained immunity in patients with STAT-1-dependent chronic mucocutaneaous candidiasis. *Clin Exp Immunol* 181(3):434–440, 2015.

Jorgensen GH, et al: Clinical symptoms in adults with selective IgA deficiency: a case-control study. *J Clin Immunol* 33(4):742–747, 2013.

Kohn DB: Eliminating SCID row: new approaches to SCID. *Hematology Am Soc Hematol Educ Program* 1:475–480, 2014. December 5, 2014.

Kwan A, et al: Newborn screening for severe combined immunodeficiency in 11 screening programs in the United States. *JAMA* 312(7):729–738, 2014.

Lanternier F, et al: Primary immunodeficiencies underlying fungal infections. *Curr Opin Pediatr* 25(6):736–747, 2013.

Leitner J, et al: CD58/CD2 is the primary costimulatory pathway in human CD28-CD8+ T cells. *J Immunol* 195(2):477–487, 2015.

Ochs HD: Common variable immunodeficiency (CVID): new genetic insight and unanswered questions. *Clin Exp Immunol* 178(Suppl 1):5–6, 2014.

Rivers L, Gaspar HB: Severe combined immunodeficiency: recent developments and guidance on clinical management. *Arch Dis Child* 100:667–672, 2015.

Schmidt FM, et al: Inflammatory cytokines in general and central obesity and modulating effects of physical activity. *PLoS ONE* 10(3):e0121971, 2015.

Sokolov DI, et al: Changes in functional activity of JEG-3 trophoblast cell line in the presence of factors secreted by placenta. *Arch Med Res* 46(4):245–256, 2015.

Zheng J, et al: Gain-of-function STAT1 mutations impair STAT3 activity in patients with chronic mucocutaneous candidiasis (CMC). *Eur J Immunol* 45(10):2834–2846, 2015.

11

Malignant Disorders of White Blood Cells

Marie L. Kotter and Jacquelyn L. Banasik

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- How do the various types of leukemia, lymphoma, and plasma cell myelomas differ based on the type of malignant transformation?
- How do the clinical presentations, prognosis, and management of types of acute and chronic leukemia differ?
- Why are malignant disorders of white blood cells commonly associated with bone marrow depression?
- How is Hodgkin disease clinically and histologically differentiated from other types of lymphoma?
- What is the purpose and process of staging procedures for lymphomas?
- What clinical and laboratory findings would suggest a diagnosis of plasma cell myeloma?

CHAPTER OUTLINE

Classification of Hematologic Neoplasms, 215
Etiology of Myeloid and Lymphoid Neoplasms, 216
General Principles of Management, 217
Diagnosis of Hematologic Neoplasms, 217
Principles of Treatment, 218
Prevention and Management of Complications, 219
Myeloid Neoplasms, 221
Chronic Myeloid Leukemia, 221
Acute Myeloid Leukemia, 222
Lymphoid Neoplasms, 223
Chronic Lymphoid Leukemia, 223
Acute Lymphoblastic Leukemia/Lymphoma, 223
Hairy Cell Leukemia, 224
Plasma Cell Myeloma (Multiple Myeloma), 224
Hodgkin Disease, 227
B-Cell, T-Cell, and NK-Cell Lymphoma (Non-Hodgkin), 229

Leukemia, lymphoma, and plasma cell myeloma (multiple myeloma) are common neoplastic disorders of the bone marrow and lymphoid tissues. Depending on the location and specific types of white blood cells involved, these malignancies can be further divided into a number of specific subtypes. Leukemias can be conceptualized as circulating tumors that are disseminated from the beginning of the disease process and primarily involve the blood and bone marrow. Lymphoma tends to localize in lymph tissues, but is often disseminated to other sites at the time of diagnosis. Plasma cell myeloma is a malignant transformation of B-lymphocyte plasma cells and has a predilection to form localized tumors in bony structures.

Malignancies of the blood-forming tissues and lymphatic structures often present with nonspecific symptoms. Malaise, weakness, unexplained fever, night sweats, and recurrent infections should raise suspicion of malignancy. Enlarged, nontender lymph nodes (lymphadenopathy) are a common finding in lymphoma and some leukemias. Often, white blood cell malignancies are found by chance during routine assessment of the complete blood cell count (CBC). A very high total white blood cell count or the presence of abnormal cell types should precipitate an assessment for hematologic cell malignancy. In general, earlier detection of malignancy is associated with a better prognosis for cure.

CLASSIFICATION OF HEMATOLOGIC NEOPLASMS

Various classification schemes have been used to group hematologic neoplasms, with clinicians favoring schemes that use clinical findings and pathologists preferring morphologic criteria. With the advent of technologies to identify specific genetic alterations and molecular characteristics of neoplastic cells, the traditional classification systems have become less useful. However, many clinicians and organizations, such as the American Cancer Society, continue to use traditional groupings to collect statistics and to provide information to the public. The approach used in this chapter incorporates the most recent World Health Organization (WHO) classifications for hematologic neoplasms and includes common clinical terminology. A major force behind the adoption of the WHO classification is the recognition that lymphoid leukemias and lymphomas are not separate disorders, but represent different stages of the same biological disease. Thus the major categories of the WHO system are

based on the cell type of the neoplasm rather than its location in the body. Neoplasms involving cells of the myeloid lineage (Box 11.1) are separated from those of the lymphoid lineage (Box 11.2). The myeloid lineage includes red blood cells, platelets, monocytes, and granulocytes; the lymphoid lineage includes B cells, T cells, and natural killer (NK) cells (Fig. 11.1). There are four major categories of myeloid neoplasms: myeloproliferative diseases, myelodysplastic/proliferative diseases, myelodysplastic syndromes, and acute myeloid leukemia (AML). There are three major categories of lymphoid neoplasms: B-cell neoplasm, T-cell and NK-cell neoplasm, and Hodgkin disease. The term *non-Hodgkin lymphoma* is still in clinical usage and refers to lymphomas of B-cell, T-cell, and NK-cell origin. Non-Hodgkin lymphoma includes such a large and diverse group of malignancies that it has little relevance to prognosis or treatment. The WHO classification does not use this term. There are many etiologic, pathogenic, and treatment similarities among the hematologic malignancies, and these are addressed in a general way first, followed by sections concentrating on specific diseases.

ETIOLOGY OF MYELOID AND LYMPHOID NEOPLASMS

As in other malignant processes, the exact cause of hematologic neoplasms is unknown. The basic mechanism of malignant transformation involves mutation of cells, which disrupts growth control and differentiation pathways. These processes are thought to be similar to those described for solid tumors (see Chapter 7). A number of specific genetic mutations have been found for several hematologic malignancies. In some cases, these specific mutations indicate prognosis of the disease, and in others they may provide a target for therapy. Monoclonal antibodies have been developed to attach to products of some of these mutated genes, thus allowing therapy to be more targeted to the abnormal cells.

Viruses have long been suspected as mutagenic agents in some neoplasms, particularly retroviruses and herpesviruses. Close associations have been found between a small number of viruses and particular malignancies. For example, human T-cell leukemia virus (HTLV-1) is

BOX 11.1 WHO Classification of Myeloid Neoplasms

Myeloproliferative Neoplasms (MPN)
Chronic myeloid leukemia (CML), *BCR-ABL1*$^{+}$
Chronic neutrophilic leukemia (CNL)
Polycythemia vera (PV)
Primary myelofibrosis (PMF)
  PMF, prefibrotic/early stage
  PMF, overt fibrotic stage
Essential thrombocythemia (ET)
Chronic eosinophilic leukemia, not otherwise specified (NOS)
MPN, unclassifiable
Mastocytosis

Myeloid/Lymphoid Neoplasms With Eosinophilia and Rearrangement of *PDGFRA, PDGFRB,* or *FGFR1,* or with *PCM1-JAK2*
Myeloid/lymphoid neoplasms with *PDGFRA* rearrangement
Myeloid/lymphoid neoplasms with *PDGFRB* rearrangement
Myeloid/lymphoid neoplasms with *FGFR1* rearrangement
Provisional entity: Myeloid/lymphoid neoplasms with PCM1-JAK2

Myelodysplastic/Myeloproliferative Neoplasms (MDS/MPN)
Chronic myelomonocytic leukemia (CMML)
Atypical chronic myeloid leukemia (aCML), *BCR-ABL1*$^{-}$
Juvenile myelomonocytic leukemia (JMML)
MDS/MPN with ring sideroblasts and thrombocytosis (MDS/MPN-RS-T)
MDS/MPN, unclassifiable

Myelodysplastic Syndromes (MDS)
MDS with single lineage dysplasia
MDS with ring sideroblasts (MDS-RS)
  MDS-RS and single lineage dysplasia
  MDS-RS and multilineage dysplasia
MDS with multilineage dysplasia
MDS with excess blasts
MDS with isolated del(5q)
MDS, unclassifiable
Provisional entity: Refractory cytopenia of childhood
Myeloid neoplasms with germ line predisposition

Blastic Plasmacytoid Dendritic Cell Neoplasm

Acute Myeloid Leukemia (AML) and Related Neoplasms
AML with recurrent genetic abnormalities
AML with t(8;21)(q22;q22.1);*RUNX1-RUNX1T1*
AML with inv(16)(p13.1q22) or t(16;16)(p13.1;q22);*CBFB-MYH11*
APL with *PML-RARA*
AML with t(9;11)(p21.3;q23.3);*MLLT3-KMT2A*
AML with t(6;9)(p23;q34.1);*DEK-NUP214*
AML with inv(3)(q21.3q26.2) or t(3;3)(q21.3;q26.2); *GATA2, MECOM*
AML (megakaryoblastic) with t(1;22)(p13.3;q13.3);*RBM15-MKL1*
Provisional entity: AML with BCR-ABL1
AML with mutated *NPM1*
AML with biallelic mutations of *CEBPA*
Provisional entity: AML with mutated RUNX1
AML with myelodysplasia-related changes
Therapy-related myeloid neoplasms
AML, NOS
  AML with minimal differentiation
  AML without maturation
  AML with maturation
  Acute myelomonocytic leukemia
  Acute monoblastic/monocytic leukemia
  Pure erythroid leukemia
  Acute megakaryoblastic leukemia
  Acute basophilic leukemia
  Acute panmyelosis with myelofibrosis
Myeloid sarcoma
Myeloid proliferations related to Down syndrome
  Transient abnormal myelopoiesis (TAM)
  Myeloid leukemia associated with Down syndrome

Acute Leukemias of Ambiguous Lineage
Acute undifferentiated leukemia
Mixed phenotype acute leukemia (MPAL) with t(9;22)(q34.1;q11.2); *BCR-ABL1*
MPAL with t(v;11q23.3); *KMT2A* rearranged
MPAL, B/myeloid, NOS
MPAL, T/myeloid, NOS

From Arber DA et al: The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* 2016;127(20):2391–2404.

BOX 11.2 WHO Classification of Lymphoid Neoplasms

Precursor Lymphoid Neoplasms
Precursor B lymphoblastic leukemia/lymphomas (precursor B-cell ALL)
Precursor T lymphoblastic leukemia/lymphoma

Mature (Peripheral) B-Cell Neoplasms*
B cell chronic lymphocytic leukemia/small lymphocytic lymphoma
B cell prolymphocytic leukemia
Lymphoplasmacytic lymphoma
Splenic marginal zone B-cell lymphoma (6 villous lymphocytes)
Hairy cell leukemia variant
Plasma cell myeloma/plasmacytoma
Waldenström macroglobulinemia
Heavy-chain diseases
Extranodal marginal zone B-cell lymphoma of MALT type
Nodal marginal zone B-cell lymphoma (6 monocytoid B cells)
Follicular lymphoma
Mantle cell lymphoma
Epstein-Barr virus positive DLBCL of the elderly
Large B-cell lymphoma arising in HHV8-associated multicentric Castleman disease
Primary mediastinal (thymic) large B-cell lymphoma
Primary effusion lymphoma
Burkitt lymphoma/Burkitt cell leukemia

Mature T-Cell and NK-Cell Neoplasms
Mature (peripheral) T-cell neoplasms
  T-cell prolymphocytic leukemia
  T-cell granular lymphocytic leukemia
Aggressive NK-cell leukemia
Adult T-cell lymphoma/leukemia (HTLV1$^+$)
Extranodal NK/T-cell lymphoma, nasal type
Enteropathy-type T-cell lymphoma
Hepatosplenic γδ T-cell lymphoma
Subcutaneous panniculitis-like T-cell lymphoma
Mycosis fungoides
Sézary syndrome
Primary cutaneous anaplastic large cell lymphoma
Peripheral T-cell lymphoma, not otherwise characterized
Angioimmunoblastic T-cell lymphoma
Anaplastic large cell lymphomas (ALK positive or ALK negative)

Hodgkin Lymphoma (Hodgkin Disease)
Nodular lymphocyte predominance Hodgkin lymphoma
Classical Hodgkin lymphoma
Nodular sclerosis Hodgkin lymphoma (Grades 1 and 2)
Lymphocyte-rich classical Hodgkin lymphoma
Mixed cellularity Hodgkin lymphoma
Lymphocyte depletion Hodgkin lymphoma

From Hoffbrand AV, Moss PAH: World Health Organization classification of tumours of the haematopoietic and lymphoid tissues. In *Essential haematology,* ed 6, Oxford, England, 2011, Blackwell, pp 428–429.
*B- and T/NK-cell neoplasms are grouped according to major clinical presentations (predominantly disseminated/leukemic, primary extranodal, predominantly nodal).

linked to the development of adult T-cell lymphoma/leukemia, and human immunodeficiency virus (HIV) is linked to B-cell lymphomas. Epstein–Barr virus (EBV) has been implicated in both Hodgkin disease and Burkitt lymphoma. Effective immune surveillance is thought to keep proliferation in check and prevent progression in immunocompetent individuals.

Radiation exposure is an important etiologic factor for leukemia and lymphoma. Because of the relatively high turnover of hematologic cells, they are more susceptible to radiation-induced damage than most other cell types. An acute whole-body dose of radiation, like that which occurs with nuclear explosions, is known to increase the risk of leukemia. In Japanese survivors of the atomic bomb, the estimated lifetime risk of leukemia is 0.85%, sixfold higher than the norm. There are substantial uncertainties about the risk of low-level, long-term exposure to radiation. The average annual exposure from usual sources, including cosmic rays and medical procedures, is very low and is estimated to account for less than 5% of leukemia cases.

Despite intensive scrutiny only a small number of chemicals have been shown unequivocally to increase the risk of hematologic malignancies. Benzene has been implicated in numerous studies, as has cigarette smoking. Other suggested carcinogens have failed to be confirmed, including exposure to hair dye, alcohol, and marijuana. On the other hand, a study from the Children's Cancer Group found a link between high maternal intake of products high in bioflavonoids (beans, fresh vegetables, and fruit) and an increased incidence of infant leukemia. These bioflavonoids were enzyme inhibitors (topoisomerase II inhibitors) that caused DNA cleavage and chromosome translocations. Many of the antineoplastic drugs used to treat cancers, especially the alkylating agents, are significant factors in the development of posttreatment hematologic neoplasia. Any drugs that suppress the bone marrow or immune function are also believed to predispose to the emergence of malignancies.

A number of disease conditions have been linked to the development of leukemia, although the mechanisms are unclear. A reduction or alteration in normal hematopoiesis, as occurs in such disorders as Fanconi anemia and aplastic anemia (see Chapter 13), is associated with a higher incidence of leukemia. A higher risk also has been noted in some genetic diseases, including Down syndrome and Klinefelter syndrome (see Chapter 6).

GENERAL PRINCIPLES OF MANAGEMENT

Diagnosis of Hematologic Neoplasms

Manifestations of hematologic neoplasms vary somewhat, depending on the cell type involved. Common manifestations are shown in Box 11.3. Clinical symptoms are related to bone marrow suppression and organ dysfunction secondary to leukemic infiltration. Bone marrow suppression results in varying degrees of leukopenia, anemia, and thrombocytopenia. These three deficiencies cause the most common clinical manifestations and may prompt the patient to seek care.

Anemia, with a hematocrit level of 25% to 30% or a hemoglobin level of 8 to 10 g/dL, may manifest with pallor, fatigue, malaise, shortness of breath, and decreased activity tolerance. The severity of symptoms is determined by the rate of red blood cell decrease, as well as the absolute deficiency. Chronically low hemoglobin and hematocrit values may be better tolerated than a drastic drop in these measurements. Depending on symptoms, transfusion may be indicated when the hematocrit level falls below 30%.

Thrombocytopenia, with a platelet count less than 20,000 cells/μL, can manifest as petechiae, easy bruising, bleeding gums, occult hematuria,

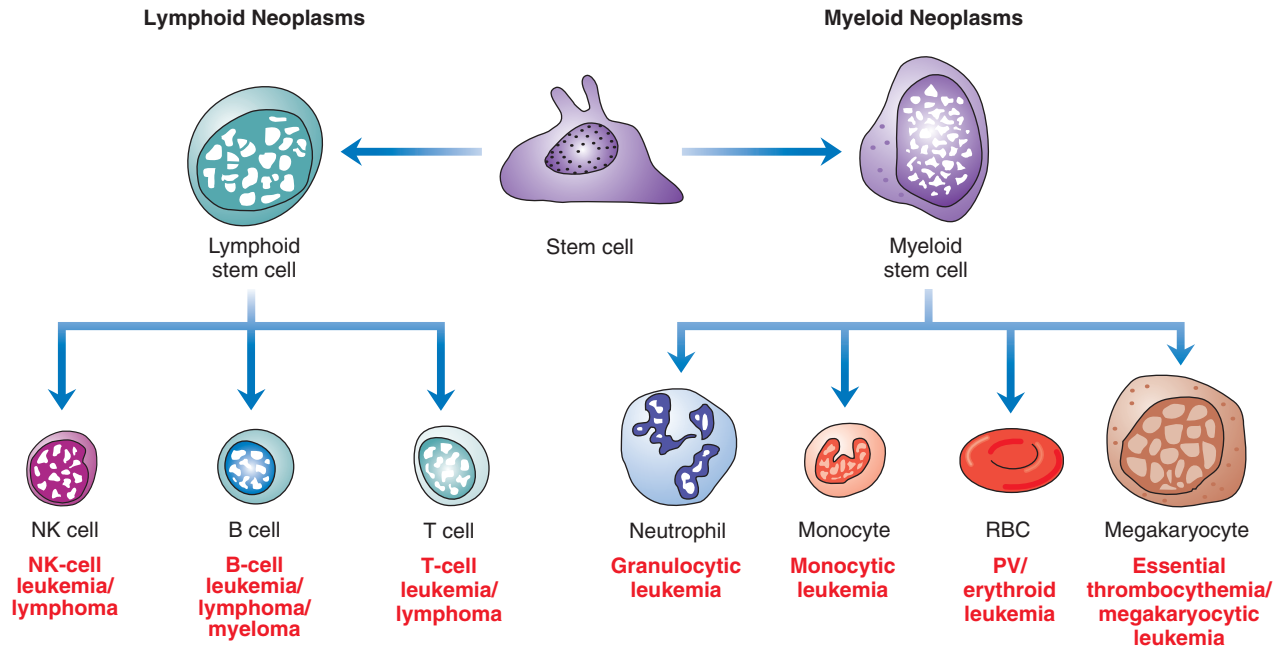


FIG 11.1 Division of hematologic neoplasms into myeloid and lymphoid lineages. *PV,* Polycythemia vera.

BOX 11.3 Common Manifestations of Hematologic Malignancies

History
Fever
Weight loss
Night sweats
Itching (pruritus)
Fatigue
Bone pain (sternum, tibia, femur, back)
Abdominal fullness
Bleeding episodes (epistaxis, menorrhagia)
Bruising, petechiae
Frequent infections
Headache, nausea, vomiting

Physical
Enlarged spleen
Enlarged liver
Enlarged lymph nodes
Hyperplasia of gums

Laboratory
Anemia or polycythemia
Thrombocytopenia or thrombocythemia
Leukopenia or leukocytosis
Blasts on peripheral blood smear
Elevated uric acid level
Elevated alkaline phosphatase level
Hypercalcemia

or retinal hemorrhages. Spontaneous intracranial bleeding can occur and may be fatal. In general, the risk of bleeding increases proportionately to the fall in platelet count. Platelet transfusion may be given when the risk of bleeding is high.

Insufficient numbers of functional leukocytes leave the patient at high risk for development of infection, and the CBC is routinely monitored. Neutropenia is an absolute neutrophil count less than 500 cells/μL, and an affected patient requires protective isolation (neutropenic precautions) to prevent infection. Infections may be caused by bacterial, viral, fungal, or protozoal organisms. Often, the microorganisms are of the opportunistic variety. That is, they are part of the patient's own flora, which normally do not cause disease unless the host's immune system becomes incompetent. It is very difficult to protect patients from their own flora, and infection is the most common cause of death in the immunocompromised leukemic patient. The presence of infection is suspected if fever develops. Infections are managed aggressively with antibiotic agents to prevent development of life-threatening sepsis.

Infiltrative manifestations include lymphadenopathy, joint swelling and pain, weight loss, anorexia, hepatomegaly, and splenomegaly. Sternal tenderness is frequently present in chronic myeloid leukemia (CML). Gingival hyperplasia occurs in AML. Meningeal involvement is frequently encountered in children with acute lymphoid leukemia (ALL). Central nervous system (CNS) infiltration can occur with any type of leukemia and may be difficult to manage because of the poor ability of chemotherapeutic agents to cross the blood–brain barrier. CNS involvement can present with increased intracranial pressure, seizures, or changes in mental ability. Increased intracranial pressure should be suspected in the leukemia patient who complains of nausea, vomiting, headache, and visual changes.

A key aspect of diagnosis is the evaluation of a peripheral blood sample. Blood cell number and morphologic evaluation are indicative; however, definitive diagnosis is usually made after bone marrow aspiration or lymph node biopsy. Malignant cells can be subtyped according to genetic and molecular characteristics to better determine prognosis and choice of treatment.

Principles of Treatment

To make informed treatment decisions, patients and their families need information about the nature and prognosis of their disease, as well as about the risks and benefits of various treatment options. Many treatment protocols are experimental, and the outcomes may be uncertain. Sometimes the side effects of treatment, as well as its limited efficacy, will weigh in favor of palliative care. Treatment decisions are complex

and stressful for all concerned. A great deal of support must be available during the diagnostic, treatment, and monitoring phases.

The management of hematologic malignancies relies primarily on the use of combination chemotherapy to eradicate malignant cells and stem cell transplant to rescue and restore bone marrow function. In some cases radiation and tissue-specific drug therapy may be indicated. Unfortunately, the treatment regimen usually causes many serious side effects that must be monitored and treated.

The goal of chemotherapy is to induce long-term remission, that is, the absence of any detectable neoplastic cells in the body. A complete remission (CR) is defined as a return to normal hematopoiesis with normal red blood cell, neutrophil, and platelet counts and no detectable neoplastic cells. For leukemia, the bone marrow must have less than 5% blasts, which are the most immature bone marrow cells, and be maintained for at least 4 weeks. CR is not synonymous with cure. Therefore most treatment protocols include several cycles of chemotherapy to eradicate the undetected cells. The choice of antineoplastic agents varies with the type of neoplasia and the stage of clinical disease. Most chemotherapeutic agents work by disrupting some aspect of DNA synthesis or cell replication and induce apoptosis (cell death; see Chapter 4). In general, rapidly dividing cells are more susceptible to apoptosis because they have less time for repair. Neoplasms with genetic defects that impair apoptotic pathways may be more difficult to eradicate and require more intense therapy. Unfortunately, these high doses are toxic to normal stem cells as well and can produce fatal bone marrow failure. Therefore to effect a cure, high-dose chemotherapy is often followed by bone marrow "rescue" with transplantation of functional stem cells.

Chemotherapy usually includes two or three treatment phases: (1) remission induction phase, (2) postremission or consolidation phase, and (3) remission maintenance phase. The aim of treatment during the remission induction phase is to eliminate all detectable neoplastic cells and achieve a CR. Postremission consolidation therapy begins after CR is attained in an attempt to eliminate the population of undetected cells that may have escaped initial induction phase treatment. Maintenance phase treatment is used in the management of some neoplasms to prolong the remission interval. Intermittent chemotherapy may be continued for 2 to 3 years after initial induction of remission. Drugs that target the neoplastic cells specifically, such as monoclonal antibodies or molecular therapies, are generally less toxic than other agents and may be used for long-term maintenance in patients with residual disease.

In children and adults, the CNS can act as a sanctuary for neoplastic cells in diseases such as ALL and AML. This makes conventional routes of chemotherapy unsuccessful, because they do not permit drugs to cross the blood–brain barrier efficiently. Chemotherapeutic agents administered into the cerebrospinal fluid via lumbar puncture (intrathecal route) can effectively eliminate leukemic cells in the CNS. This therapy carries significant risk for temporary or permanent neurologic damage. A number of different chemotherapeutic agents can be administered safely by the intrathecal route, including methotrexate.

Prevention and Management of Complications

Maintenance of adequate nutrition in patients with hematologic malignancy is a major challenge. Anorexia, weight loss, nausea, vomiting, and stomatitis are common findings, especially during the treatment phase. Children and adolescents receiving chemotherapy may experience significant growth delay, and measures to maintain protein and caloric intake are necessary. Newer antiemetic agents have been helpful in reducing nausea, vomiting, and anorexia associated with chemotherapy and should be considered in patients experiencing these symptoms.

Infection is the most troublesome complication for the patient who is immunosuppressed by either disease or treatments. Constant vigilance in prevention, early detection, and rapid management of infections can profoundly affect the outcome of chemotherapy. The length of time that a patient remains neutropenic can be shortened with the use of growth factors to stimulate bone marrow production of granulocytes.

Bone marrow transplantation has been an important part of the management of certain leukemias for many years. The intense chemotherapy used to induce remission can lead to bone marrow failure. Stem cells can be reintroduced into the host's bone marrow by bone marrow transplantation. The transplanted cells are given intravenously; they find their way to the host's bone marrow, where they establish residence and begin to produce functional white blood cells, red blood cells, and platelets. A close match between donor and host is necessary for a successful transplantation. Otherwise, the transplanted cells can mount an immune attack on the host's tissues—a life-threatening problem called *graft-versus-host disease* (see Chapter 10). In past years, the transplant was obtained by aspiration from the marrow of a suitable donor, and this may still be done in some cases. *Peripheral stem cell transplantation* allows stem cells to be harvested from the circulating bloodstream. This procedure can be used to collect stem cells from the patient's own blood to be stored and then reinfused after chemotherapy and irradiation. This type of transplant is called *autologous,* whereas a transplant from a closely matched relative is called *allogeneic* (Fig. 11.2). Use of autologous transplants eliminates the problem of graft-versus-host disease and reduces transplant-related mortality, but the potential for disease recurrence is higher than with allogeneic transplants.

It has been noted that in AML and CML, transplantation with allogeneic cells is much more successful in curing leukemia than is autologous transplantation. Transplanted cells in the allograft are thought to detect and kill leukemic cells in a process termed *graft versus leukemia.* Autologous transplants are appropriate in some cases, especially when a matched donor is not available, because they may extend life even though cure is unlikely. Autologous transplants are well tolerated and cause fewer complications than allografts. Methods to purify a patient's collected peripheral blood by selectively removing neoplastic cells are available to reduce the risk of reintroducing malignant cells during autologous transplantation. Increased availability of stem cell transplants allows patients to undergo more intensive chemotherapy, aimed at cure rather than palliation, based on the knowledge that bone marrow rescue is possible.

Anemia is a common complication of leukemia and chemotherapy. Red blood cell production by the bone marrow is suppressed, but the size and shape of red blood cells present in the blood are normal. This is called *normocytic,* normochromic anemia. Administration of erythropoietin growth factors can enhance red blood cell production and moderate anemic episodes. However, patients frequently require red blood cell transfusion therapy to maintain adequate red blood cell counts. Patients with frequent or significant bleeding episodes are also predisposed to severe anemia, and efforts to prevent bleeding will help minimize anemia.

Platelet deficiency (thrombocytopenia) with resultant hemorrhage can be a life-threatening complication of leukemia and chemotherapy. In patients at high risk of bleeding, fresh frozen plasma or pooled platelets may be given to inhibit bleeding. Patients must be protected from trauma and may be placed on activity restrictions.

Pain is a common complication of the diagnostic and treatment protocols used in the cancer patient as well as of the disease process itself. Pain most commonly involves the bones and joints, and is due to pressure caused by infiltration and accumulation of neoplastic cells in the bone marrow. *Hemarthrosis* (bleeding into joints) can cause acute episodes of joint pain. Chemotherapy may help reduce bone pain, as the number of neoplastic cells is reduced drastically. Patients are subjected to numerous painful procedures during diagnosis, treatment, and monitoring. Frequent collection of blood and bone marrow samples,

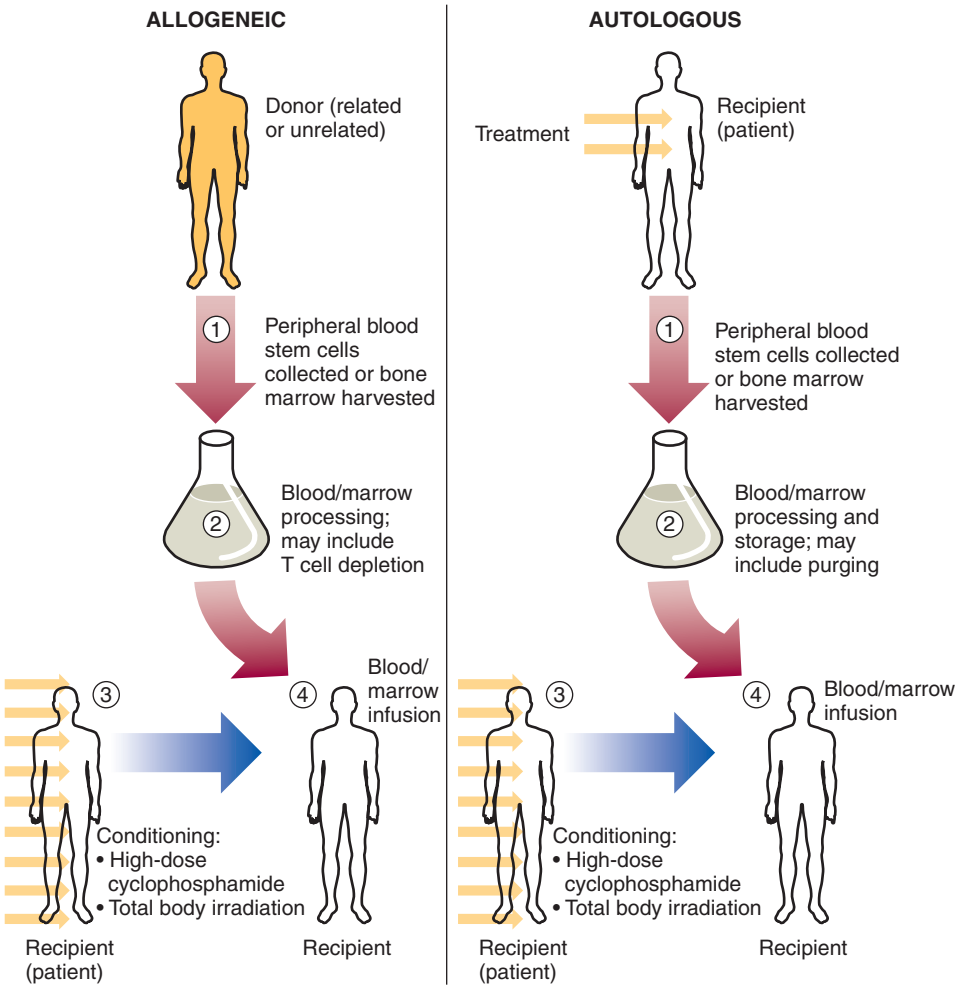


FIG 11.2 Procedures for allogeneic and autologous stem cell transplantation. For nonmyeloblative (reduced intensity) allogeneic stem cell transplantation, lower doses of chemotherapy, with or without radiotherapy, are used. (From Rodak B et al: *Hematology: clinical principles and application*, ed 4, Philadelphia, 2012, Saunders.)

placement of intravenous access lines for drug administration, and manifestations of unpleasant drug side effects all contribute to the pain experience. Nausea and mouth pain (stomatitis) are frequent complaints during chemotherapy. Pain management with a variety of strategies, including narcotic and nonnarcotic drugs, distraction, and biofeedback, is generally helpful. (See Chapter 47 for a discussion of pain and pain management.)

Epithelial cells, with normally high rates of turnover, are particularly susceptible to damage by radiation and chemotherapy. Sloughing of skin and mucous membranes and hair loss (alopecia) are common. Loss of skin and mucous membrane integrity increases the risk of infection and can contribute significantly to the pain and discomfort of treatment. Abnormalities in growth, development, and fertility are complications of particular concern in children undergoing radiation and chemotherapy.

KEY POINTS

- Classification of the types of leukemia is based on cell type involved (lymphoid or myeloid) and degree of cell maturation. Common myeloid neoplasms include Chronic myeloid leukemia (CML), polycythemia vera (PV), essential thrombocythemia (ET), and acute myeloid leukemia (AML). Common lymphoid neoplasms include chronic lymphoid leukemia (CLL), acute lymphoid leukemia (ALL), plasma cell myeloma, Hodgkin disease, and various forms of non-Hodgkin lymphoma.
- Risk factors for the development of hematologic neoplasms include exposure to chemical, viral, and radiation mutagens; consequences of chemotherapy drugs; and effects of immunodeficiency disorders.
- Common manifestations of hematologic neoplasia are due to insufficient production of normal white blood cells, red blood cells, and platelets, as evidenced by leukopenia, anemia, and thrombocytopenia. Anemia manifests as pallor, fatigue, dyspnea, and decreased activity tolerance. Thrombocytopenia causes petechiae, bleeding gums, hematuria, and prolonged bleeding time. Leukopenia manifests as frequent, recurrent infections. Other manifestations may occur with infiltration of tissues and organs. These include weight loss, anorexia, lymphadenopathy, bone pain, and central nervous system (CNS) dysfunction.
- Chemotherapy is the mainstay of management for most hematologic neoplasms. Several courses may be needed to kill neoplastic stem cells. Most chemotherapeutic agents interfere with some aspect of DNA replication and cell division to induce apoptosis (programmed cell death).
- Treatment is associated with a number of potential complications, including anemia, infection, and bleeding. Rapidly dividing hair cells and mucous membranes are also affected, leading to alopecia and stomatitis. Transfusion of blood products or stimulation of endogenous production with colony-stimulating factors and erythropoietin may be necessary. Bone marrow transplantation may be undertaken in some cases to restore stem cell function.

MYELOID NEOPLASMS

Myeloid neoplasms result from transformation and proliferation of a precursor stem cell in the bone marrow (Fig. 11.3). The progeny of the aberrant stem cell clone accumulate in the bone marrow and are released into the circulation. In many cases, the abnormal stem cell is multipotent and causes the overproduction of more than one cell type, resulting in myeloproliferative disease. The cells produced in myeloproliferative diseases are usually functional and have a normal morphologic appearance. The common myeloproliferative diseases are CML, PV, and ET, referring, respectively, to an excess of granulocytes, red blood cells, and platelets. These insidious and indolent disorders have few clinical symptoms and are commonly discovered on routine CBC analysis. Common features of CML, PV, and ET include involvement of a multipotent hematopoietic progenitor cell; hypercellularity of marrow; overproduction of one or more functional blood cells; chromosomal abnormalities involving chromosomes 1, 8, 9, 13, and 20; and eventual spontaneous conversion to AML or development of marrow fibrosis.

In contrast to the myeloproliferative diseases, the myelodysplastic syndromes and AML are characterized by neoplastic cells that are morphologically and functionally abnormal. The prognosis for myelodysplastic syndromes and AML is poor, and intensive treatment is necessary to extend life. AML and CML are described next. A discussion of PV and ET can be found in Chapters 13 and 14, respectively.

Chronic Myeloid Leukemia

Pathogenesis and clinical manifestations. CML represents approximately 15% of all cases of leukemia in the United States. The average age of onset is between 40 and 50 years, and CML occurs only occasionally in childhood and adolescence. The majority of CML cases are characterized by malignant granulocytes that carry a unique chromosomal abnormality, the Philadelphia chromosome (Ph^+). The Philadelphia chromosome is formed because of a balanced translocation between chromosomes 9 and 22 (Fig. 11.4). The translocation causes two genes to be juxtaposed, resulting in a new fusion gene called *bcr-abl.* This mutation is thought to be critical in the development of CML. Molecular studies have revealed that the protein product of the fusion gene is a functional enzyme that spurs cell proliferation and reduces apoptotic cell death. CML is unusual among human cancers because a single oncogene *(bcr/abl)* is capable of conferring a malignant state. Numerous mutations are necessary for development of most other cancers (see Chapter 7). The cells in CML are more mature than those found in AML, as noted by the greater degree of nuclear segmentation (Fig. 11.5).

The usual clinical presentation of CML includes a high granulocyte count on the CBC and splenomegaly. Symptoms, when present, may include fatigue, weight loss, diaphoresis, bleeding, and abdominal discomfort from the enlarged spleen.

Prognosis and treatment. CML does not respond well to chemotherapy. Although most patients will achieve a temporary remission, the overall survival time is poor. In untreated patients the median survival

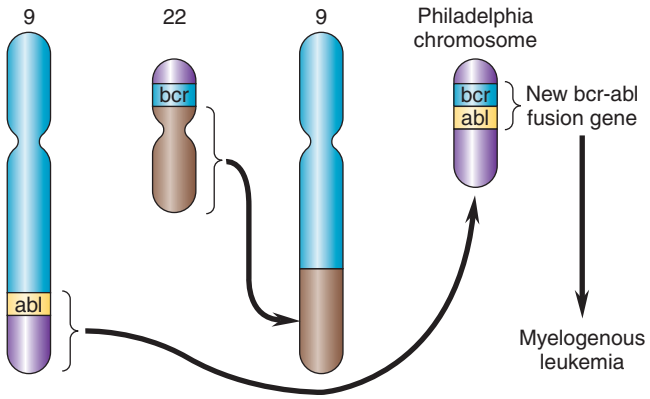


FIG 11.4 Balanced translocation between chromosomes 9 and 22 results in the formation of a Philadelphia chromosome. The translocation causes two genes, *abl* and *bcr,* to become juxtaposed, resulting in a fusion gene. This fusion gene, *bcr-abl,* is thought to be essential for the development of chronic myeloid leukemia.

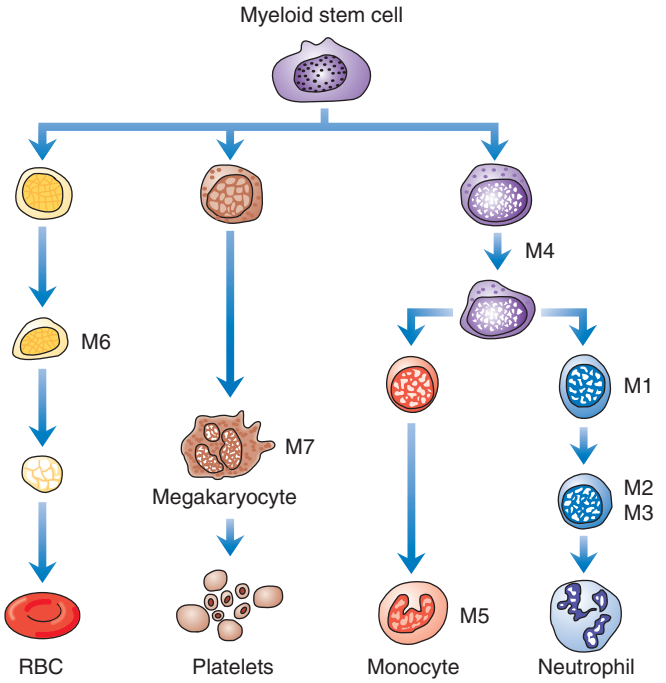


FIG 11.3 Maturation pathways of myeloid cells showing the different types of leukemia associated with various stages of development. M1, M2, and M3 types result in granulocytic leukemia; M4 has characteristics of monocytic and granulocytic leukemia; M5 is monocytic; M6 is associated with erythroid leukemia; and M7 is associated with megakaryocytic leukemia. *RBC,* Red blood cell.

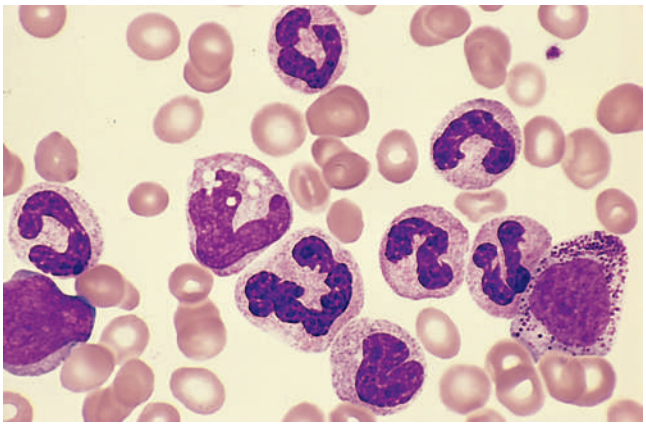


FIG 11.5 Peripheral blood smear from a patient with chronic myeloid leukemia (CML). Note that a greater degree of neutrophil segmentation is found in CML than in acute myelogenous leukemia (see Fig. 11.6), reflecting a more advanced stage of development. (From Hoffman R, et. al. *Hematology: basic principles and practice,* ed 6, Philadelphia, 2013, Saunders.)

is about 2 years. After discovery of the fusion protein *(bcr/abl),* drugs targeted to inhibit its action were developed (e.g., imatinib). The goal of anti-*bcr/abl* therapy is to reduce the number of leukemic cells with the *bcr/abl* phenotype to undetectable levels. It is not known whether imatinib can cure CML or what duration of treatment is necessary to permanently suppress the leukemic cell population. Some patients with CML have developed drug resistance against imatinib, and research is ongoing. The only known curative treatment is allogeneic bone marrow transplantation from a suitable donor. It is believed that the donor cells detect and kill the host's leukemic cells. Even with a human leukocyte antigen (HLA)–identical sibling donor, the probability of transplant-related mortality is about 25%, and the likelihood of long-term disease-free survival is 50% to 60%. Transplant-related mortality is about 50% if the HLA-matched donor is unrelated. Bone marrow transplantation with cells harvested from the patient's own blood (autografting) during the early stages of CML may also be done, but it is less effective in curing the disease. For those not able to undergo stem cell transplantation, standard chemotherapy during the chronic phase may be instituted. Once CML has progressed to the blast phase, essentially becoming AML, the prognosis is poor, regardless of treatment, with an expected median survival of 3 to 4 months.

Acute Myeloid Leukemia

Pathogenesis and clinical manifestations. AML is primarily a disease of adults, comprising 80% of cases of acute leukemia in this population but accounting for only 20% of the cases of acute leukemia of childhood. The median age at presentation is 64 years. Like CML, AML is a malignant disorder associated with transformation of a myeloid stem cell. The bone marrow aspirate must have more than 20% blasts to be classified as AML. AML can present in a variety of ways because of the potential for myeloid stem cells to produce different cell types. Thus AML has a number of subtypes. Acute granulocytic leukemia is the most common type of disorder, and the term is often used interchangeably with AML. Myeloblastic cells have a large, nonsegmented nucleus and fine chromatin (Fig. 11.6). AML is also subtyped according to genetic abnormalities. Worse outcomes are noted with loss of *P53* or *RB* tumor suppressor gene function. Most are chromosomal translocations or inversions. If cytogenetic class is not apparent, then AML is classified by morphologic characteristics, including myeloid cell of origin and degree of differentiation or maturation. Correct classification increases the accuracy of prognosis and may influence the choice of treatment.

AML presents in a manner very similar to that of ALL, and the two are difficult to distinguish by clinical findings alone. Acute leukemia causes bone pain, anemia, thrombocytopenia, and increased susceptibility to infection. The skin, the genitourinary and gastrointestinal systems, and the respiratory tract are common infection sites. The onset of symptoms is abrupt, with most patients seeking care within a few weeks of disease onset. The prognosis is much worse for AML than for ALL, with fewer than 50% of children and only about 25% of adults achieving long-term survival. An exception is the promyelocytic subtype of AML. Although it accounts for only 10% to 15% of AML cases, acute promyelocytic leukemia (APL) deserves special consideration because it is the most curable of all AML subtypes, with a 70% to 80% 5-year disease-free survival. APL is characterized by a chromosomal translocation between chromosomes 8 and 21, which forms a fusion gene called *PML/RARα.* The PML/RARα protein binds to a repressor complex in the cell nucleus and inhibits myeloid cell differentiation. RARα is a retinoic acid receptor that can be induced to release its inhibitory hold on differentiation when *all-trans*-retinoic acid (ATRA) is administered. Addition of ATRA to the chemotherapy management of patients with APL significantly improves disease-free survival (see Table 11.1 for a comparison of acute and chronic leukemias).

Prognosis and treatment. Treatment protocols for AML are increasingly incorporating the cytogenetic profile of the leukemic cells to individualize therapy and monitor response. Traditionally the management of AML has two phases: remission induction and consolidation/postremission. A CR is attempted in the remission induction phase, with an attempt to eliminate any undetected residual leukemic cells during the consolidation/postremission phase. Patients with AML who are able to complete only one or two cycles of their chemotherapy because of toxicity almost invariably have recurrence of leukemia even when CR was achieved. To induce remission, most protocols use two cycles of a combination of agents. Postremission therapy commonly includes high-dose Ara-C in younger patients, whereas the elderly require lower-dose regimens. Drug treatment is constantly being evaluated and altered to obtain better outcomes. With the advent of allogeneic stem cell transplantation, the chances for a cure may improve; however, procedure-related mortality is 10% to 25%. New therapies using monoclonal antibodies to detect and destroy leukemic cells have been used and show promise for improving outcomes in AML.

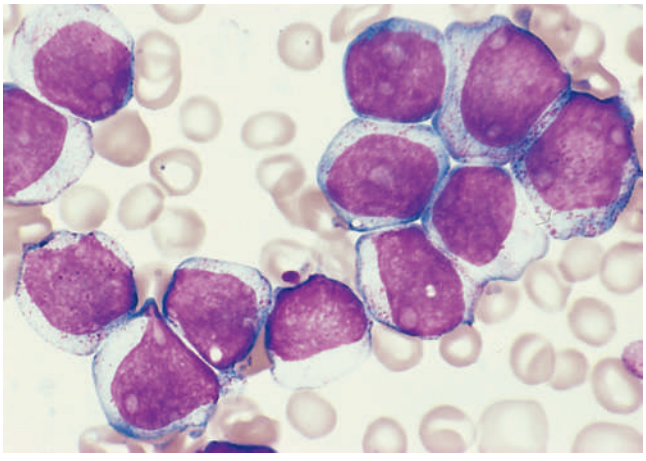


FIG 11.6 Peripheral blood smear showing typical cells of acute myelogenous leukemia. (From Eric Jeandidier et. al, A cytogenetic study of 397 consecutive acute myeloid leukemia (AML) cases identified three with a t(7;21) associated with 5q abnormalities and exhibiting similar clinical and biological features, suggesting a new, rare AML entity. *Cancer Genetics,* 2012;205(7):365–372.)

TABLE 11.1 Comparison of Acute and Chronic Leukemias

	Acute	Chronic
Age	All ages	Adult
Clinical onset	Sudden	Insidious
Course of untreated disease	Weeks to months	Months to years
Predominant cell	Blasts, some mature forms	Mature forms
Anemia	Mild to severe	Mild
Thrombocytopenia	Mild to severe	Mild
WBC count	Variable	Increased

Adopted from McKenzie S: *Clinical laboratory hematology,* Upper Saddle River, NJ, 2004, Pearson, p 482.

KEY POINTS

- Chronic myeloid leukemia (CML) is a myeloproliferative disorder that primarily affects adults, has an insidious onset, and responds poorly to chemotherapy. Most cases of CML are characterized by the presence of a gene translocation (Philadelphia chromosome) that produces a fusion gene called *bcr-abl*. The product of this gene is thought to be responsible for producing the malignancy. This molecular abnormality is the target of new drug therapies to eliminate neoplastic cells. The prognosis of CML remains relatively poor, but may be improved with allogeneic bone marrow transplantation in which transplanted cells destroy leukemic cells.
- AML affects adults primarily, has an acute onset, responds fairly well to treatment, and has a prognosis somewhat worse than that of ALL. AML is usually a malignancy of granulocytes, although other myeloid cell types may be affected. Several gene abnormalities have been identified in AML that may indicate better or worse prognosis. The median survival of patients younger than 60 years is 30% to 40% at 4 years.

LYMPHOID NEOPLASMS

The lymphoid neoplasms include malignant transformations of B cells, T cells, and NK cells. When present in blood and bone marrow, lymphoid neoplasms are called *leukemias*, and when they are localized in lymphoid tissues, they are called *lymphomas*. The location of lymphoid neoplasms is a consequence of the stage of the disease. The WHO classification uses cell type rather than stage to classify the lymphoid neoplasms, resulting in some difficulty with the traditional conceptualization of leukemias and lymphomas. The factors that determine whether a particular neoplastic cell will present as leukemia or as lymphoma are not presently known. Subcategories of the B-cell and T-cell/NK-groups are based on the maturity of the neoplastic cells (see Box 11.2). The *precursor cell* neoplasms are characterized by cells that have arrested development in the early blast stage, whereas the *mature cell* neoplasms are more differentiated and often located in peripheral sites (Fig. 11.7).

Chronic Lymphoid Leukemia

Pathogenesis and clinical manifestations. CLL accounts for about 30% of all cases of leukemia in the United States. In 95% of cases, a malignant B-cell precursor is at fault. Only 5% of cases of CLL are associated with T-cell transformation, but this type is more aggressive. In general, B-cell CLL follows an indolent course, which is usually asymptomatic. Often CLL is found by accident on routine blood count examinations. When CLL becomes symptomatic, patients may experience fatigue, weight loss, and anorexia. Because the leukemic B cells do not produce antibodies normally, an increased susceptibility to certain types of infection may occur. Malignant lymphocytes invade lymphoid tissues and bone marrow, disrupting function. Lymphoid invasion often presents as enlarged, painless lymph nodes (lymphadenopathy) or enlarged spleen (splenomegaly). Bone marrow infiltration reduces the production of other cells, including red blood cells and platelets. A typical slide of a bone marrow aspirate from a patient with CLL is shown in Fig. 11.8. Note the preponderance of lymphoid cells. CLL cells are characterized by defective apoptosis and therefore have extended life spans. They are derived from mature peripheral B cells (see Table 11.2 for a comparison of acute and chronic leukemias).

Prognosis and treatment. Certain genetic mutations confer better or worse prognosis. A mutation in the variable region of the immunoglobulin gene (IgV) is associated with a median survival of 24 years or more; those without this mutation have a median survival of less than 8 years. CLL cell types demonstrating short telomere lengths and P53 dysfunction have poor outcomes. Because the average age of patients with CLL is about 65 to 70 years, those with indolent disease may not be treated; they are more likely to die of another disorder rather than CLL. Patients with cell types likely to progress rapidly may receive chemotherapy to induce remission. Those without complete response may consider stem cell transplantation to prolong the duration of remission.

Acute Lymphoblastic Leukemia/Lymphoma

Pathogenesis and clinical manifestations. ALL is a malignant disorder of the lymphoid cell lineage. The great majority of cases are

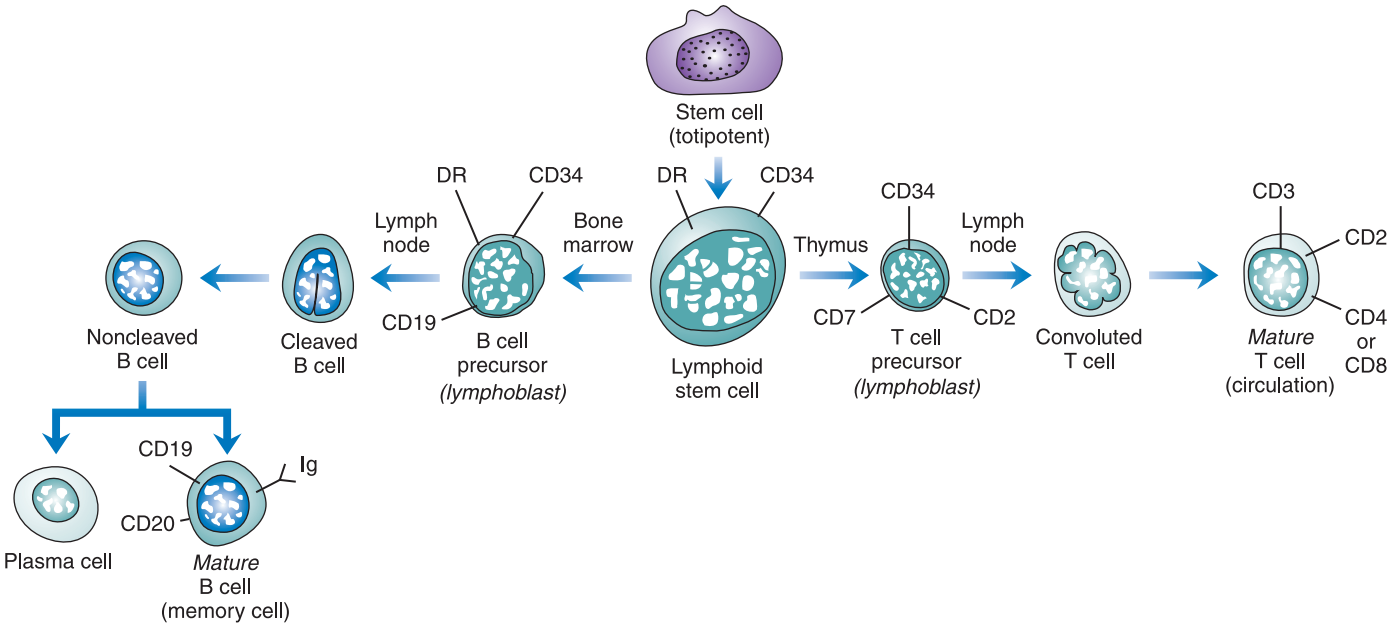


FIG 11.7 Maturation pathways of T and B lymphocytes showing the stages at which lymphocyte development is typically arrested in leukemia. Different markers are present on the surface of B cells and T cells at progressive stages of development, which are helpful in identifying the neoplastic cell type and maturity.

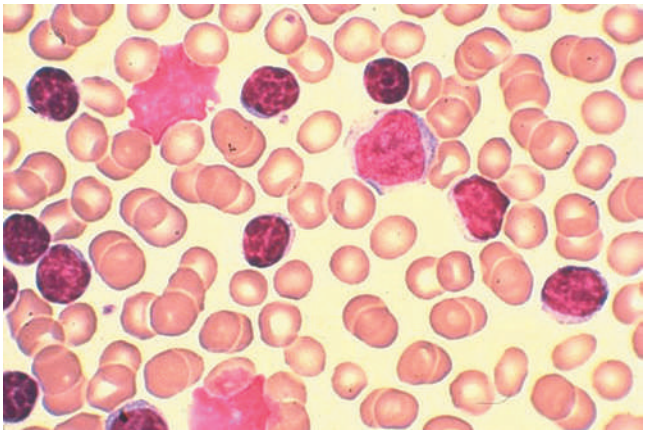


FIG 11.8 Bone marrow aspirate showing small lymphocytes with condensed nuclear chromatin typical of chronic lymphocytic leukemia, B-cell type. (From Henderson ES et al, editors: *Leukemia*, ed 7, Philadelphia, 2003, Saunders, color plate 11-28.)

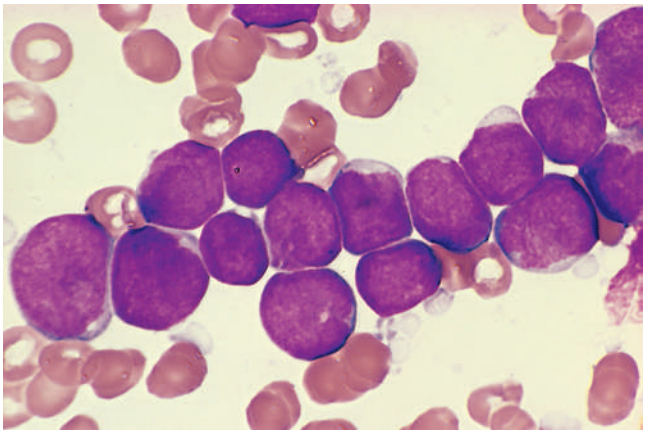


FIG 11.9 Peripheral blood smear showing typical cells of acute lymphocytic leukemia. (From Onciu M: Acute lymphoblastic leukemia. *Hematol Oncol Clin North Am* 2009;23(4):655–674.)

TABLE 11.2 Comparison of Acute Lymphocytic (ALL) and Acute Nonlymphocytic Leukemia (ANLL)

	ALL	ANLL
Age	Common in children	Common in adults
Hematology results	Anemia, neutropenia, thrombocytopenia	Anemia, neutropenia, thrombocytopenia
Cell morphology	Small to medium lymphoblasts, fine chromatin, indistinct nucleoli	Medium to large myeloblasts, distinct nucleoli, clear chromatin, Auer rods
Cytochemistry	PAS positive, peroxidase negative, Sudan black B negative	PAS negative, peroxidase positive, Sudan black B positive

Adopted from McKenzie S: *Clinical laboratory hematology,* Upper Saddle River, NJ, 2004, Pearson, p 487.
PAS, Periodic acid–Schiff reaction.

the result of malignant transformation of B cells (80%), with the remainder involving T cells. The abnormal cells resemble immature lymphocytes, called *lymphoblasts* (Fig. 11.9). Most lymphoblastic neoplasms present as leukemias, but lymphoblastic lymphomas are thought to be the same disease at a different stage. B-cell leukemias are categorized into cytogenetic groups based on common chromosomal translocations. One of these transformations results in the *bcr/abl* fusion gene discussed previously in the context of CML. Three other types of translocations also form fusion genes that produce abnormal signaling components. These gene derangements have different prognoses and may respond differently to alternative treatment protocols.

Lymphoblasts do not mature and accumulate in large numbers in the blood and bone marrow. At least 20% of the bone marrow cells must be leukemic lymphoblasts to meet the diagnostic criteria for ALL. The space occupied by the accumulation of leukemic cells in the bone marrow prohibits the production of normal red blood cells, platelets, and leukocytes. Circulating blasts are poorly functioning cells and do not provide effective immunocompetence (see Table 11.2 for a comparison of ALL and acute nonlymphocytic leukemia [ANLL]).

ALL primarily affects children. It is the most common malignancy and the second-leading cause of death in this population. The peak incidence occurs between the ages of 3 and 7 years. A second peak occurs in middle age. The onset of symptoms is abrupt, with common complaints of bone pain, bruising, fever, and infection. Children may refuse to walk, and their parents may report loss of appetite, fatigue, and abdominal pain. The spleen, liver, and lymph nodes may be enlarged from leukemic infiltration. A small number of children (3%) may present with CNS signs from leukemic infiltration of brain tissues.

Prognosis and treatment. ALL is highly curable in the pediatric population, but less so in adults. The 5-year survival rate is 85% in children and 30% to 50% in adults. Certain forms of ALL are more responsive to therapy. For example, children with the pre–B-cell type have a 90% cure rate, whereas those with mature B-cell or immature T-cell leukemia have a poorer prognosis.

Chemotherapy is used for remission induction. Postremission chemotherapy with or without stem cell transplantation is indicated for most patients. In general, adults with ALL require more intense therapy than children to achieve complete remission (CR). Monoclonal antibodies may be used in patients whose tumors express specific antigens.

Hairy Cell Leukemia

Pathogenesis and clinical manifestations. Hairy cell leukemia is a rare, chronic type of leukemia. The disease represents about 2% of adult leukemias, but it is of interest because of its highly treatable nature. The median age at presentation is about 55 years, and there is a 5-to-1 predominance of males. Hairy cell leukemia has a B-cell phenotype and is characterized by the presence of peculiar cells with hairlike projections on their surface (Fig. 11.10). At diagnosis, patients have hairy cells in the peripheral blood as well as reduced numbers of granulocytes, platelets, and red blood cells. Splenomegaly is a common finding, being present in 90% of patients.

Prognosis and treatment. Treatment may be instituted when a patient becomes symptomatic with an enlarged spleen, recurrent infection, bleeding disorder, or anemia. An appropriate chemotherapeutic protocol produces CR rates of 80%.

Plasma Cell Myeloma (Multiple Myeloma)

Pathogenesis and clinical manifestations. Plasma cell myeloma, also known as *multiple myeloma,* is a malignant disorder of mature, antibody-secreting B lymphocytes, called *plasma cells.* Malignant plasma cells have a predilection to invade bone and form multiple tumor sites.

Other tissues may be targeted also, including lymph nodes, liver, spleen, and kidneys. Plasma cell myeloma occurs exclusively in the adult population, usually affecting individuals older than 40 years, with a median age at presentation of 65 years. Men are affected more often than women.

As with other forms of neoplasia, the exact etiologic process of plasma cell myeloma is unknown, but abnormalities in chromosome structure and number are commonly found. The malignant plasma cells all belong to a single clone, and the excessive antibodies they produce are identical monoclonal antibodies. These accumulate in the bloodstream and can be detected by serum protein electrophoresis. Normally, serum antibodies are of many forms (polyclonal) and show a varied distribution of size on the electrophoresis test. In plasma cell myeloma, there is a large amount of one type of antibody, which forms a characteristic spike (Fig. 11.11). Excessive production of light-chain antibody fragments by malignant plasma cells results in their accumulation in blood and urine. When found in urine, these light-chain fragments are called Bence Jones protein**.** In addition to helping confirm the diagnosis, Bence Jones protein is important to the pathogenesis of plasma cell myeloma because it can accumulate in the kidneys and cause kidney damage. Malignant plasma cells tend to accumulate in bone, where they enhance osteoclastic activity and produce bone lesions. Pathologic fractures, especially compression fractures of the vertebral column, are common. Bone destruction releases calcium into the bloodstream, with resultant hypercalcemia.

Most of the clinical manifestations of multiple myeloma are due to bone and renal damage. The diagnosis of plasma cell myeloma is suspected based on the monoclonal antibody peak, the presence of Bence Jones protein, hypercalcemia, and evidence of bone lesions. The diagnosis is confirmed by bone marrow biopsy. Normally, the plasma cell component of the marrow comprises about 5%. In multiple myeloma, plasma cells may occupy 30% to 95% of the bone marrow (Fig. 11.12). A minimum of at least 10% to 15% of bone marrow plasma cells is necessary for the diagnosis of plasma cell myeloma. The likelihood of bone marrow dysfunction increases as the plasma cell component increases. Normal production of erythrocytes, platelets, and leukocytes can be impaired to varying degrees.

The onset of plasma cell myeloma is generally slow and insidious. A premalignant stage is apparent in some individuals who have excess production of monoclonal antibodies but no evidence of bone lesions or Bence Jones protein in the urine. This stage is called *monoclonal gammopathy of undetermined significance (MGUS).* Approximately 25% of patients with MGUS progress to malignant disease. Affected individuals remain asymptomatic until the disease is fairly advanced. The asymptomatic stage often lasts for many years after malignant transformation. During this time the only complaint may be frequent infections. Diagnosis during the asymptomatic phase is usually made because protein in the urine or high serum calcium levels are found on routine examination. Bone pain is usually the first symptom. Sometimes the evaluation of a fracture or back pain leads to the identification of myeloma. Anemia, recurrent infections, and bleeding tendencies are suggestive of bone marrow depression.

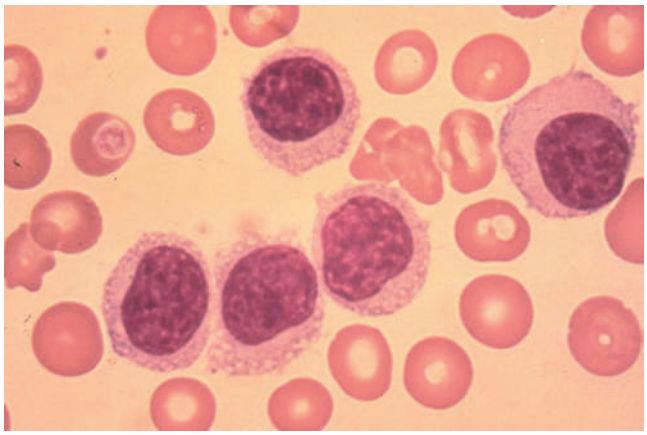


FIG 11.10 Peripheral blood smear showing cells typical of hairy cell leukemia. (From Skarin AT: *Atlas of diagnostic oncology,* ed 4, Philadelphia, 2010, Saunders.)

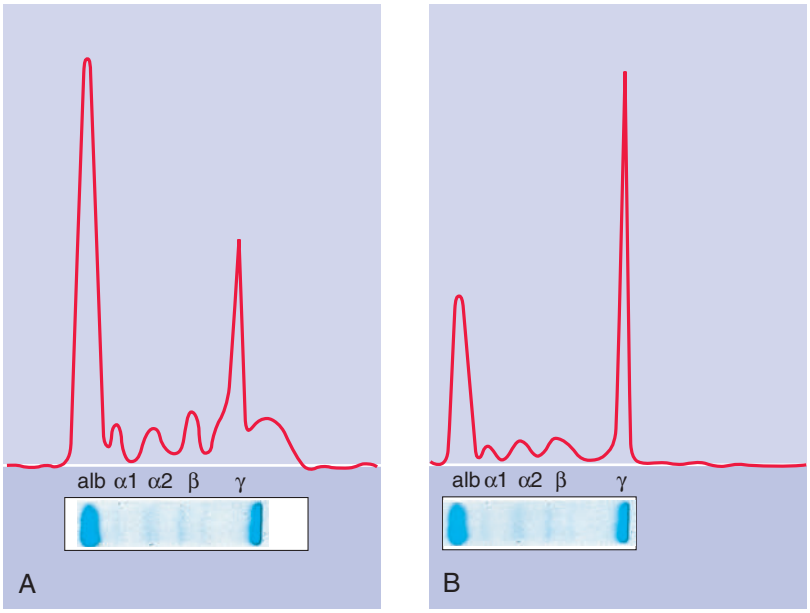


FIG 11.11 Serum protein electrophoresis comparing abnormal myeloma protein in the γ region typical of benign monoclonal gammopathy **(A)** with the large quantity of monoclonal antibody *(spike)* γ typical of plasma cell myeloma **(B).** (From Skarin AT: *Atlas of diagnostic oncology,* London, 2003, Gower Medical, pp 536–537.)

Renal insufficiency is a complication experienced by approximately 50% of patients with plasma cell myeloma. Impairment of renal function is due to a combination of factors, including hyperproteinemia, high levels of Bence Jones protein, hypercalcemia, and hyperuricemia. Renal function may continue to decline over time, culminating in end-stage renal disease (chronic renal failure).

Bone involvement is a consistent feature of plasma cell myeloma. Radiologic studies of ribs, spine, skull, and pelvis show a characteristic "honeycomb" appearance, attributable to lucid areas of demineralized bone (Fig. 11.13). Minimal trauma is likely to result in fractures. Sometimes fractures occur with no known trauma; these are called *pathologic* fractures.

Prognosis and treatment. Antineoplastic agents may be used to induce and maintain a remission in plasma cell proliferation. The best chemotherapy regimen has not yet been determined. The remission induction rate is about 60%, with a median survival of about 3 years after initiation of therapy. High-dose chemotherapy followed by allogeneic bone marrow transplantation is becoming more common and offers a better CR rate. However, the death rate associated with transplantation is high (approximately 40% to 50%). Autologous stem cell transplantation is considered to be the optimal initial therapy for most patients. Pharmacologic management of renal dysfunction is often necessary.

Chronic bone pain is a common problem in the myeloma patient that may require use of multiple remedies. Narcotic and nonnarcotic pain relievers are often necessary. Localized application of radiation to bone lesions may reduce bone pain in some cases.

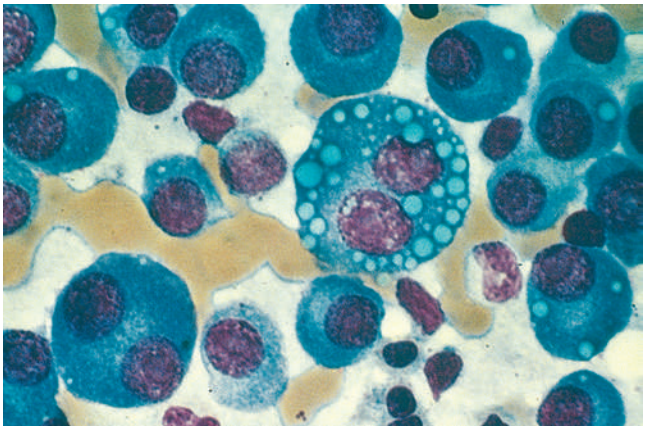


FIG 11.12 Bone marrow aspirate from a patient with multiple myeloma showing a large number of abnormal plasma cells with multiple nuclei and cytoplasmic droplets. (From Kocjan G: *Diagnostic cytopathology essentials*. Churchill Livingstone, 2013, Elsevier.)

KEY POINTS

- Chronic lymphoid leukemia (CLL) is a neoplastic transformation of a mature, peripheral B cell that affects adults primarily and has an insidious onset. CLL is usually asymptomatic. Disease in certain genotypes is associated with long survival times and does not require therapy; in other cases, disease is progressive and may be managed with stem cell transplantation or administration of monoclonal antibodies.
- Acute lymphoid leukemia (ALL) affects children primarily, has an acute onset, responds well to therapy, and has a good prognosis. ALL is associated with transformation of precursor "blasts" in the bone marrow. ALL often manifests with bone pain, infections, and a tendency for bleeding. A significant number of children with ALL have CNS involvement, and intrathecal chemotherapy is necessary.
- Plasma cell myeloma is due to malignant transformation of antibody-secreting B lymphocytes. It primarily affects older adults. The onset of symptoms is insidious, with most patients experiencing a 4- to 10-year period of clinical latency. Some patients have a preneoplastic phase called MGUS. When present, symptoms include bone pain, pathologic fractures, anemia, thrombocytopenia, leukopenia, and renal insufficiency. Malignant plasma cells all secrete the same monoclonal antibody, and detection of this antibody in the blood or urine (Bence Jones protein) aids in diagnosis.

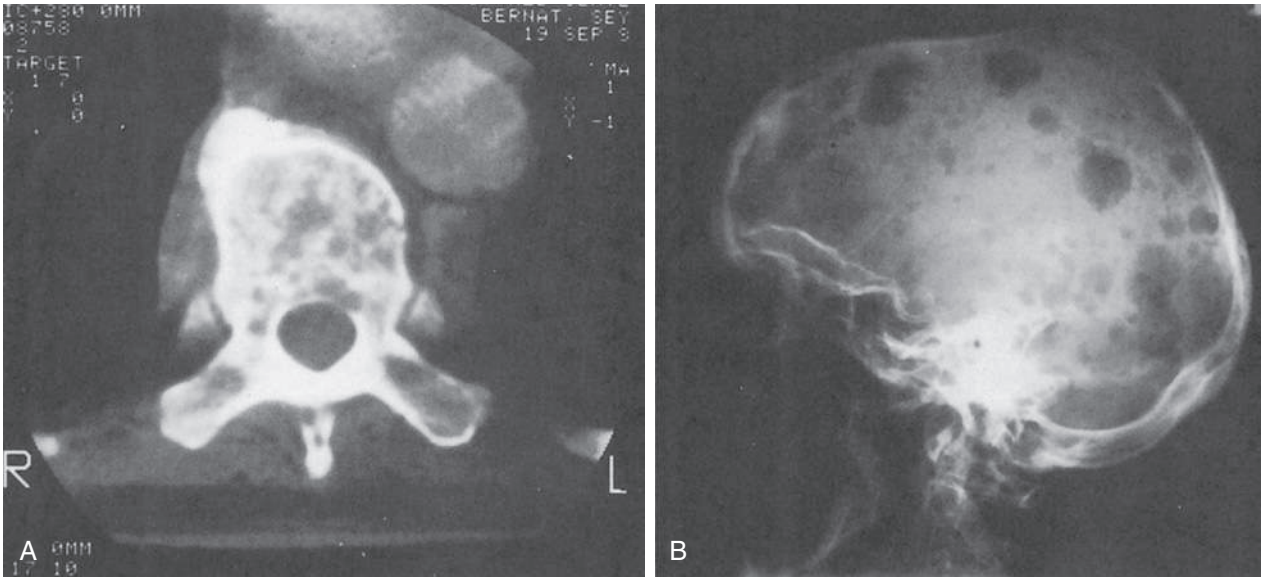


FIG 11.13 Vertebral body **(A)** and skull **(B)** radiographs showing the characteristic "honeycomb" appearance of demineralized bone associated with multiple myeloma. (Courtesy Marvin J. Stone, MD, Sammons Cancer Center, Baylor University Medical Center, Dallas, TX.)

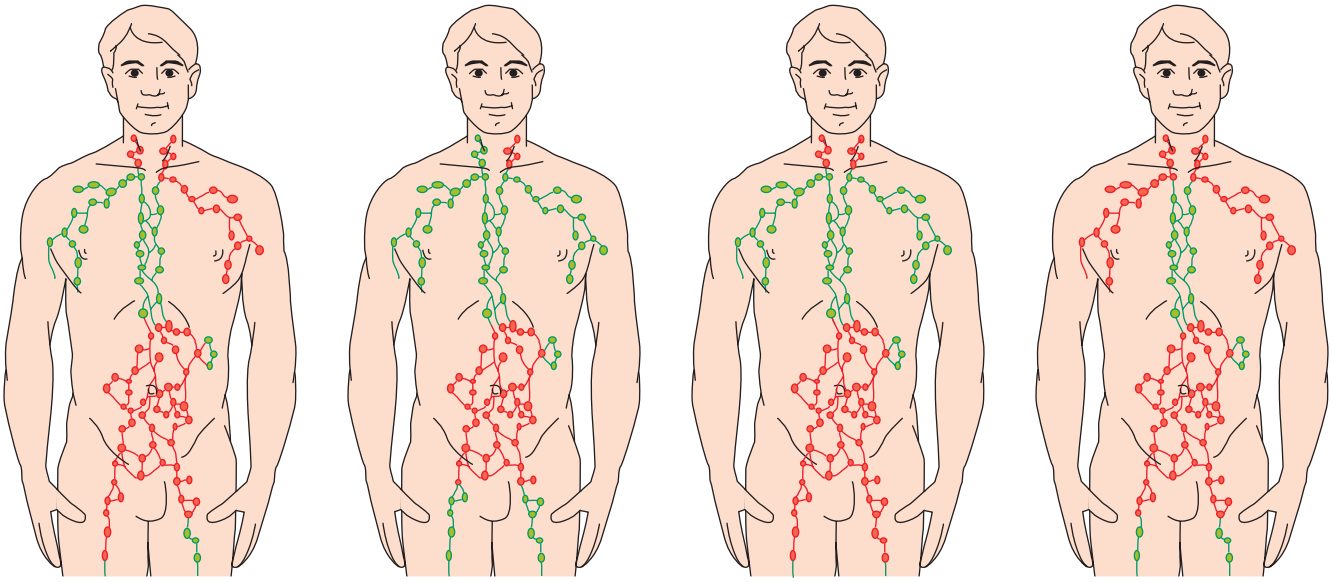


FIG 11.14 Schematic drawing showing the orderly, contiguous, and predictable spread of Hodgkin disease. (Redrawn and modified from Rosenberg SA: Hodgkin disease: no stage beyond cure. *Hosp Pract* 1986;21(8):97. After original illustrations by Bunji Tagawa.)

Hodgkin Disease

Hodgkin disease represents about 30% of all cases of malignant lymphoma, accounting for approximately 8500 new cases annually in the United States. It occurs across the age continuum, with half of cases occurring in persons between the ages of 20 and 40 years. The overall incidence of Hodgkin disease is higher in males, who have a worse prognosis. The overall 5-year survival rate for treated Hodgkin disease, including all stages, is about 85%.

Pathogenesis and clinical manifestations. Hodgkin disease is a malignant disorder of the lymph nodes characterized by the presence of Reed-Sternberg cells on histologic examination. Reed-Sternberg cells originate from B cells in the germinal centers of lymph nodes. Reed-Sternberg cells are malignant, but they tend to grow and spread in a predictable manner. This predictability differentiates Hodgkin disease from other types of lymphoma. Hodgkin disease usually metastasizes along contiguous lymphatic pathways (Fig. 11.14). EBV is frequently found in the genome of transformed Reed-Sternberg cells and is thought to be important in the pathogenesis of Hodgkin disease. The malignant cells are clonal, originating from a single mutant precursor cell, and usually present in a single node or localized chain of nodes. In addition to malignant Reed-Sternberg cells, inflammatory cells accumulate within the node (Fig. 11.15) such that Reed-Sternberg cells constitute only a small minority (2%) of the cells in the lymph node tumor.

There are two types of Hodgkin disease: (1) the rare lymphocyte predominance type, which accounts for 5% of cases, and (2) the classical type (cHD) representing the other 95%. The classical type can be divided further into four subtypes according to the relative number of reactive cells in the tumor. The histologic pattern does not seem to predict the prognosis. The stage of Hodgkin disease is more relevant.

Clinical manifestations of Hodgkin disease are dependent on the site of origin, as well as on the stage of dissemination. Lymphomas often are asymptomatic in the early stages. The usual clinical presentation includes painless lymphadenopathy that may be accompanied by fever, night sweats, pruritus, weight loss, and malaise. Usually enlargement occurs in lymph nodes above the diaphragm, with the cervical nodes being the most common site (Fig. 11.16). Other supradiaphragmatic nodes are the supraclavicular, axillary, and mediastinal nodes. Less commonly, nodes below the diaphragm are the primary site. The inguinal nodes are the most common subdiaphragmatic site. As the disease spreads from the site of origin, other lymph nodes and lymphatic tissues may become involved, including the spleen and bone marrow. Staging procedures are performed to determine the extent of metastasis at the time of diagnosis. Staging dictates the treatment modality best suited to provide the patient with the greatest chance for long-term survival.

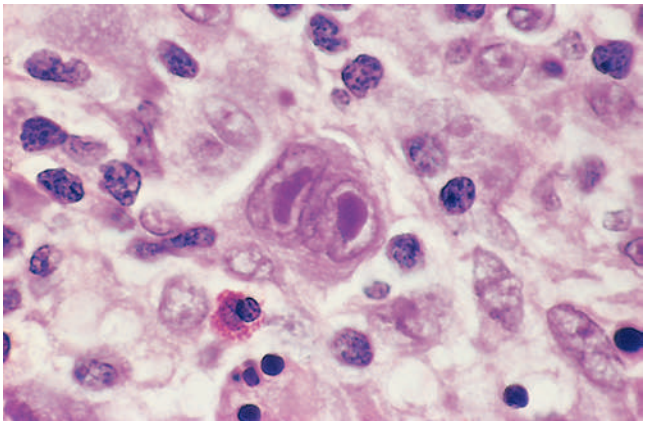


FIG 11.15 Hodgkin/Reed-Sternberg cells. (From Vandenberghe P: Noninvasive detection of genomic imbalances in Hodgkin/Reed-Sternberg cells in early and advanced-stage Hodgkin's lymphoma by sequencing of circulating cell-free DNA: a technical proof-of-principle study. *Lancet Haematol* 2015;2(2):e55–e65.)

Prognosis and treatment. The staging protocol commonly used today was first adopted in 1971 at the Ann Arbor symposium and modified later in 1989 at the Cotswold meeting (Box 11.4). The same procedure is also used for staging non-Hodgkin lymphomas. This protocol uses the presence or absence of certain clinical symptoms as well as the locations of affected nodes to determine the clinical stage of disease. The four stages are shown in Fig. 11.17. The letter A denotes

FIG 11.16 Hodgkin disease (stage IIA). Marked enlargement of cervical lymph nodes is present in this patient. It is usually painless and may be confined to only one area or may affect two or more areas. (From Skarin AT: *Atlas of diagnostic oncology,* London, 2003, Gower Medical, p 482.)

BOX 11.4 Ann Arbor Staging System for Lymphomas With Cotswold Modifications*

Stage I
Involvement of a single lymph node region or lymphoid structure (e.g., spleen, thymus, Waldeyer ring)

Stage II
Involvement of two or more lymph node regions on the same side of the diaphragm. The number of anatomic regions is indicated by a subscript (e.g., II_3).

Stage III
Involvement of lymph node regions or structures on both sides of the diaphragm
- III_1: with or without splenic, hilar, celiac, or portal nodes
- III_2: with paraaortic, iliac, mesenteric nodes

Stage IV
Involvement of extranodal site(s) beyond that designated "E"
The site is indicated by a letter code followed by a plus sign (+)

Modifying Characteristics
For Stages I to III
E: Involvement of a single, extranodal site contiguous or proximal to known nodal site

For All Stages
A: No symptoms
B: Fever (temperature >38°C), drenching sweats, weight loss (>10% body weight over 6 months)
X: Bulky disease
- >One-third widening of mediastinum
- >10 cm maximal dimension of nodal mass

*Clinical stage (CS) is based on history, physical examination, laboratory studies, and CT scans. Pathologic stage (PS) is based on tissue sampling obtained through invasive procedures such as laparotomy and biopsy.

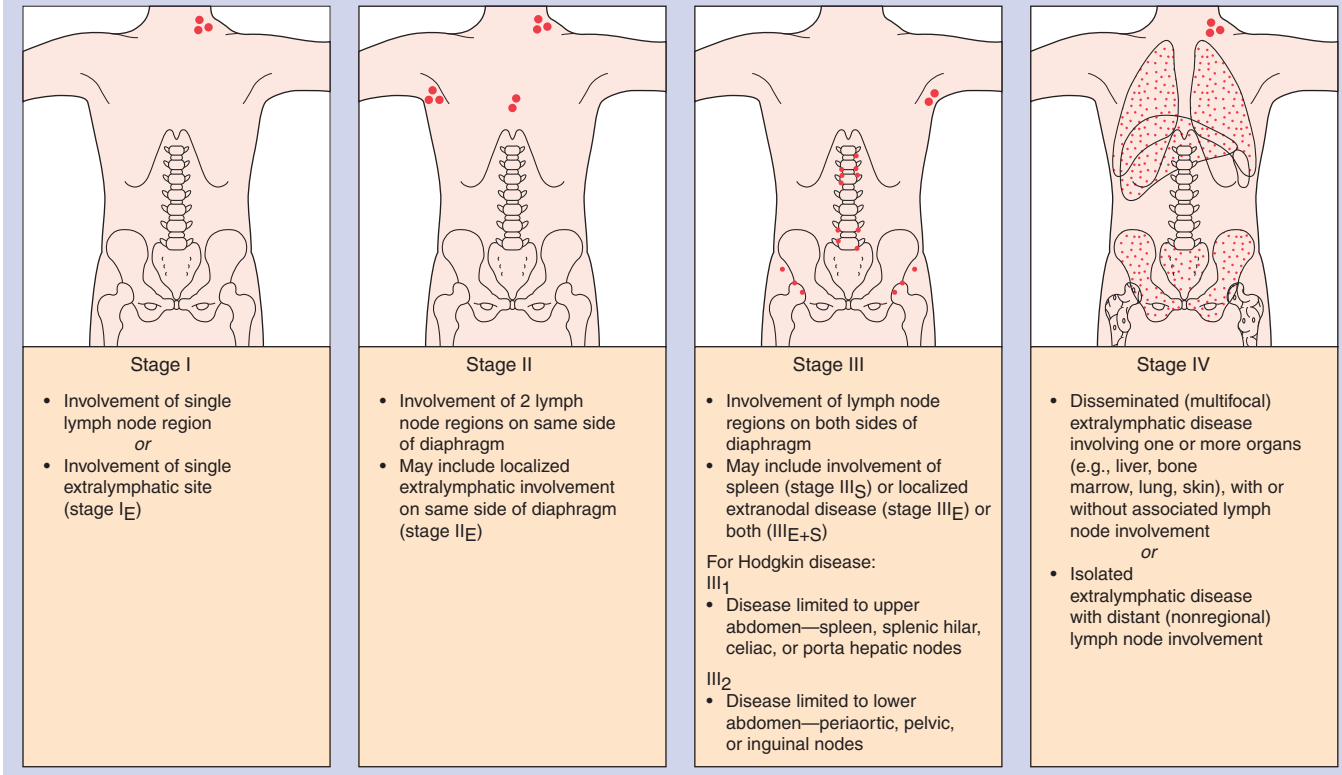


FIG 11.17 Depiction of the locations of malignant cells in the various stages of lymphoma using the Ann Arbor staging system. (From Skarin AT: *Atlas of diagnostic oncology,* London, 2003, Gower Medical, p 479.)

the absence of clinical symptoms, whereas the letter B is used when symptoms are present at the time of staging. These symptoms include loss of more than 10% of body weight, unexplained fevers, and night sweats. The clinical stage (CS) is based on history, physical examination, and noninvasive procedures such as computed tomography (CT) scanning. The pathologic stage (PS) is determined by the results of invasive procedures such as laparotomy and tissue biopsy. The stage dictates the treatment modalities used. In general, localized tumors are more amenable to application of radiation therapy, whereas disseminated disease responds better to systemic chemotherapeutic agents. Because Hodgkin disease often is detected while localized, radiation therapy is commonly used, with good results (Fig. 11.18).

Patients with nonbulky, stage IA or IIA disease may be candidates for radiation as sole therapy. However, a relatively high rate of relapse has been noted, and combined chemotherapy with limited-field radiation is often used. Patients with bulky disease, "B" symptoms, or stage III and IV disease require chemotherapy with or without radiation. In early-stage disease, appropriate treatment produces a 90% 10-year disease-free survival. More aggressive chemotherapy is indicated for patients with advanced Hodgkin disease.

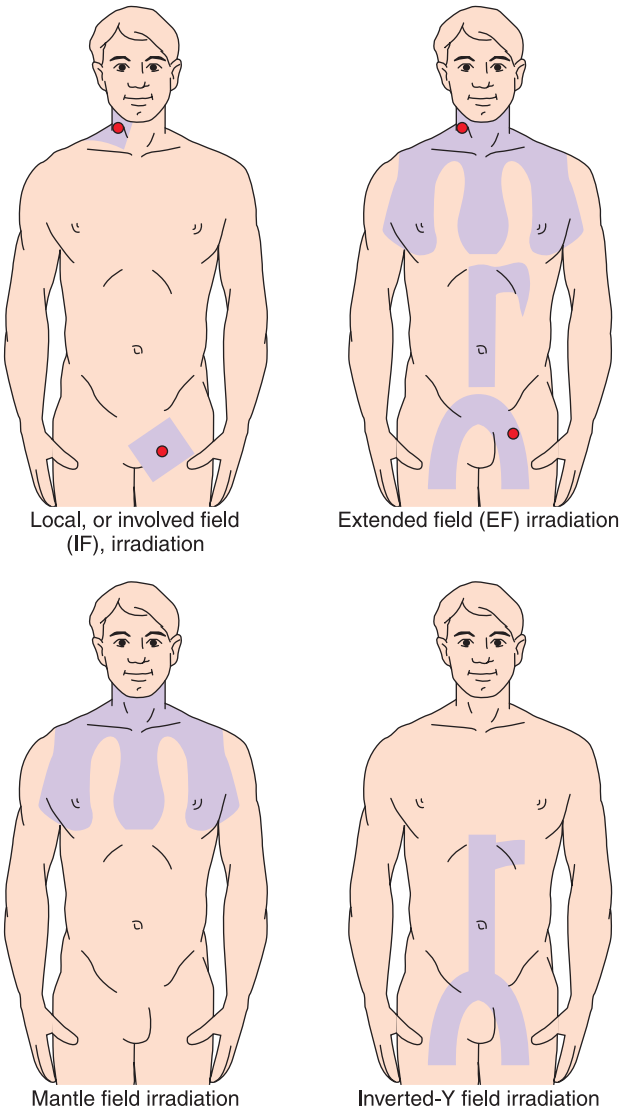


FIG 11.18 Typical radiation fields for lymphoma. Different fields of radiation may be used depending on the location of disease.

B-Cell, T-Cell, and NK-Cell Lymphoma (Non-Hodgkin)

The malignancies included in the classification of non-Hodgkin lymphoma are those that do not have the characteristic Reed-Sternberg cells found in Hodgkin disease. The majority of cases of non-Hodgkin lymphoma arise from lymph nodes, but they can originate in any lymphoid tissue. With the exception of a few subtypes, most cases of non-Hodgkin lymphoma occur in older adults (95%), and males are at a slightly higher risk than females. The incidence of non-Hodgkin lymphoma is on the rise, particularly in areas with large AIDS populations. More than 72,000 new cases of non-Hodgkin lymphoma are diagnosed annually in the United States. The lifetime risk of developing this disease is about 1 in 50.

Most cases of non-Hodgkin lymphoma arise from B cells, T cells, or NK cells. Some of the more common types of non-Hodgkin lymphoma are summarized in Table 11.3. The prognosis and recommended treatment protocols vary according to type. A general schema for grouping non-Hodgkin lymphoma according to indolent or aggressive types is in common usage. Generally, indolent disease is associated with longer survival times, whereas aggressive lymphomas tend to be disseminated at presentation and carry a generally poorer prognosis. As a group, the non-Hodgkin types of lymphoma are more likely to spread early and unpredictably in comparison with Hodgkin disease.

Pathogenesis and clinical manifestations. The etiologic process of non-Hodgkin lymphoma is thought to be similar to that of other malignant transformations. The tumor cells are all derived from a single mutant precursor cell and are clonal. Viruses are suspected in the development of some types of lymphoma. In particular, Burkitt lymphoma is strongly associated with the presence of EBV. Adult T-cell lymphomas are associated with infection by HTLV-1. The overall 5-year survival rate for all types of non-Hodgkin lymphomas combined is about 70%.

Most patients with non-Hodgkin lymphoma present with advanced disease (stage III or IV). Clinical manifestations may include painless lymphadenopathy, fever, night sweats, weight loss, malaise, and pruritus (similar to Hodgkin disease). A comparison of the features of Hodgkin disease and non-Hodgkin lymphoma is shown in Table 11.4. Extranodal involvement occurs early in the course of non-Hodgkin lymphoma, and patients may present with infiltrative disease of the skin, gastrointestinal tract, bone, or bone marrow. Complications occur more frequently than in Hodgkin disease. Two of the most serious oncology emergencies are obstruction of the superior vena cava and spinal cord compression. Infection, bone metastasis, and joint effusions are also common. Staging of non-Hodgkin lymphoma is done in the same way as for Hodgkin disease, and the classification system is not different. Earlier clinical stages are associated with the best prognosis for survival.

Prognosis and treatment. The effectiveness of therapy for non-Hodgkin lymphoma is variable. Favorable outcomes are likely in stage I and II disease. However, non-Hodgkin lymphoma is likely to present as stage III or IV disease, which has a poor prognosis. Therapeutic management is determined by the clinical stage, histologic type, patient age, and bone marrow integrity at the time of diagnosis. Radiation, chemotherapy, and tissue-specific therapies such as monoclonal antibodies and bone marrow transplants may all be applicable.

TABLE 11.3 Summary of Major Types of Lymphoid Leukemias and Non-Hodgkin Lymphomas

Diagnosis	Cell of Origin	Genotype	Salient Clinical Features
Neoplasms of Immature B and T Cells			
B-cell acute lymphoblastic leukemia/lymphoma*	Bone marrow precursor B cell	Diverse chromosomal translocations; t(12;21) involving *RUNX1* and *ETV* 6 present in 25%	Predominantly children; symptoms relating to marrow replacement and pancytopenia; aggressive
T-cell acute lymphoblastic leukemia/lymphoma	Precursor T cell (often of thymic origin)	Diverse chromosomal translocations, *NOTCH1*mutations (50%–70%)	Predominantly adolescent males; thymic masses and variable bone marrow involvement; aggressive
Neoplasms of Mature B Cells			
Burkitt lymphoma*	Germinal-center B cell	Translocations involving *MYC*and Ig loci, usually t(8;14); subset EBV-associated	Adolescents or young adults with extranodal masses; uncommonly presents as "leukemia"; aggressive
Diffuse large B-cell lymphoma†	Germinal-center or postgerminal center B cell	Diverse chromosomal rearrangements, most often of*BCL6* (30%), *BCL2* (10%), or *MYC* (5%)	All ages, but most common in older adults; often appears as a rapidly growing mass; 30% extranodal; aggressive
Extranodal marginal zone lymphoma	Memory B cell	t(11;18), t(1;14), and t(14;18) creating *MALT1-IAP2, BCL10-IgH,* and *MALT1-IgH* fusion genes, respectively	Arises at extranodal sites in adults with chronic inflammatory diseases; may remain localized; indolent
Follicular lymphoma†	Germinal-center B cell	t(14;18) creating *BCL2-IgH*fusion gene	Older adults with generalized lymphadenopathy and marrow involvement; indolent
Hairy cell leukemia	Memory B cell	Activating *BRAF* mutations	Older males with pancytopenia and splenomegaly; indolent
Mantle cell lymphoma	Naive B cell	t(11;14) creating *CyclinD1 -IgH* fusion gene	Older males with disseminated disease; moderately aggressive
Multiple myeloma/solitary plasmacytoma†	Postgerminal-center bone marrow homing plasma cell	Diverse rearrangements involving *IgH;* 13q deletions	Myeloma: older adults with lytic bone lesions, pathologic fractures, hypercalcemia, and renal failure; moderately aggressive Plasmacytoma: isolated plasma cell masses in bone or soft tissue; indolent
Small lymphocytic lymphoma/ chronic lymphocytic leukemia	Naive B cell or memory B cell	Trisomy 12, deletions of 11q, 13q, and 17p	Older adults with bone marrow, lymph node, spleen, and liver disease; autoimmune hemolysis and thrombocytopenia in a minority; indolent
Neoplasms of Mature T Cells or NK Cells			
Adult T-cell leukemia/ lymphoma	Helper T cell	HTLV-1 provirus present in tumor cells	Adults with cutaneous lesions, marrow involvement, and hypercalcemia; occurs mainly in Japan, West Africa, and the Caribbean; aggressive
Peripheral T-cell lymphoma, unspecified	Helper or cytotoxic T cell	No specific chromosomal abnormality	Mainly older adults; usually presents with lymphadenopathy; aggressive
Anaplastic large-cell lymphoma	Cytotoxic T cell	Rearrangements of *ALK*(anaplastic large cell lymphoma kinase) in a subset	Children and young adults, usually with lymph node and soft-tissue disease; aggressive
Extranodal NK/T-cell lymphoma	NK-cell (common) or cytotoxic T cell (rare)	EBV-associated; no specific chromosomal abnormality	Adults with destructive extranodal masses, most commonly sinonasal; aggressive
Mycosis fungoides/Sézary syndrome	Helper T cell	No specific chromosomal abnormality	Adult patients with cutaneous patches, plaques, nodules, or generalized erythema; indolent
Large granular lymphocytic leukemia	Two types: cytotoxic T cell and NK cell	Point mutations in *STAT3*	Adult patients with splenomegaly, neutropenia, and anemia, sometimes, accompanied by autoimmune disease

EBV, Epstein-Barr virus; *HIV,* human immunodeficiency virus; *Ig,* immunoglobulin; *NK,* natural killer.
*Most common tumors in children
†Most common tumors in adults.
Modified from Kumar V et al, editors: *Pathologic basis of disease,* ed 9, Philadelphia, 2015, Saunders, p 591.

TABLE 11.4 Clinical Differences in Hodgkin Disease and Non-Hodgkin Lymphoma

Characteristic	Hodgkin Disease	Non-Hodgkin Lymphoma
Pattern of spread	Contiguous spread	Noncontiguous spread
Extranodal disease	Uncommon	More common involvement of gastrointestinal tract, testes, bone marrow
Site of disease	Mediastinal involvement common	Mediastinal involvement less common
	Bone marrow involvement uncommon	Bone marrow involvement common
	Liver involvement uncommon	Liver involvement common
Extent of disease	Often localized	Rarely localized
B symptoms	Common	Uncommon

KEY POINTS

- Hodgkin disease is characterized by malignant transformation of B cells in lymph nodes, called *Reed-Sternberg cells*. The spread of malignant cells occurs along predictable, contiguous pathways. Most commonly, a single cervical lymph node is involved initially, with slow progression to nearby nodes.
- Non-Hodgkin lymphoma constitutes a diverse group of malignant diseases of lymphoid tissue. The characteristic Reed-Sternberg cell of Hodgkin disease is not present. Non-Hodgkin lymphoma is unpredictable in its spread and is often disseminated at diagnosis.
- Manifestations of Hodgkin disease and non-Hodgkin lymphoma are similar. These include painless lymph node enlargement, fever, night sweats, and weight loss. Early-stage disease is often asymptomatic.
- Staging is done to determine the degree of dissemination of disease. When affected lymph nodes are localized to one area (stage I) or one side of the diaphragm (stage II), the prognosis for cure is very good. Dissemination to lymph nodes above and below the diaphragm (stage III) or to extralymphatic organs or tissues (stage IV) carries a poorer prognosis.
- Radiation of the involved field is commonly used for malignant lymphoma in early stages. More disseminated disease may be treated with chemotherapeutic protocols. Non-Hodgkin lymphoma is routinely treated with chemotherapy because the disease is often well advanced at the time of diagnosis. Treatment may lead to bone marrow suppression and may predispose the patient to anemia, thrombocytopenia, and leukopenia.

SUMMARY

Malignant disorders of white blood cells are classified according to cell type and fall into two major categories: myeloid neoplasms and lymphoid neoplasms. Myeloid neoplasms commonly present as leukemia and usually involve transformation of granulocytes. Lymphoid neoplasms may present as leukemia, lymphoma, or plasma cell myeloma. Leukemia is a malignant neoplasm of immature stem cells that is characterized by diffuse replacement of the bone marrow by neoplastic blasts. In most cases the leukemic cells overflow into the bloodstream, where they may be seen in large numbers. These cells may also infiltrate the liver, spleen, lymph nodes, and other tissues throughout the body. Lymphoma is characterized by malignancy of cells found in lymphoid tissues and usually arises in the lymph nodes. Hodgkin disease is a special category of malignant lymphoma that is characterized by the presence of Reed-Sternberg cells. Hodgkin disease is more predictable in its spread than the non-Hodgkin types of lymphoma, and it is generally curable in the early stages. Non-Hodgkin lymphoma types comprise a large number of different disorders that involve malignant transformation of B cells, T cells, or NK cells. As with Hodgkin disease, earlier stages are more easily cured. However, the non-Hodgkin lymphomas tend to be unpredictable in their dissemination, and the prognosis is less certain. Plasma cell myeloma is a malignant transformation of mature, antibody-secreting B cells. Malignant plasma cells are monoclonal, and all produce identical antibodies, which accumulate in the blood. These cells have a predilection to settle in skeletal structures, where they cause bone demineralization and destruction. Hypercalcemia, bone fractures (pathologic fractures), and renal damage are common complications of plasma cell myeloma. Treatment for the various types of hematologic neoplasms continues to evolve, and excellent disease-free survival is commonly achieved when the disease is diagnosed in the early stages.

RESOURCES

Etiology and Classification of Hematologic Neoplasms

Henderson ES, Lister TA, Greaves MF, editors: *Leukemia*, ed 7, Philadelphia, 2002, Saunders.

Jaffe ES, Harris NL, Stein H, Vardiman JW, editors: *Pathology and genetics of tumours of haematopoietic and lymphoid tissues*, Lyon, France, 2001, IARC Press.

Lowenberg B, Griffin JD, Tallman MS: Acute myeloid leukemia and acute promyelocytic leukemia. *Hematology* 82–101, 2003. doi:10.1182/asheducation-2003.1.82.

McKenzie SB, Williams L: *Clinical laboratory hematology*, Upper Saddle River, NJ, 2015, Pearson Education, 3/E.

National Cancer Institute: https://seer.cancer.gov/statfacts/.

Ross JA, Potter JD, Reaman GH, et al: Maternal exposure to potential inhibitors of DNA topoisomerase II and infant leukemia (United States): a report from the Children's Cancer Group. *Cancer Causes Control* 7(6):581–590, 1996.

Tessoulin B, Eveillard M, Lok A, et al: *p53 dysregulation in B-***cell** *malignancies: More than a single gene in the pathway to hell. Blood Rev*, 2017. S0268-960X(16)30059-5. doi: 10.1016/j.blre.2017.03.001.

Myeloid Neoplasms

American Cancer Society: *Cancer facts and figures, 2015*, Atlanta, GA, 2003, Author.

Aster JC: The hematopoetic and lymphoid systems. In Kumar V, Abbas AK, Fausto N, Mitchell R, editors: *Robbins basic pathology*, ed 8, Philadelphia, 2007, Saunders, pp 421–478.

Barille-Nion S, et al: Advances in biology and therapy of multiple myeloma. *Hematology* 248–278, 2003. doi:10.1182/asheducation-2003.1.248.

Henderson ES, Lister TA, Greaves MF, editors: *Leukemia*, ed 7, Philadelphia, 2002, Saunders.

Hoffbrand AV, Moss PAH: The World Health Organization (WHO) classification of the tumors of haematopoietic and lymphoid tissues. In *Essential haematology*, ed 6, Oxford, England, 2011, Blackwell.

Keating MJ, et al: Biology and treatment of chronic lymphocytic leukemia. *Hematology* 153–175, 2003. doi:10.1182/asheducation-2003.1.153.

Linker CA: Blood. In Tierney LM, McPhee SJ, Papadakis MA, editors: *Current medical diagnosis and treatment*, ed 46, New York, 2007, McGraw-Hill, pp 493–547.

McKenzie SB, Williams L: *Clinical laboratory hematology*, Upper Saddle River, NJ, 2015, Pearson Education. 3/E,.

Melo JV, Hughes TP, Apperley JF: Chronic myeloid leukemia. *Hematology* 132–152, 2003. doi:10.1182/asheducation-2003.1.132.

Spivak JL, et al: Chronic myeloproliferative disorders. *Hematology* 200–224, 2003. doi:10.1182/asheducation-2003.1.200.

Lymphoid Neoplasms

Carbone PP, Kaplan HS, Musshoff K, et al: Report of the Committee on Hodgkin's Disease Staging Classification. *Cancer Res* 31:1860–1861, 1971.

Diehl V, Stein H, Hummel M, et al: Hodgkin's lymphoma: biology and treatment strategies for primary, refractory and relapsed disease. *Hematology* 225–247, 2003. doi:10.1182/asheducation-2003.1.225.

Iacobucci I, Mullighan CG: Genetic Basis of Acute Lymphoblastic Leukemia. *J Clin Oncol* 35(9):975–983, 2017.

Lister TA, et al: Report of a committee convened to discuss the evaluation and staging of patients with Hodgkin's disease: Cotswolds Meeting. *J Clin Oncol* 7(11):1630–1636, 1989.

Rytting ME, Jabbour EJ, O'Brien SM, Kantarjian HM: Acute lymphoblastic leukemia in adolescents and young adults. *Cancer* 2017. doi:10.1002/cncr.30624.

Vose JM, Chiu BCH, Cheson BD, et al: Update on epidemiology and therapeutics for non-Hodgkin's lymphoma. *Hematology* 241–262, 2002. doi:10.1182/asheducation-2002.1.241.

12

HIV Disease and AIDS

Faith Young Peterson

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- What are the common modes of HIV transmission, and how can infection be prevented?
- What is the scope of the HIV/AIDS epidemic in the United States and the world?
- How does infection with HIV lead to progressive immunodeficiency and AIDS?
- How has knowledge of the HIV life cycle led to the development of multidrug treatment strategies?
- How are $CD4^+$ cell counts and various clinical findings used to classify the stages of HIV disease and AIDS?
- What are the common systemic manifestations of AIDS and associated opportunistic infections?
- What are the current treatment recommendations for HIV disease and AIDS?

CHAPTER OUTLINE

Epidemiology, 234
History, 234
Types of HIV, 234
Transmission, 236
Prevention of Transmission, 238
Etiology, 239
HIV Structure, 239
HIV Binding and Infection, 239
Pathogenesis, 242
Effect of HIV on Immune Cells at the Cellular Level, 242
Viral Production and Cell Death, 242
Progression of HIV Infection from Seroconversion to AIDS, 244
CDC HIV Classification System, 245
Diagnostic Testing, 245
Monitoring the Progression of HIV, 246
Clinical Manifestations, 248
Systemic Manifestations, 248
Gastrointestinal Manifestations, 249
Pulmonary Manifestations, 249
Mucocutaneous Manifestations, 250
Gynecologic Manifestations, 251
Neurologic Manifestations, 252
Ocular Manifestations, 253
Cardiovascular Manifestations, 253
Manifestations in Other Systems, 253
Manifestations in Children, 254
Treatment, 254
Antiretroviral Therapy Recommendations, 254
Nucleoside Reverse Transcriptase Inhibitors, 255
Nucleotide Reverse Transcriptase Inhibitors, 255
Nonnucleoside Reverse Transcriptase Inhibitors, 255
Protease Inhibitors, 257
Fusion Inhibitors, 257
CCR5 Inhibitors, 257
Integrase Strand Transfer Inhibitors, 257
Other Treatments and Vaccines, 257

This chapter focuses on HIV disease and AIDS—from epidemiology to pathogenesis and management. Human immunodeficiency virus (HIV) is the prototypical public health infectious disease of the late twentieth century. It does not act like other infectious organisms that overwhelm the immune system. Although originally thought to be a rapid killer, it is a long-term, complex, chronic disease caused by persistent replication of HIV that leads to immunodeficiency. It triggers chronic widespread and diverse organ involvement with varying signs and symptoms. It encompasses all the armamentarium of a viral infection that has completed the evolutionary progression from animal to human. HIV has done more than just confuse and captivate scientists and health professionals; it also has mobilized risk groups and placed medicine and society at a crossroads of opinion. In this epidemic, the lines between privacy and public health and between morality and compassion have been debated. HIV disease is complex. However, in its complexity, it has opened the door to better understanding of the immune system.

Human immunodeficiency virus (HIV) infection and acquired immunodeficiency syndrome (AIDS) are acquired immunodeficiency disorders resulting in defective cell-mediated immunity, especially the decrease in $CD4^+$ or T helper/inducer lymphocytes. $CD4^+$ T cells are necessary for appropriate immune responsiveness because they are the cells that mediate between the antigen-presenting cells and other immune cells, such as B cells and other T cells. $CD4^+$ lymphocytes are characterized by the presence of the CD4 receptor.

EPIDEMIOLOGY

HIV infection is a primary immunodeficiency disease caused by the retroviruses HIV type 1 and HIV type 2. Despite research and public health surveillance and prevention activities, the virus has continued mutating and spreading globally. HIV infects people worldwide. Since its identification in the early 1980s, the HIV global epidemic continues with an estimated 36.9 million people living with HIV worldwide as of 2014 and a total of 1.2 million AIDS deaths worldwide in 2014. The number of adults living with HIV is estimated to be 34.3 million. Of those, 17.4 million people are women. There are also 2.6 million children (<15 years) estimated to be living with HIV globally. AIDS-related deaths worldwide have fallen by 42% since 2004. Selected aspects of the global impact of HIV and AIDS are illustrated in Table 12.1.

According to the World Health Organization (WHO), the number of new HIV infections globally declined 35% since 2000, attributable to expanded and improved HIV programs. There were approximately 2 million people estimated to be newly infected with the HIV virus in 2014. New HIV infections in children have decreased 58% since 2000 with 220,000 children newly infected with HIV and 150,000 deaths in children in 2014. Of the newly infected people with HIV, 15 countries account for more than 75%. These 15 countries include Brazil, Cameroon, China, India, Indonesia, Kenya, Mozambique, Nigeria, Russian Federation, South Africa, Uganda, United Republic of Tanzania, United States of America, Zambia, and Zimbabwe. HIV infection rates are influenced by location and population and as such continue to increase in sub-Saharan Africa, which has the highest number of people living with HIV (24.8 million), as well as in Eastern Europe (1.5 million) and central Asia (5.0 million). In those countries, people infected with HIV have more limited access to testing and medication for treatment, as well as limited information for prevention because of the effects of gender inequity and harmful social norms that drive transmission.

In the United States it is estimated by the Centers for Disease Control and Prevention (CDC) that there are more than 1.2 million persons living with HIV/AIDS in the United States, with 47,352 people newly diagnosed with HIV infections in 2013. In the United States and other industrialized countries where access to medication, care, and prevention is greater, the number of patients diagnosed with and dying from AIDS is stable or declining (Fig. 12.1). In 2013 26,688 people were diagnosed with AIDS. However, the CDC estimates that approximately 12.8% of the people who are infected with HIV in the United States may be unaware that they are infected. Since 1981, 658,507 people with AIDS in the United States have died.

Current statistics show that of the people diagnosed with HIV/AIDS in the United States, men who have sex with men (MSM), injection drug users, racial and ethnic minorities, and women of color are disproportionately affected. Among racial and ethnic groups, 41% of people living with HIV infection and most of the new cases are African Americans, especially those in the youngest age group. Hispanics account for approximately 20% of people living with HIV infection and 21% of new HIV infections. According to the CDC, 63% of new HIV infections occur in MSM. New HIV infections occur in approximately 8% of injection drug users, who represent 15% of those living with HIV.

The rate of persons living with an HIV/AIDS diagnosis also varies by geographic region. The Northeast and South report the highest numbers of HIV/AIDS diagnoses, particularly in urban areas. For example, the South reported 45% of new AIDS diagnoses followed by 24% in the Northeast, 19% in the West, and 13% in the Midwest. The highest rates of new diagnoses of HIV are in African American men and women in all geographic regions except in the West.

Approximately 20% of new HIV infections occur in women, and 23% of those living with HIV are women. HIV infections in women are usually due to heterosexual contact (84%) or injection drug use (16%). It is thought that noninjection drugs (such as crack cocaine or methamphetamine) may also contribute to the spread of HIV/AIDS because of sex trading for drugs, shelter, or money. African American and Hispanic women account for 80% of all HIV diagnoses in women in the United States.

People living with HIV older than age 50 comprise 20% of all HIV infections in the United States. Initial infection in this age group is often as a result of a decreased perception of risk. Despite improved diagnoses and treatment, people older than 50 comprise 50% of HIV deaths. In the 50 to 54 age group, HIV is one of the top causes of death in the United States.

History

In 1981 the first descriptions of immunodeficiency disease in previously healthy persons appeared in the medical literature. At that time, previously healthy young homosexual men contracted unusual diseases for their age group in increasing numbers, such as *Pneumocystis jiroveci (carinii)* pneumonia (PCP) and Kaposi sarcoma, that researchers identified as HIV. The first evidence of alternative forms of transmission of the virus by blood and blood products appeared in 1982. All these early patients were shown to have a type of HIV virus called *HIV-1.* It was at this time that the term *acquired immunodeficiency syndrome (AIDS)* was first used. However, the specific retrovirus causing HIV infection and AIDS was not isolated until the early to mid-1980s. The timeline of HIV history is found in Table 12.2.

Types of HIV

HIV is a type of retrovirus from the subfamily Lentivirinae, with *Lentivirus* being its only genus. This subfamily is so named from the Latin word *lentus,* meaning "slow," because infection develops gradually. HIV-2, a related but distinct retrovirus, was later identified in 1986 and is most closely related to simian immunodeficiency virus. HIV-2 is differentiated from HIV-1 by a longer clinical latency period from the onset of infection to the development of symptoms. It is also characterized by having lower plasma HIV-2 RNA viral loads and lower mortality rates. HIV-2 infection can progress to AIDS, even though it appears to be less virulent than HIV-1. It is also possible to be coinfected with both HIV-1 and HIV-2.

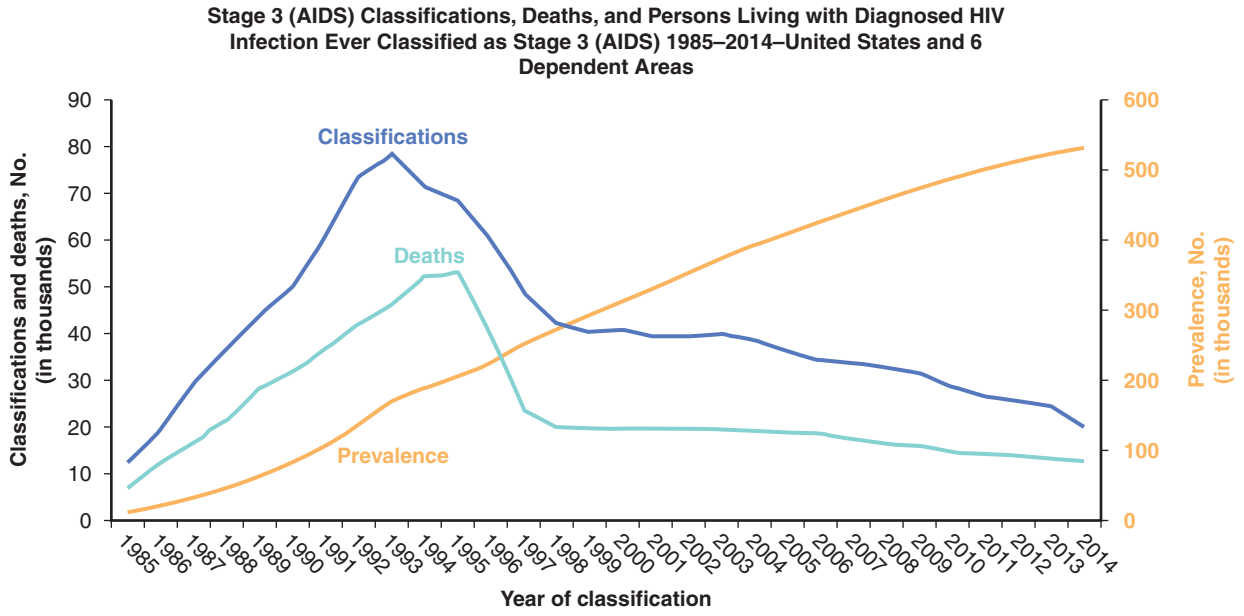
HIV-1 and HIV-2 are found worldwide. They are similar in structure and function, but are differentiated from each other by their envelope glycoproteins, point of origin, and latency periods. The point of origin for HIV-1 is Central Africa, and for HIV-2 it is West Africa. HIV-1 is the causative organism of most cases found in Central Africa, the United States, Europe, and Australia. HIV-2 is found primarily in West Africa or in countries with strong socioeconomic ties to West Africa (e.g., France, Spain, Portugal, and former Portuguese colonies).

Many subspecies or strains of HIV also exist because of the rapid rate of HIV virion mutation. The subspecies may exist in different hosts, as well as within an individual host. Currently, at least 10 subtypes of HIV-1 have been identified: group N (YBF30), group O, and group M with 8 subtypes (A, B, C, D, E, F, G, and H). Research is currently focusing on the identification of HIV subtypes and strains in different

TABLE 12.1 Global Health Considerations for HIV/AIDS

Country/ Region	Prevalence of Disease	Cultural Factors	Most Common Means of Transmission	Treatment	Economic/Social Impact
Sub-Saharan Africa	25.8 million people affected. This region carries most of the world's HIV/AIDS burden. Nigeria, South Africa and Uganda account for 48% of all new HIV infections. Adult prevalence % = 4.8% Adult and child deaths due to AIDS = 790,000	Gender inequalities: Males dominate sexual decision making and adolescent girls and young women are disproportionately infected.	Heterosexual sex and mother-to-child transmission	Treatment is available but it depends on the country. Treatment coverage = 90% of those who test positive for HIV. 76% of people on HAART achieve viral suppression. However, Central African Republic, Democratic Republic of the Congo, Nigeria and South Sudan have high HIV burden, low treatment coverage, and no or little decline in new HIV infections.	Life expectancy depends on access to treatment. Untreated populations have decreased life expectancy. Stigma associated with disease causes people to lose property.
Middle East and North Africa	240,000 people affected. Adult prevalence % = 0.1%. Adult and child deaths due to AIDS = 12,000	Gender inequalities: Males dominate sexual decision making. Marked stigma.	Heterosexual sex and mother-to-child transmission	Only 11% of infected people are receiving antiretroviral therapy New infections have risen by 7%. AIDS-related deaths have risen by 66% since 2005.	
Asia/Pacific	5.0 million people affected. Adult prevalence % = 0.2%. Adult and child deaths due to AIDS = 240,000	Growing male demographic has led to a growing sex industry.	Intravenous drug use, prostitution, and transmission through migrant workers	Medical coverage in rural areas is poor, so a large proportion of the population must pay out of pocket. Treatment coverage: 31%	Cost of therapy generally exceeds annual income. Stigma is so great that many people do not disclose their HIV status to their families and thus do not seek treatment.
Western and Central Europe and North America	2.4 million people affected. Adult prevalence % = 0.3%. Adult and child deaths due to AIDS = 26,000.	Continued stigma	Men having sex with men (MSM), intravenous drug use, heterosexual transmission, and prostitution	Treatment available to most people. The percentage of people receiving treatment is highest at 51%. New infections have risen by 8% by 2005.	Cost of therapy is high but programs are available for assistance
Eastern Europe and Central Asia	1.5 million people affected. Adult prevalence % = 0.9%. Adult and child deaths due to AIDS = 62,000	Continued stigma	MSM, intravenous drug use, heterosexual transmission and prostitution	Treatment availability varies. Russian Federation has low treatment coverage and no or little decline in new HIV infections. New infections have risen by 5% since 2005. AIDS-related deaths have risen by 5% since 2005.	
Caribbean	280,000 people affected. Adult prevalence % = 1.1%. Adult and child deaths due to AIDS = 8,800	Continued stigma	MSM, intravenous drug use, heterosexual transmission, and prostitution	New infections declined most = 40% since 2005	
Latin America	1.7 million people affected. Adult prevalence % = 0.4%. Adult and child deaths due to AIDS = 41,000	MSM are highly stigmatized, so prevention efforts overlook this group. Drug use is commonplace	MSM, intravenous drug use, and prostitution	45% of infected people are receiving antiretroviral therapy. The government provides therapy for free.	Government has had a strong and positive response to the epidemic: its efforts have reduced stigma, improved social reintegration, and reduced HIV prevalence among high-risk populations.

UNAIDS Press Release. UNAIDS report shows that 19 million of the 35 million people living with HIV today do not know that they have the virus. Geneva, July 16, 2014.



Note. Deaths of persons with HIV infection, stage 3 (AIDS) may be due to any cause.

FIG 12.1 Estimated AIDS incidence and deaths among persons with AIDS, by year of diagnosis or death and year-end prevalence, United States, 1985 to 2014. Data were adjusted for delays in reporting of cases and deaths. (From Centers for Disease Control and Prevention. https://www.cdc.gov/hiv/library/slideSets/index.html.)

populations and geographic areas. For example, in the United States, Europe, and Australia, most infected persons have HIV-1, subtype B, whereas in India, HIV-2 is found near Goa, and HIV-1 strains A, B, and C are also present.

Transmission

HIV-1 and HIV-2 are relatively weak viruses outside of the body. HIV viruses can infect people through three major types of transmission: sexual transmission via semen or vaginal and cervical secretions through homosexual, bisexual, or heterosexual intercourse; parenteral transmission via blood, blood products, or blood-contaminated needles or syringes; and perinatal transmission in utero, during delivery, or in breast milk. Of these forms of transmission, sexual transmission through unprotected vaginal or anal intercourse is the most common mode of infection globally. HIV is found in every bodily fluid and is broadly disseminated in the body. In very low titers, HIV is known to be present but has not been shown to be transmitted via urine, saliva, tears, cerebrospinal fluid, amniotic fluid, and feces. HIV is not known to be transmitted via aerosol routes. In the United States those at greatest risk of HIV infection include (1) MSM; (2) intravenous drug users (IVDs) who share needles or syringes; (3) sexual partners of those in high-risk groups, particularly heterosexual women; and (4) infants born to infected mothers. Heterosexual intercourse with infected partners, contact with contaminated blood, and prenatal or perinatal exposure of the infant are the major routes of transmission of HIV in Africa, South and Southeast Asia, and developing countries. In these countries, an equal proportion of males and females are infected.

Common modes of transmission include needle/syringe sharing between IVDs, unprotected sex with infected partners, recipients of HIV-contaminated blood or blood products or infected semen during artificial insemination, unanticipated needle or scalpel injury during care or surgical treatment of infected patients, and neonatal transmission from an infected mother to her infant. In both men and women, concomitant sexually transmitted diseases or genital lesions increase the risk of HIV infection. In women, high-risk heterosexual contact is influenced by lack of HIV knowledge, low socioeconomic status, low perception of risk, concomitant drug or alcohol use, relationship dynamics such as fear of abuse or loss of relationship, and the increased biological vulnerability of HIV contraction during vaginal intercourse, especially in the presence of other sexually transmitted diseases or vaginal inflammation. In both men and women, the use of noninjection drugs (such as crack cocaine) contributes to HIV transmission by decreasing inhibition, allowing the person to engage in risky sexual behaviors or to trade sexual relations for drugs or money.

Blood bank screening and testing procedures have nearly eliminated the transmission of HIV-contaminated blood in the United States. In the United States all blood and blood products have been tested by HIV-1, HIV-2, and HIV-1 p24 antigen tests since 1996. However, this route of transmission continues in third-world countries, where there is a high number of HIV-infected persons and much of the blood and blood products are not screened before use.

Health care workers who are exposed to blood or infected body fluids or needles/sharp instruments are at risk of contracting HIV. The risk of developing HIV is greatest for those health care workers who have a deep injury with visible blood from a contaminated needle or sharp instrument or who have a direct puncture into an artery or vein. They also are at risk if they have prolonged blood–skin contact, especially if extensive. The risk of infection is much lower when both universal precautions and postexposure prophylaxis are employed.

Transmission from an infected mother to her infant may occur in the intrauterine period, in the intrapartum period at the time of delivery, or in the postpartum period via breast feeding; it may also be transmitted in some cultures from saliva attributable to premastication. Of these, intrapartum transmission at the time of delivery is thought to be the most common. HIV infection does not cause any specific congenital abnormalities, but there is an increased risk of spontaneous

TABLE 12.2 HIV History Timeline Synopsis

1900	Retroviruses identified as cause of cancer in chickens
1930s	HIV-1 precursor virus crossed species from chimpanzees to humans during hunting from contact with infected chimpanzee blood
1950s	Positive tests on serum from man in Leopoldville (now Kinshasa) Congo
1960s	Virus reaches Haiti
1968	First case in United States in sexually active 15-year-old African American male in St. Louis
1980	First retrovirus identified in humans affecting T cells (human T-cell lymphocytic virus, HTLV-1)
1981	Syndrome of HIV first reported in healthy young homosexual men in Los Angeles and New York. First clinical article in MMWR.
1982	Identification of HIV transmission by blood/blood products with first use of term AIDS. Syndrome also identified in women, infants, Haitians, and persons who had received blood or blood products
1983	CDC publishes first Blood and Body Fluid Precautions.
	First WHO meeting on AIDS.
1985	First HIV-1 antibody testing (ELISA and Western Blot).
	First international conference on AIDS
1987	HIV-2 identified in visitor to United States from West Africa.
	First anti-HIV drug approved.
	CDC revises guidelines with identification of mucocutaneous exposure – "Universal Blood and Body Fluid Precautions."
1988	World Health Organization (WHO) declares December 1 as World AIDS Day.
1989	FDA approves aerosolized pentamidine for PCP prophylaxis.
	Second federal commission on AIDS established.
1990	Ryan White Comprehensive AIDS Resources Emergency Act signed into law.
1991	Earvin (Magic) Johnson announces his HIV-positive status.
	OSHA mandates Universal Precautions.
1992	FDA approves the use of AZT and zalcitabine (ddC, Hivid) for concurrent treatment of HIV infection.
1993	Scientists reveal AZT-resistant HIV strains are being transmitted to newly infected individuals.
1994	CDC announces that AIDS is leading cause of death among Americans 25–44 years old
1995	FDA approves the first protease inhibitor (saquinavir).
	Establishment of the Presidential Advisory Council of HIV/AIDS.
	U.S. PHS recommends that pregnant women be offered AZT to reduce the possibility of perinatal transmission.
1996	Combination antiretroviral therapy (HAART) is introduced.
1997	CDC warns that the number of AIDS cases among people over age 50 is slowly increasing.
1999	Researchers discover origins of HIV-1 from chimpanzee species (see 1930s).
	Behavioral studies report recidivism in regard to safer sex practices among MSM in both Europe and the United States.
2006	WHO declares March 8 as National Women's & Girls HIV/AIDS Awareness Day.
	SMART trial found that episodic antiretroviral therapy more than doubles risk of AIDS or death in people with HIV infection
2014	Worldwide, HIV/AIDS is the leading cause of death for women in their reproductive years.
	WHO report indicates that 19 million of the 35 million people living with HIV do not know their HIV status.
2015	WHO issues new guidelines for treatment of all HIV-infected persons at the time of diagnosis regardless of $CD4^+$ T cell count.

Compiled from the following sources:

Balter M: Virus from 1959 sample marks early years of HIV, *Science* 279:801, 1998.
Garry RF et al: Documentation of an AIDS virus infection in the United States in 1968, *JAMA* 260(14):2085-2087, 1988.
Lemey P et al: Tracing the origin and history of the HIV-2 epidemic, *Proc Natl Acad Sci USA* 10.1073/pnas.0936469100, 2003.
Ungvarski PJ: The past 20 years of AIDS. *AJN,* 101(6):26-29, June, 2001.
Varmus H: Retroviruses. *Science* 240:1427-1435.
UNAIDS: Press Release: *UNAIDS report shows that 19 million of the 25 million people living with HIV today do not know that they have the virus and Core Epidemiology Slides,* Geneva, July 16, 2014, World Health Organization (WHO).
Division of HIV/AIDS Prevention: *HIV in the United States: At a Glance,* Atlanta, GA, July, 2014, Centers for Disease Control and Prevention (CDC).
Division of HIV/AIDS Prevention: *HIV and AIDS in the United States by Geographic Distribution,* Atlanta, GA, May 11, 2015, Centers for Disease Control and Prevention (CDC).

abortion. The overall risk to the fetus of HIV transmission is estimated to be between 15% and 40% for each pregnancy, with increasing risk in subsequent pregnancies for each HIV-positive fetus born. Increased risks of antepartum transmission include increased maternal viral load or high viremia during early infection, advanced maternal clinical disease as evidenced by low $CD4^+$ counts, and breaks in the placental barrier. Increased risks of intrapartum transmission include high maternal viral load at the time of delivery, prolonged ruptured membranes (more than 4 hours), infant exposure to blood/secretions, abruptio placentae, infant prematurity, and the presence of coinfections. The rate of HIV perinatal transmission is reduced with the use of antiretroviral therapy during pregnancy and during the first months of the infant's life.

Routine social contact with people who are HIV positive does not increase one's risk of HIV infection. The following examples are safe practices and will not cause exposure to HIV infection: using public restrooms, swimming in public swimming pools, touching or hugging someone who is HIV positive, and eating with community utensils or in restaurants. Insects such as mosquitoes cannot transmit the HIV virus to humans.

Exposure to HIV does not mean that one will contract HIV or AIDS, and it does not mean rapid progression. HIV-1 transmission rates are generally low and range from 0.1% to 10% per exposure. The interacting forces between viral and host factors influence whether a person will contract HIV infection, particularly the amount and virulence of the virus and the host's response by T-cell–mediated cytotoxicity or by cytokines. In studies of patients with hemophilia who received tainted blood products, 10% to 25% of the individuals evaded infection. Because of genetic differences that either increase or decrease susceptibility to the infection, the risk of acquiring HIV and the response to infection also vary within populations. Some individuals, despite high-risk exposure, do not exhibit any signs of infection or immunodeficiency.

Researchers have identified an HIV resistance mutation of the *CCR5* gene, called *CCR5-delta 32,* which is associated with natural resistance to HIV infection in certain people. When inherited from both parents, the mutant *CCR5-delta 32* gene appears to protect individuals from infection even after multiple exposures. When only one gene is inherited, the progression to AIDS tends to be slower. The *CCR5* gene is not equally distributed among people. Persons of Caucasian American and Caucasian-European descent have the highest number of mutant allele genes, approximately 10%, whereas Native American, African, and East Asian people have the lowest number of mutant alleles. A recent study on the protective effect of the *CCR5-delta 32* allele in Mexican women found that there was an inverse relationship between allele frequency and the risk of HIV-1 transmission.

Researchers have also found that other mutations in the *CCR5* gene can delay progression of HIV infection, such as polymorphism-2459 (A/G). Other proteins bind to *CCR5* and demonstrate antiviral activity, such as macrophage inflammatory protein-1α (MIP-1α) and MIP-β, human neutrophil peptide 1 (hNP-1), as well as other chemokines. People who have fewer genes encoding CCL3L1, a potent HIV-blocking protein that interacts with CCR5, are more susceptible to HIV infection and have more rapid progression to AIDS, whereas individuals with natural killer (NK) cells that produce interferon-γ, tumor necrosis factor-α (TNF-α), CCL3, CCL4, and CCL5 are less likely to develop HIV infection. Because the susceptibility to and progression of HIV are multifactorial and exhibit significant interindividual variation, many aspects of the pathogenesis of HIV are still unknown.

Prevention of Transmission

Prevention is essential, because effective management of HIV is expensive and a cure is not yet possible. However, one-time exposure to information or a single message is usually less successful than programs that teach prevention skills and reinforce positive behavior. The primary way to prevent transmission is to use safe sex practices. Safe sex practices include abstaining from sex, using a condom (barrier protection) during sexual intercourse, avoiding multiple sexual partners, and knowing the HIV status of all sexual partners. It is important that education regarding safe sex practices be tailored to appropriate age groups, ethnicity, culture, and sexual preference. Patient visits to health care providers are an excellent opportunity to encourage individual HIV protection.

Spermicides such as nonoxynol 9 or C31G do not inactivate HIV or other sexually transmitted microorganisms. No studies suggest any benefit from using progestins such as levonorgestrel (Norplant) or medroxyprogesterone (Depo-Provera), the diaphragm, or oral contraceptives to prevent HIV transmission. The early use of antepartum and intrapartum antiretroviral therapy and avoidance of breast feeding can prevent maternal–child HIV infection.

HIV infection in drug users can be prevented with the use of sterile needles via improved access to clean needles and avoidance of dirty or shared needles. Such intervention includes needle/syringe exchange programs for IVDs and cleaning of dirty needles with bleach before use. When using bleach, the user must rinse out all blood first; then fill the needle and syringe with full-strength bleach at least three times for 30 to 60 seconds.

Medical and health care personnel are at risk through occupational exposure to blood and body fluids. The degree of occupational risk varies with the type and severity of exposure. For example, the average risk for HIV transmission after a needle-stick exposure to HIV-infected blood is estimated to be approximately 0.3%. Self-protection through the use of standard precautions can decrease risk by reducing exposure. Health care providers should carefully wash their hands before and immediately after patient contact even when using gloves. It is essential to wear disposable gloves for any actual or potential contact with blood or body secretions, when handling items contaminated with blood or body fluids, when performing finger sticks or heel sticks, or when the health care provider has scratches or cuts on the hand.

Gowns or plastic aprons, masks, goggles, or face shields should be worn to protect the face and clothing when there is risk of splashes and airborne droplets of blood or body fluids. Protective gear should be changed between patients. Careful prevention of parenteral exposure when using needles or other equipment should be emphasized. Needles and sharp implements should be disposed in rigid, puncture-proof containers. Such implements should not be bent, broken, or recapped before disposal. In combative patients who must have blood drawn or injections given, careful use of humane and limited restraint devices may be necessary to prevent injury to the involved health care workers. Resuscitation bags and masks should be readily available to minimize the need for mouth-to-mouth procedures.

Unfortunately, accidents necessitating the development of postexposure prevention protocols do occur. If a health care worker sustains an injury with significant exposure to blood or body fluids such as a needle stick and the source patient is HIV positive, prophylaxis should be given immediately after a workplace postexposure prevention protocol. If the source patient tests negative, if there is concern about new infection, or if the HIV status is unknown, the health care worker should be offered prophylaxis. The selection of a drug regimen for HIV postexposure prophylaxis must balance the risk for infection against potential toxicities and side effects of the medication(s). The postexposure protocol usually involves the administration of three medications—emtricitabine/tenofovir (Truvada) and raltegravir (Isentress)—for 4 weeks. This combination of medications is usually well tolerated and has fewer drug interactions. The length of administration of the agents depends on multiple factors and may be longer. This same protocol has been advocated for use as postsexual exposure prophylaxis.

KEY POINTS

- HIV disease is a primary immunodeficiency disorder caused by viral infection of $CD4^+$ cells. It is a major health concern because it causes chronic, severe, long-term disease in industrialized countries with access to medication, care, and prevention. In third-world countries where therapy is unavailable, the prognosis for HIV is very poor and death is more likely.
- HIV types 1 and 2 are retroviruses that primarily infect $CD4^+$ lymphocytes and macrophages. HIV-1 is the primary causative virus infecting persons in Central Africa, the United States, Europe, and Australia.
- HIV is acquired primarily through sexual transmission via semen and vaginal and cervical secretions; through parenteral transmission via blood, blood products, and contaminated needles/syringes; and through perinatal transmission from an infected mother to her infant antepartum, intrapartum, and postpartum via breast milk.
- HIV is known to be present in but is not believed to be transmitted via urine, saliva, tears, cerebrospinal fluid, amniotic fluid, feces, or aerosols.

- Those at greatest risk of HIV infection include homosexual and bisexual men, IVDs who share needles or syringes, sexual partners of those in high-risk groups, and infants born to infected mothers.
- The use of safe sex practices (such as condoms) and safe parenteral practices (such as sterile needles/syringes) decreases the risk of infection.
- Exposure to blood and body fluids of infected individuals through skin, mucous membranes, and accidental needle sticks is the primary risk factor for health care workers. The universal use of standard precautions decreases the risk of infection.
- After significant accidental exposure to HIV-infected blood or body fluids, it is recommended that health care workers receive postexposure antiretroviral medication as soon as possible after exposure and as needed for 4 weeks after the incident.

ETIOLOGY

HIV Structure

HIV is an RNA retrovirus that causes a defect in cell-mediated immunity that may progress to AIDS. The viral RNA must be converted to DNA before the viral genes can be expressed to make copies of the RNA virus. Like other retroviruses, HIV differs from DNA viruses in that the RNA genome cannot replicate without undergoing conversion into DNA.

HIV consists of a core or nucleocapsid containing two strands or chains of RNA, protein, and enzymes surrounded and protected by a spherical lipid bilayer viral envelope that is 0.0001 mm in diameter. Between the envelope and core is a protein layer called *p17*. The nucleocapsid or core is composed of a protein called *p24*. Within the nucleocapsid, the two strands of RNA compose the HIV genome (Fig. 12.2). The HIV genome consists of at least nine genes. The *gag* gene encodes the core antigen proteins that form the virus particle. The *pol* gene encodes reverse transcriptase proteins. The *env* gene encodes the viral envelope protein glycoprotein gp160, which is split into two fragments, gp120 and gp41, by cellular protease.

Several other genes have been identified, including *tat, rev, nef, vif, vpr,* and *vpu*. These genes are primarily regulatory genes. The *tat* gene encodes proteins that regulate HIV replication and can accelerate HIV viral protein production. It is controlled by *tat*-binding protein. The *rev* gene encodes proteins that regulate viral messenger RNA expression. Rev proteins inhibit regulatory proteins, allowing the transport of HIV RNA from the nucleus. Rev proteins also enhance viral structural gene production. The *vif* (virion infectivity factor) gene appears to increase the ability of the virus to infect other cells. It suppresses the human protein (CEM 15) that inhibits HIV-1.

The HIV genome contains all the information regulating the virus's structural format and growth during its life cycle. The enzymes within the core also are very important because they facilitate the conversion of RNA to DNA. This conversion is the means of information transfer. The enzymes include reverse transcriptase, integrase, and protease. Reverse transcriptase is composed of two associated enzymes called *polymerase* and *ribonuclease*. It is the unique enzyme in HIV that allows the virus to copy RNA into DNA. Protease is a complex enzyme that works as a "molecular scissors." It splits the other viral components by a process known as *autocatalysis*. Immature, noninfectious virions containing inactive gag/pol, a long precursor protein, are released in the plasma, where they are cleaved by protease into smaller active units. Protease also clips p55, the core gag viral protein precursor, into smaller molecules and is needed to facilitate final mature viral assembly for HIV to be infectious. In other words, HIV infection does not occur unless *protease* activates the virions. Protease also cleaves and inactivates other cellular proteins, including receptor-interacting protein kinases 1 and 2 (RIPK1/RIPK2). The inactivation of these proteins, which are involved in innate immunity and apoptosis, is another way the virus counters host defense mechanisms.

The viral envelope consists of a membrane derived from the host cell. Viral glycoprotein studs protruding from the cell membrane make it look like a studded ball (Fig. 12.3). Gp120 and gp41 are the two HIV envelope proteins that cover the viral particle surface. Gp120 is the most external and distal part of each "stud," whereas gp41 is the bridge that holds it onto the virion surface. It is thought that the interaction between gag and the long cytoplasmic tail of gp41 play an important role in viral envelope glycoprotein incorporation. The surface envelope also contains other cell-surface proteins derived from the host cell containing adhesion molecules. Although the viral particle (virion) is usually nearly spherical, great diversity is found in size and shape, such as comet-shaped virions.

HIV Binding and Infection

At the time of exposure to the HIV virus, infection occurs when the virus moves across the epithelium or mucosal membrane of the body. Once inside the body, HIV particles are attracted to cells with receptors on their surface called CD4 cells. The HIV envelope protein gp120 specifically binds to the CD4 receptor in the mucosal or other tissue at or near the site of exposure. For example, during heterosexual

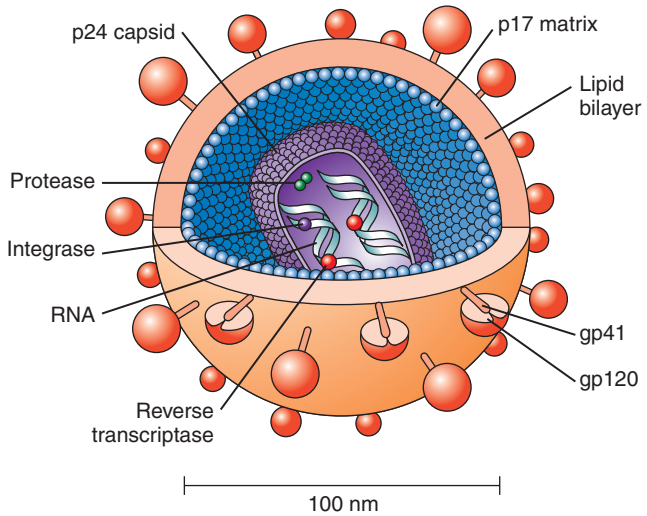


FIG 12.2 HIV particle showing the p24 capsid protein surrounding the two strands of viral RNA.

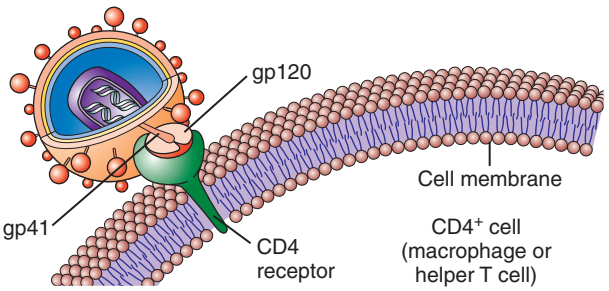


FIG 12.3 Schematic view of a retrovirus particle. The core is surrounded by an envelope that is derived from host membranes enriched with viral glycoproteins (gp120, gp41). Interaction of the envelope glycoproteins with a host-encoded cell surface receptor (CD4) is shown.

transmission of HIV, the virus is attracted to $CD4^+$ cells in the mucosal membranes, such as Langerhans cells, or dendritic cells, which capture the virus and transport it to $CD4^+$ T cells or proximal lymphoid tissues.

HIV virus is initially attracted to the $CD4^+$ cells that have the highest number of CCR5 coreceptors on their surfaces. The numbers of CD4 receptors on cell surfaces vary depending on the kind of $CD4^+$ cells—for example, $CD4^+$ T helper/inducer cells have large numbers of CD4 receptors while macrophages have lower levels of surface CD4. The CD4 receptor is found on many types of cells, including T cells; microglial cells; monocyte-macrophages; follicular dendritic cells; immortalized B cells; retinal cells; Langerhans cells in the skin; bone marrow stem cells; cervical cells; bone marrow–derived circulating dendritic cells; and enterochromaffin cells in the colon, duodenum, and rectum. Of these cells, the $CD4^+$ T helper/inducer cells are most often implicated and involved in the process of infection. Fig. 12.4 illustrates a group of HIV-infected $CD4^+$ cells imaged by scanning electron micrography.

The kind of $CD4^+$ cells that are attracted to the virus change over time. When the virus primarily affects T cells with high levels of surface CD4, it is known as *T tropic*. When the virus is attracted to macrophages with lower levels of surface CD4, the virus is called *M tropic*. For example, M tropic viruses occur especially in the brain, infecting microglial cells and perivascular macrophages that migrate into the brain. When the virus affects both macrophages and $CD4^+$ T cells, it is called *dual tropic*. Most often M tropic and dual tropic variants occur when the host is markedly immunodeficient late in the disease course.

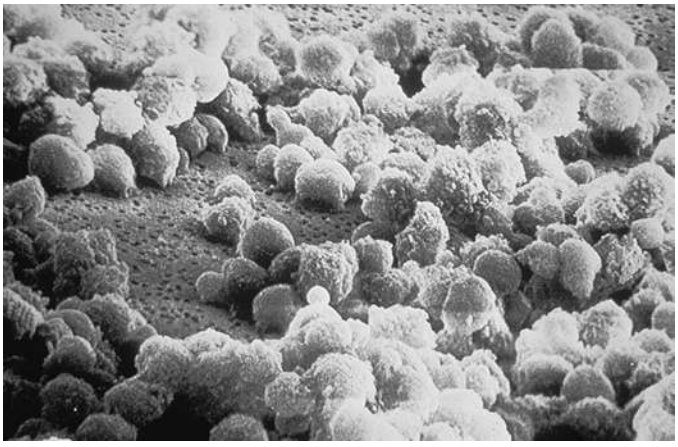


FIG 12.4 Scanning electron micrograph (low magnification) of a population of HIV-infected lymphocytes. (Courtesy Centers for Disease Control and Prevention, Atlanta, Georgia.)

Usually T cells are infected before the onset of symptoms. $CD4^+$ T cells are composed of two subsets: T helper-1 (T_H1) and T helper-2 (T_H2). The T_H1 subset produces interferon-γ and interleukin-2 (IL-2). The T_H2 subset produces IL-4, IL-6, and IL-10. Of these two subsets, the one that is markedly decreased in advanced disease is T_H1.

However, CD4 alone is not sufficient for fusion of the virion and host cell. A number of important coreceptors on the target cells called *chemokines* are necessary for the virus to gain entry into cells. These important chemokine coreceptors must be present for the virion to fuse with the host cell. The chemokine called *CCR5* must be present for the HIV particles to bind to the $CD4^+$ cells, especially in early infection, and another chemokine receptor named *CXCR4* is often present in later infection. Since 1996 when the coreceptors were first discovered, a number of other coreceptors have been identified, including APJ, CCR2b, CCR3, CCR8, CCR9, CX3CR1, CXCR4, GPR1, GPR15, STRL33, US28, and V28. The function of most of these coreceptors is unknown. It is hypothesized that some of the coreceptors may be needed for various strains of HIV, for HIV infection in infants and children, or for infection of the brain and nervous system.

The gp120 portion of the virion envelope must combine with the first receptor, CD4. This is done primarily through the V3 variable loop of gp120. It then changes shape by refolding. In the second shape, it combines with the second receptor, either CCR5 or CXCR4, to fuse with the cell. Once the HIV particle is bound to both the CD4 receptor and the chemokine receptor on the host cell, gp41 implants itself in the cell membrane (Fig. 12.5). This sequence of events causes the viral particle and the cell to fuse. The binding of the virion and host cell increases adenosine triphosphate release with activation of a variety of purinergic receptors, which increases intracellular free calcium and opens ion channels (Panx-1 hemichannels). This facilitates viral entry into the host cell. The core of the virus is then injected into the cytoplasm of the host cell, and infection is produced.

Once in the cytoplasm, a single-stranded DNA copy is made by reverse transcriptase from the viral RNA. Using the single-stranded DNA as a template, DNA polymerase copies it to make a second DNA strand and destroys the original RNA strands. The accuracy of DNA transcription is poor, with mutations occurring frequently. This tendency to mutate makes HIV resistant to antiviral medications.

Once formed, the new viral DNA, called *viral provirus* or *preintegration complexes (PICs)*, migrates to the cell nucleus and is actively transported in the nuclear compartment. Inside the nucleus, integrase splices the viral DNA, or provirus, into the host cell's DNA. Once in the host cell's DNA, the viral DNA is replicated together with the host cell's DNA during every cell division. Now the viral DNA is permanently part of the host cell's DNA (Fig. 12.6).

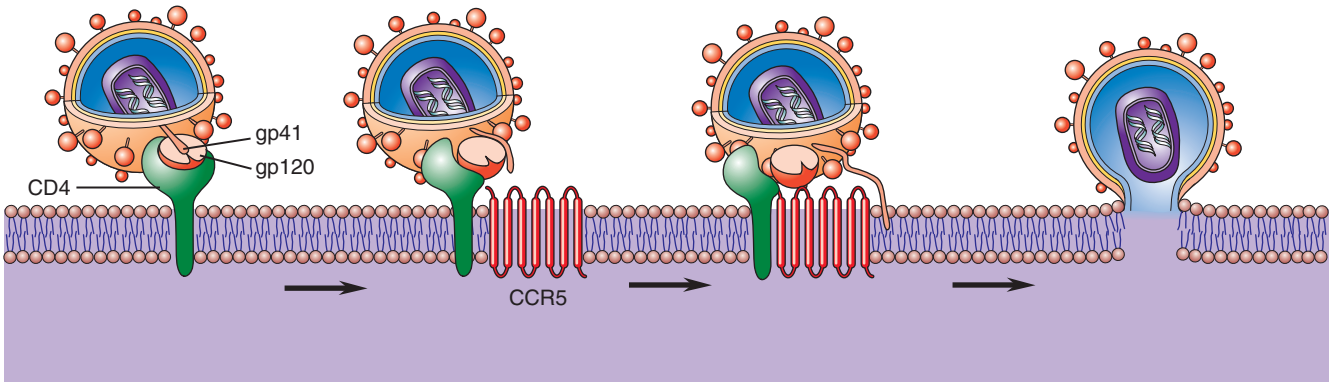


FIG 12.5 Early HIV infection, M tropic. In HIV infection, the virus must bind both a CD4 receptor and a coreceptor to fuse with the host cell. In the M tropic phase, the key coreceptor is CCR5.

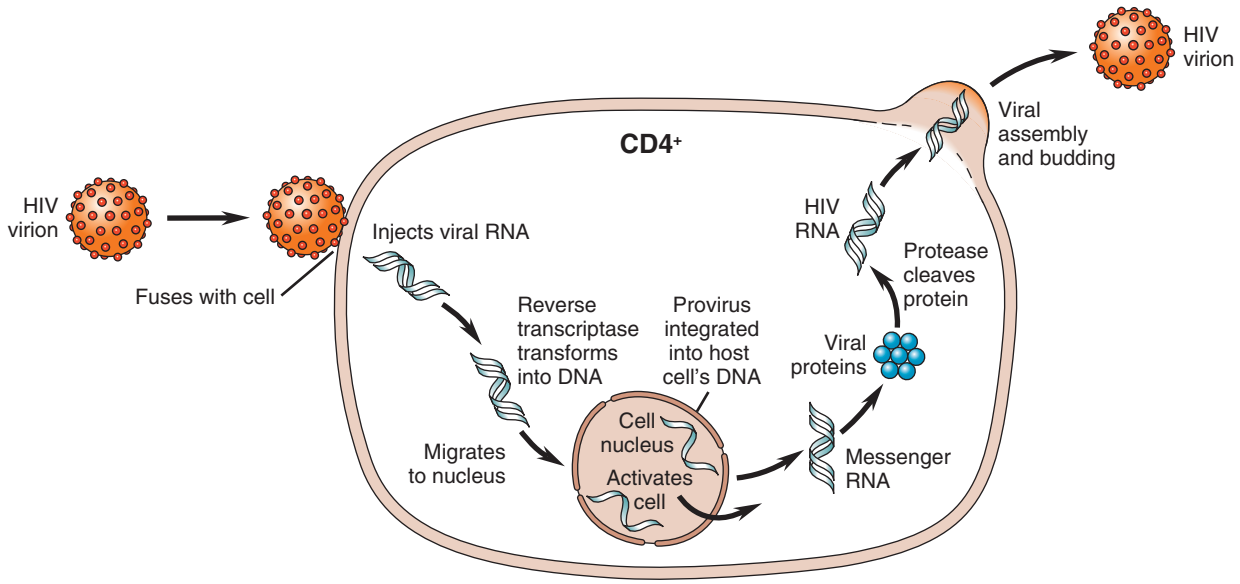


FIG 12.6 HIV life cycle. A schematic representation from the time of initial HIV fusion with a host cell to integration into the host cell's DNA and ending with the replication of a new virion.

The process of building new virus particles begins within the host cell's DNA. Segments at the end of the viral genome instruct the host cell to make RNA copies of the viral DNA. Some of the genes direct the host cell to manufacture viral envelope proteins (gene name: *env*) and enzymes (gene name: *pol*), whereas other RNA strands become future genetic material (gene name: *gag*). The HIV DNA then hijacks cellular protein pathways to produce the proteins needed for replication of HIV. This complex interaction between the host cellular kinases and factors affects the HIV replication cycle.

To cause further infection, the viral RNA must be produced, leave the nucleus, and migrate to the cell surface. Rev proteins along with a human protein called *CRM1* aid in the process of transporting the viral RNA proteins from the cell nucleus. A human RNA helicase enzyme, DDX3, helps to straighten HIV's twisted strand of RNA before threading it through a small pore in the nucleus. The viral RNAs are then exported from the cell nucleus to the cytoplasm, where viral proteins are synthesized—particularly gag/pol. Once gag/pol is synthesized in the cytoplasm, it is moved to the inner portion of the plasma membrane.

The assembly of new virus particles, called virions, occurs at the cell membrane. The HIV polyproteins direct viral RNA movement to the membrane, promote the incorporation of the viral envelope glycoproteins, and are involved in the viral particle budding out from the membrane. The protein-cutting enzyme protease separates the envelope proteins from enzymes and RNA genetic material and binds the viral core (Fig. 12.7). Therefore the completed virion has a host cell membrane from which the envelope proteins gp120 and gp41 protrude like spikes.

When $CD4^+$ cells decline, the diversity of $CD4^+$ cells is affected. With antiretroviral therapy, the naive T cells that can respond to new infections persist in low numbers despite an increase in memory T cells. Therefore persons with HIV who are receiving antiretroviral therapy can respond better to previous compared with new infections. This phase is indicative of deterioration in immune system function despite temporary increases in $CD4^+$ cell counts and decreased viral load from antiretroviral therapy.

Once infection occurs, the host's adaptive and innate immune responses to HIV infection try to control disease progression but are incapable of eliminating the virus. These processes involve multiple and complex interactions between the HIV virus and host innate immune

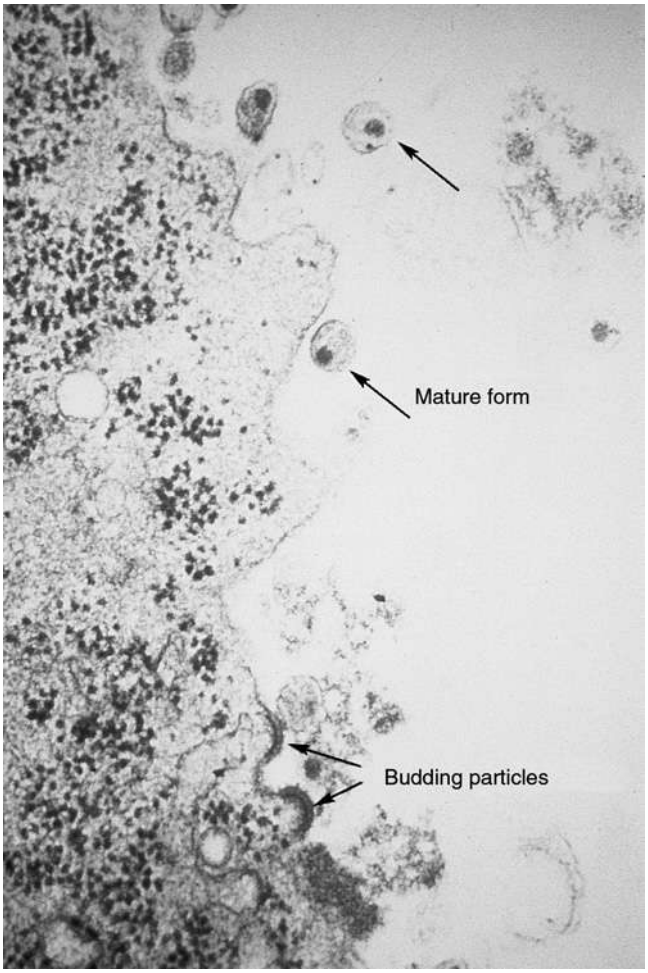


FIG 12.7 HIV-1/lymphadenopathy virus found in a hemophiliac patient with AIDS. Virus particles range in size from 90 to 120 nm. Viral budding and the production of new virions are facilitated by protease. (Courtesy Centers for Disease Control and Prevention, Atlanta, Georgia.)

system factors, leading to persistent inflammation and chronic immune system activation. HIV virus takes advantage of host metabolic pathways and proteins, allowing the virus to thrive and persist in the host.

Host recognition of HIV virus causes antigen-presenting cells and dendritic cells to initiate host activation of $CD4^+$ T cells and NK cells in the lymph nodes. Proinflammatory cytokines are activated and produced along with induction of T helper-1 cells. HIV-1 viral infection, especially during the M tropic phase, has been shown to increase IL-1β levels, which leads to increased inflammation, immune dysregulation, and $CD4^+$ T-cell depletion. TNF-α is also produced, impairing the tight epithelial junctional barrier, which allows migration of HIV and bacteria to move or translocate across mucosal surfaces. All of these events cause a pathogenic cascade in which the death of $CD4^+$ T cells releases inflammatory signals that attract more $CD4^+$ T cells, which subsequently die. This ultimately creates a state of chronic inflammation.

KEY POINTS

- HIV is an RNA virus known as a *retrovirus*. It must undergo reverse transcription within infected cells to form viral DNA.
- HIV consists of a nucleocapsid containing two strands of RNA, protein, and enzymes surrounded by a spherical lipid bilayer viral envelope. At least nine genes comprise the HIV genome.
- The HIV genome contains all the information regulating the virus's structural format and growth, including the conversion of RNA to DNA.
- The enzymes needed to convert HIV RNA to DNA include reverse transcriptase, integrase, and protease.
- HIV gains access to $CD4^+$ cells by attaching to the CD4 receptor on the cell surface. Viral envelope protein gp120 and coreceptor chemokines such as CCR5 and CXCR4 mediate attachment.

PATHOGENESIS

Effect of HIV on Immune Cells at the Cellular Level

The hallmark of HIV infection is the decrease in the number of $CD4^+$ T helper/inducer lymphocytes. T helper/inducer cells are necessary for appropriate immune responsiveness because they are the cells that mediate between the antigen-presenting cells, other immune cells such as B cells, and other T cells. During acute and chronic untreated HIV infection, the immune system is in a hyperactive state with high T-cell death, nonspecific T-cell activation, polyclonal activation of B cells, and elevated levels of proinflammatory cytokines. Normally $CD4^+$ T helper lymphocytes in lymphoid tissues activate and help B cells to develop appropriate humoral immune responses to immune system threats. However, in HIV infection, chronic immune activation causes impairment of the $CD4^+$ T helper cells to adequately activate and help B cells.

Macrophages have $CD4^+$ receptors and act as both targets and reservoirs for HIV. As the infection progresses, they become more functionally impaired with defective phagocytosis and chemotaxis, abnormal antigen presentation, and abnormal cytokine production. They also contribute to the T-cell decline by increasing $CD4^+$ cell death.

Humoral immune system dysfunction is also present, although the effect of HIV on antibody-producing B cells is more poorly understood. B-cell differentiation and response to antigens are decreased, progressively suboptimal, and dysfunctional. The B-cell subset—$CD21^{-/low}$ B cells—has a decreased ability to proliferate in response to HIV viremia, which is referred to as an *exhausted cell population*. Recent findings indicate that normal production of antibodies does not develop early in HIV-1 infection. B-cell production of nonessential antibodies, as well as failure to respond appropriately to normal immune system signals, are affected by high viral load, the duration of infection, and increased viral diversity. Levels of salivary immunoglobulin A (IgA) in HIV-infected children were found to be significantly low before treatment in a recent study. The responsiveness to bacterial cell wall (polysaccharide) antigens that require $CD4^+$ cell activation of B cells is decreased. Immune complexes are increased. Autoantibodies, especially against erythrocytes, platelets, lymphocytes, neutrophils, nuclear proteins, myelin, and spermatozoa, occur either in association with disease processes (e.g., HIV-associated thrombocytopenia) or spontaneously. HIV antibodies are produced, but they are ineffective against the disease.

The envelope glycoproteins (gp120 and gp41) on the surface of HIV virions are the reason for successful HIV infection. HIV *env* (gp120 and gp41) is one of the most heavily glycosylated proteins known. The immunogenic portions of the viral envelope glycoproteins are well disguised and variable because of the protection afforded by the glycosylation. Within the human body, high-carbohydrate substances look like "self" to the immune system. Therefore the virus "hides" under the cover of the glycosylation. Another factor that allows HIV envelope proteins to escape the early antibodies is the way that gp120 and gp41 are bound together. Although the interface between gp120 and gp41 is an area that is highly immunogenic, the gp120 and gp41 molecules are noncovalently bonded together. Early antibodies cannot bind the assembled, functional envelope glycoprotein complex. Later, neutralizing antibodies are effective against the complex, but by that time the infection is well established.

Viral Production and Cell Death

A key element in the success of HIV infection is that HIV replicates prolifically from the onset. It generates so many virions that it overwhelms the body's defenses. Because HIV is primarily a mucosal disease, the gastrointestinal (GI) tract is the major site of HIV replication. Within the first 3 to 6 weeks and continuing throughout the infection, HIV replication is high in the lamina propria of $CD4^+$ T cells of the GI tract. HIV infection is characterized by a high level of virion turnover (HIV replication) and a high level of $CD4^+$ cell turnover (host cell death). HIV-infected $CD4^+$ cells undergo viral budding to generate and produce new virions (Fig. 12.8). At least 10 billion HIV particles are produced and destroyed each day, with a plasma virus half-life of 6 hours and an acutely infected T-cell half-life of 1.1 days. Total T-cell numbers in acute HIV infection decline sharply, but with continuing infection blood T-cell numbers rebound slightly as a result of antiviral immune responses,

FIG 12.8 Scanning electron micrograph of HIV-1–infected $CD4^+$ lymphocytes showing virus budding from the plasma membrane of the lymphocytes. (Courtesy Centers for Disease Control and Prevention, Atlanta, Georgia.)

whereas GI T cell numbers remain low. In children infected with HIV, the virus is more aggressive and leads more rapidly to immune system dysfunction.

The production of new virus is variable between individuals and dependent on the host's cellular activity, as well as the interaction between HIV regulatory genes *(tat, nef, rev, vif)*. In some cells, such as T cells, HIV can lie dormant until activated. In other cells, such as macrophages and monocytes, RNA copies of HIV are consistently being made and released, initially without destruction of the host cell. Other host cellular factors that influence viral production of HIV include inhibition by other proteins and low concentrations of initiation factors.

Other proteins in the body can repress or inhibit HIV-1 if present in the cell. One of these is the delta 32 mutation in the *CCR5* gene (CCR5-delta32), which involves the chemokine coreceptor needed for HIV to infect cells. Other genetic variants that can affect susceptibility to HIV infection and progression include CCR5 promoter A/G, CC-chemokine receptor 2 mutation 64 isoleucine (CCR2-64l) and stromal cell-derived factor 3′A mutation (SDF1-3′A). The prevalence of these mutations varies in different regions and populations.

Other inflammatory proteins from macrophages decrease the probability of becoming infected with HIV-1 and promote recovery of the CD4 cells after starting therapy, including macrophage inflammatory protein-1α (MIP-1α/CCL3) and macrophage inflammatory protein-1β (MIP-1β/CCL4). HLA-B57 and HLA-B27, part of the major histocompatibility complex alleles, have been found to target several gag epitopes, inducing cytolytic destruction of infected cells. NK cells and macrophages secrete soluble factors that can inhibit HIV infection, including TNF-α, interferon, and chemokines CCL3, CCL4, and CCL5. The chemokines tend to compete with the virus for cell adhesion because of their attraction to CCR5. Interferon type 1 also inhibits HIV cell adhesion and induces apoptosis of HIV-1–infected $CD4^+$ T cells. These factors have been associated with lower viral loads and improved survival.

Long-term survivors with HIV often have a lower viral load and strong $CD8^+$ killer T-cell activity. The $CD8^+$ killer T-cell activity suppresses viral replication and thus slows progression of the disease, especially in the early stages. It is also thought that a strong immunologic defense preserves the manufacture of $CD4^+$ T cells that especially recognize and react to HIV. If this ability is lost, these cells may not regenerate, even with treatment. The presence of a weaker strain of HIV, particularly during the M tropic phase, may also lead to longer survival.

Once viral production starts or restarts in activated cells, death of the infected cells may occur by a variety of mechanisms. Cells may die of the accumulation of intracellular viral DNA or from the loss of normal cellular protein synthesis because of the infection. Some cell death may occur due to the action of cytotoxic T cells. Most methods of host cell death involve the envelope protein gp120 or immune processes. Cross-linking of CD4 and gp120 can trigger automatic preprogrammed T-cell death (apoptosis). *Apoptosis is the major mechanism of CD4 T-cell depletion.* B cells also have an increased risk of cell death through apoptosis.

CD4 and gp120 cross-linking can also cause the cell to stop dividing and decreases the cell's ability to fight new infections—a condition called anergy**.** Profuse viral production with multiple CD4 receptors in close proximity can rip holes in the cell membrane and cause host cell death (Fig. 12.9). Multiple virion buds with gp120 on their surfaces attach to the surrounding host cell membrane CD4 receptors. This attachment causes tearing of the host cell membrane with subsequent cellular edema and death.

Viral proteins, such as nef (negative regulatory factor), vpr (viral protein R), tat (transactivator of transcription), *gag,* or protease, also contribute to the causes of cell death in HIV. Exosomes containing nef protein released into the blood are one of the earliest and most abundant

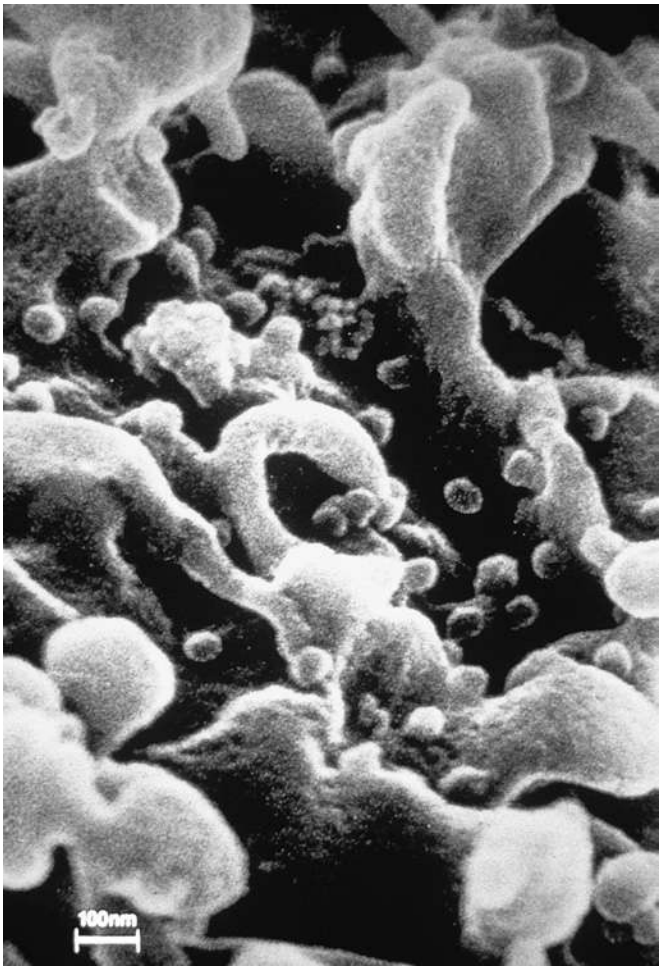


FIG 12.9 High magnification of a $CD4^+$ lymphocyte infected with HIV-1. Note the large number of budding HIV virions, which can lead to host cell death by membrane tearing or syncytium formation. (Courtesy Centers for Disease Control and Prevention, Atlanta, Georgia.)

expressed viral proteins and can cause *nef*-induced apoptosis, facilitating the depletion of $CD4^+$ T cells. *Tat* protein can disrupt mitochondrial function and trigger extrinsic and intrinsic apoptosis pathways. *Vpr* either directly or indirectly causes induction of apoptosis because it causes cell cycle arrest and/or the intrinsic pathway.

Another process of cell death occurs when multiple uninfected cells become fused together with infected cells by the virus. This mass of cells, called a syncytium**,** can lead to a large number of cell deaths from a single event. During the T tropic phase, there is a greater tendency to produce syncytia, causing even faster depletion of T cells. Myeloid-derived dendritic cells aid in the formation of syncytia when they patrol the body, engulfing the virus and presenting the virus to T cells. During this presentation, virus, receptors, and coreceptors are in close proximity, facilitating infection as well as the development of a syncytium.

Cell death can occur when the immune system makes antibodies to the viral envelope protein. When gp120 is shed, it can bind to uninfected CD4 receptors. The immune system then attacks the uninfected but antibody-coated cells via the complement system (antibody-dependent cellular cytotoxicity) or cytotoxic T cells. Cell death can also be secondary to a type III hypersensitivity reaction. As discussed earlier, gp120 and gp41 hide from the immune system because of the large amount of carbohydrate on the surface of gp120 and gp 41. The binding and glycosylation of gp120 and gp41 cause them to

have characteristics similar to those of major histocompatibility class (MHC) antigens. In this case immune system cells may fail to recognize the difference between gp120 and "self MHC markers," causing them to attack normal cells as "nonself." This phenomenon causes immune cells to attack and destroy large numbers of T cells. Cells also may be affected by T-cell–mediated cytotoxicity or by cytokines and inflammation resulting from infection.

Progression of HIV Infection From Seroconversion to AIDS

HIV disease infection progresses over time to AIDS with many manifestations. AIDS is a syndrome, not a disease, which means that the virus can express itself in many ways. No one symptom typifies either HIV infection or AIDS. However, groups of signs and symptoms are useful in staging progress of the infection. HIV is characterized by two major phases: acute and chronic.

Once the HIV virion enters the body, it rapidly replicates (Fig. 12.10). This is the acute stage of infection. HIV is present in the blood and cerebrospinal fluid but is not detected by usual laboratory tests because no antibodies have formed yet. Usually no symptoms are present. It is a time of rapid virus replication with widespread attack of $CD4^+$ T cells in mucosal membranes, especially in the GI system. The person is infectious but does not know it. This stage lasts approximately 3 weeks.

Seroconversion occurs when sufficient antibodies are detected in the blood, usually between 3 weeks and 6 months after exposure (range: 3 weeks to 14 months). At the time of seroconversion, the person experiences signs and symptoms of acute retroviral syndrome or primary HIV infection. Up to 80% of people newly infected with HIV have flulike symptoms that can mimic other viral illnesses. At this time, the symptoms of primary HIV infection include flulike or mononucleosis-like symptoms, such as fever, chills, headaches, nausea, vomiting, fatigue, weakness, arthralgias, sore throat, stiff neck, photophobia, irritability, and rash. The rash is not the same in everyone and may be maculopapular, vesicular, or urticarial. Encephalopathy may even develop. The $CD4^+$ T-cell count is greater than 400 cells/μL. The number of white blood cells, including lymphocytes, is decreased except for an increased number of $CD8^+$ T cells. The number of platelets is also decreased. The person has an elevated erythrocyte sedimentation rate. During this period, the HIV count in the blood and genital fluids is high and the individual is very infectious. Then, after 1 to 4 weeks, the symptoms disappear. However, HIV is still present, and the person continues to be infectious throughout the rest of the course of the infection.

After this period of seroconversion, the patient experiences the clinical latency period, which may last for longer than 10 years (range: 3 to 12 years). During this period, seeded HIV replicates in the lymph nodes and gradually destroys lymph tissue over time. Antiviral immune activity is ongoing. Production of virus is maintained or stabilized at a set level. The person feels well but may experience chronic lymphadenopathy (enlargement of lymph nodes for more than 3 months) or mild general symptoms, including lack of energy, weight loss, frequent fevers, and sweats. The $CD4^+$ T-cell count is greater than 400 cells/μL. Stabilization of the serum level of virus at a certain point is attributable to the antiviral response, the number of $CD4^+$ cells, and the virulence of the HIV strain. During this period of asymptomatic or mild infection, large numbers of virions are produced, destroying the body's immune system. Up to 2 million viral particles can be produced daily. The lymphadenopathy is caused by a vigorous immune response against HIV infection. The key point is that although the infection is clinically asymptomatic or mildly symptomatic, the virus is active, not latent.

Immediately after the latency period of infection, a period of rapid virus production occurs for up to 18 months. It is hypothesized that the destroyed lymph nodes are no longer capable of removing or holding

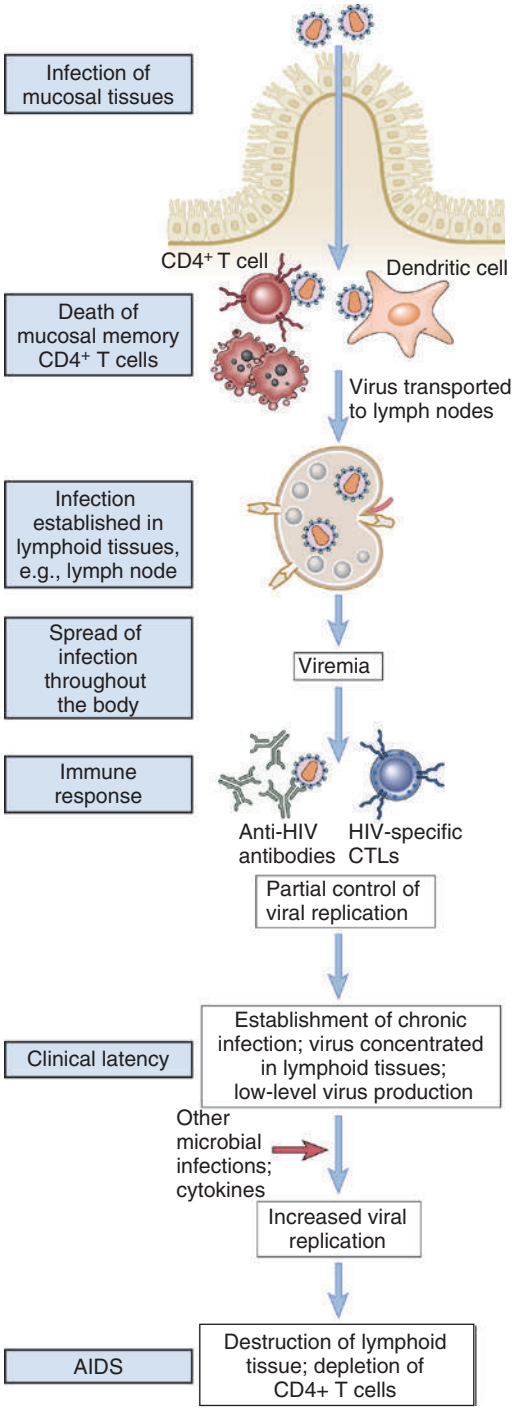


FIG 12.10 Progression of HIV infection. The clinical stages of HIV disease correlate with a progressive spread of HIV from the initial site of infection to lymphoid tissues throughout the body. The immune response of the host temporarily controls acute infection, but does not prevent the establishment of chronic infection of cells in the lymphoid tissues. Cytokine stimuli induced by other microbes serve to enhance HIV production and progression to AIDS. *CTLs,* Cytotoxic T lymphocytes. (From Abbas AK et al: *Cellular and molecular immunology,* ed 8, Philadelphia, 2015, Saunders, p 465.)

TABLE 12.3 CDC Revised Surveillance Case Definition for HIV Infection – United States, 2014

	AGE ON DATE OF CD4 T-LYMPHOCYTE TEST					
	<1 Year		1–5 Years		6 Years Through Adult	
Stage*	Cells/μL	%	Cells/μL	%	Cells/μL	%
1	≥1,500	≥34	≥1,000	≥30	≥500	≥26
2	750–1,499	26–33	500–999	22–29	200–499	14–25
3	<750	<26	<500	<22	<200	<14

*The stage is based primarily on the $CD4^+$ T-lymphocyte count; the $CD4^+$ T-lymphocyte count takes precedence over the CD4 T-lymphocyte percentage, and the percentage is considered only if the count is missing.

From Selik RM et al: Revised Surveillance Case Definition for HIV Infection—United States, 2014. *Recommendations and Reports MMWR,* April 11, 2014, 63(RR03);1-10. https://www.cdc.gov/mmwr/preview/mmwrhtml/rr6303a1.htm.

virus, thus allowing viral escape into the bloodstream (viremia). During this time, a persistent and continuous drop in the $CD4^+$ T-cell count to less than 400/μL is taking place. The antiviral innate immune activity is less effective as the viral load (level of virus in the blood) increases. Also during this time, HIV can persist for years in pools of resting memory $CD4^+$ T cells. This pool of $CD4^+$ T cells carries only provirus DNA and lacks HIV surface antigens, so they are not detected or destroyed by the immune system.

As the viral loads increase and the immune system declines, the patient enters the stage of symptomatic, chronic HIV infection. At this time, the patient progresses from partially responding to skin testing (partial anergy) to complete anergy with no response to skin testing. Severe viral or fungal infections of the skin and mucous membranes develop. Oral and genital herpes simplex infections, including shingles or candidiasis (yeast) infections, usually develop, as well as oral hairy leukoplakia. The patient also may have persistent skin rashes or flaky skin, short-term memory loss, or pelvic inflammatory disease that does not respond to treatment. Children tend to have growth delays and frequent illnesses. The person may develop cytomegalovirus (CMV) infection, Epstein–Barr virus infection, or both, as well as other opportunistic infections.

An HIV-infected person is not diagnosed with AIDS until the $CD4^+$ T-cell count is less than 200/μL. At that time, the person typically has one or more opportunistic infections, including PCP, *Toxoplasma gondii*–associated neural toxoplasmosis, cryptosporidiosis (gastroenteritis), and *Mycobacterium tuberculosis.* The person usually has one or more tumors or cancers, including Kaposi sarcoma (a connective tissue skin cancer), lymphomas, or cancer of the rectum or tongue. General symptoms of opportunistic infections include coughing or shortness of breath; difficult or painful swallowing (dysphagia); cognitive symptoms such as confusion, forgetfulness, or lack of coordination; seizures; fever; vision loss; severe headaches; and GI symptoms such as abdominal cramps, nausea and vomiting, severe and persistent diarrhea, and weight loss, causing extreme fatigue.

Disease progression in infants and children is determined by the timing of the child's infection, the viral load, the child's immune response, and the viral virulence. In general, children with HIV progress more rapidly than adults. Most children fall into two distinct groups: rapid disease progression or slower disease progression. In those with rapidly progressing disease, symptoms develop within the first 6 months of life, sustained decreases in $CD4^+$ T-cell counts are noted, and AIDS develops within the first 2 years of life. Early aggressive treatment in perinatally infected infants may slow disease progression and prolong immune function. Children with AIDS generally manifest the same opportunistic infections as adults and may have severe forms of common childhood diseases, such as conjunctivitis, otitis media, and tonsillitis.

CDC HIV Classification System

The CDC HIV classification system is a simple matrix classification system for adults and children and is the preferred method of staging. In April 2014, the CDC revised the surveillance case definition for HIV infection in the United States (Table 12.3). This case definition is designed for monitoring HIV infection and population epidemiology. The CDC has defined five stages of HIV infection based on age at the time of the $CD4^+$ T-cell testing. Stage 0 represents acute, early HIV infection based on a negative HIV test within 6 months of the first positive HIV diagnosis. This is recognition of the presence of infection before an antibody response due to more sensitive testing. In children 6 years through adult, Stage 1 occurs when the $CD4^+$ T-cell count is greater than or equal to 500/μL; in Stage 2, the $CD4^+$ T-cell counts range from 200 to 499/μL; and in Stage 3, the $CD4^+$ T-cell count is less than 200/μL. If a Stage 3–defining opportunistic illness has been diagnosed at any time, the stage is defined as Stage 3. A list of AIDS indicator conditions that are listed in the AIDS surveillance case definition is shown in Box 12.1. The CDC considers all HIV infections in the United States to be type 1 (HIV-1) unless laboratory test results are sufficient to classify the HIV infection as type 2 (HIV-2).

DIAGNOSTIC TESTING

The CDC and U.S. Preventive Task Force have recommended HIV testing as part of routine care for all adolescents and adults. Testing is also recommended for all pregnant women, as well as for those at high risk for HIV infection. The Food and Drug Administration (FDA) has approved various methods to test for HIV infection. Rapid HIV assays that use blood, serum, or oral secretions can be used. These antibody test results can be obtained in 20 to 30 minutes. However, positive results must be confirmed by further laboratory testing. It is also important to remember that false-negative tests can occur during the initial period of HIV infection before seroconversion, although new fourth-generation assay tests are able to detect seroconversion more rapidly.

To diagnose HIV infection, serum or plasma specimens are tested in approved laboratories using an HIV-1/2 antigen/antibody combination immunoassay. If the first test is positive, an antibody differentiation immunoassay is performed. Usually the enzyme-linked immunosorbent assay (ELISA) and the Western blot are used to detect the presence of HIV antibodies.

The ELISA test result is positive for HIV antibodies if the blood or oral mucosal transudate of an infected person reacts with the surface antigen of a killed HIV virus. The ELISA test uses purified viral proteins placed on plastic beads or in multiwell trays. When the test serum or

BOX 12.1 Conditions Indicative of AIDS in HIV-Infected Persons

- Candidiasis of bronchi, trachea, or lungs
- Candidiasis, esophageal
- Cervical cancer, invasive*
- Coccidioidomycosis, disseminated or extrapulmonary
- Cryptococcosis, extrapulmonary
- Cryptosporidiosis, chronic intestinal (greater than 1 month's duration)
- Cytomegalovirus disease (other than liver, spleen, or nodes)
- Cytomegalovirus retinitis (with loss of vision)
- Encephalopathy, HIV-related
- Herpes simplex: chronic ulcer(s) (greater than 1 month's duration); or bronchitis, pneumonitis, or esophagitis
- Histoplasmosis, disseminated or extrapulmonary
- Isosporiasis, chronic intestinal (greater than 1 month's duration)
- Kaposi sarcoma
- Lymphoma, Burkitt (or equivalent term)
- Lymphoma, immunoblastic (or equivalent term)
- Lymphoma, primary, of brain
- *Mycobacterium avium complex* or *M. kansasii,* disseminated or extrapulmonary
- *Mycobacterium tuberculosis,* any site (pulmonary* or extrapulmonary)
- *Mycobacterium,* other species or unidentified species, disseminated or extrapulmonary
- *Pneumocystis carinii (jirovecii)* pneumonia
- Pneumonia, recurrent*
- Progressive multifocal leukoencephalopathy
- Salmonella septicemia, recurrent
- Toxoplasmosis of brain
- Wasting syndrome attributable to HIV

From Appendix B of the 1993 Revised Classification System for HIV Infection and Expanded Surveillance Case Definition for AIDS Among Adolescents and Adults, *MMWR* 41(RR-17).
*Added in the 1993 expansion of the AIDS surveillance case definition.

oral mucosal transudate from a patient is contacted by the purified viral proteins, an antigen–antibody reaction occurs. Anti–human antibody added to the reaction can be detected colorimetrically and indicates whether any antigen–antibody compounds have formed. This test is highly sensitive (more than 99%) and specific (more than 99%) in high-risk populations. For the test to be specific, however, it must be performed with both HIV-1 and HIV-2 viral antigens.

When the ELISA test result is positive, a second test, the Western blot, is used to confirm the presence of HIV antibodies. The Western blot test uses an expensive process called *electrophoresis,* so usually it is used only as a confirmatory test. This test identifies specific antibodies against the HIV protein antigens. The specificity of this test in combination with the ELISA is greater than 99.9%. The problem with this additional testing is that the patient must wait up to 1 to 2 weeks for confirmation.

Testing neonates for HIV is difficult because of maternal transmission of IgG antibodies against HIV. These passive maternal antibodies cross the placenta and can last as long as 15 months. Therefore the best method to determine whether a neonate has HIV is to culture the virus from blood and peripheral tissue.

MONITORING THE PROGRESSION OF HIV

After initial diagnosis of HIV, other laboratory tests need to be performed to stage the disease and assist in the selection and monitoring of appropriate drug treatments. One of the most important laboratory tests is the $CD4^+$ count, which is the specific indicator of disease progression of HIV to AIDS. As the $CD4^+$ T-cell count declines, the risk of progression to AIDS and development of opportunistic infections and malignancies increases. Highly virulent communicable diseases can still occur when the $CD4^+$ count is high. However, when the $CD4^+$ T-cell count drops below 200 cells/μL, the number and severity of low-virulence diseases and opportunistic infections increase. It is at this level that many patients begin taking prophylactic medications to prevent opportunistic and other infections. Also used is the $CD4^+$ lymphocyte percentage, which is more stable and has less variation over time. When the $CD4^+$ lymphocyte percentage is less than 14%, the risk of developing AIDS-defining illnesses is higher.

Another useful test is the plasma HIV nucleic acid test, which either looks for viral RNA directly or viral load. This test indicates the amount of viral replication and helps predict disease progression. The level of HIV RNA in plasma is the strongest predictor of outcome over time. When the plasma HIV RNA content is low, the risk of disease progression declines. The plasma viral load helps the clinician assess the effectiveness of various therapies and is the basis for initiating more aggressive therapies to decrease the viral load. Usually, HIV RNA levels should drop after the onset of therapy and by 6 months should be undetectable in the plasma. The viral load assay counts copies of HIV RNA in 1 mL of plasma and is either a reverse transcriptase polymerase chain reaction or a branched DNA assay. Because each virion contains two strands of RNA, the actual virion level is half the HIV RNA counted. Tests are currently sensitive to 50 copies per milliliter.

Genotypic resistance testing is now part of routine management of HIV infection and is usually part of the initial evaluation of HIV. Genotypic testing identifies viral mutations, whereas phenotypic testing identifies the concentration of antiretroviral drug needed to inhibit viral replication in culture medium. Ideally this testing can help determine the best drugs to be given to the patient. Genotypic resistance tests specifically examine protease and reverse transcriptase and are now used even when the viral loads are less than 1000 copies/mL. Testing for HLA-B*-5701 is a baseline test performed before starting the medication abacavir, which can cause fatal hypersensitivity reaction in those who are HLA-B*-5701 positive. These tests are expensive, which can limit their use in some populations.

Another common test is an anergy test or a delayed hypersensitivity (type IV) test for such organisms as *M. tuberculosis* or mumps or measles virus. In early HIV infection, these skin test results are normal. However, in advanced cases, the patient will have no response to testing because of the loss of macrophage and $CD4^+$ T-cell functioning.

Other tests include β_2-microglobulin and p24 antigen. β_2-Microglobulin is a cell-surface protein that indicates macrophage stimulation. Levels greater than 3.5 mg/L are associated with rapid progression of the disease. P24 antigen is indicative of active HIV replication and confirms the diagnosis of HIV infection. It is positive before seroconversion and may be used before confirmation of the disease by ELISA. It is also elevated in later stages of the disease, a period when antibody testing may be unreliable. It is usually part of the antigen–antibody tests for HIV.

Other laboratory and diagnostic tests can assist in the identification, monitoring, and treatment of persons infected with HIV, including a complete blood cell count (CBC), a chemistry panel or screen, and chest radiograph. These tests are routinely used to detect infections and changes in a patient's physiologic status (Box 12.2). The CBC detects the development of anemia (as a result of infection, chronic illness, or secondary to therapy), neutropenia, and thrombocytopenia, which may occur in advanced stages of the disease. In patients with chronic diarrhea, stool testing is usually indicated, including fecal lactoferrin, fecal

BOX 12.2 Tests Used to Evaluate Diagnosis and Progression of HIV Infection

HIV Antibody
Confirm diagnosis at entry to care

Complete Blood Cell Count With Differential and Platelets
At entry into care and every 3 to 6 months
White blood cell count normal to decreased
Lymphopenia (<30% of normal number of WBCs)
Thrombocytopenia (decreased platelet count)

$CD4^+$ Count/Percentage
At entry into care and every 3 to 6 months
Reduced $CD4^+/CD8^+$ T-cell ratio
$CD4^+$ (helper) lymphocytes decreased
$CD8^+$ lymphocytes increased

Quantitative Plasma HIV RNA (HIV Viral Load)
Perform at entry into care, at initiation of HAART, and every 4 to 8 weeks until viral load suppressed, then every 3 to 6 months
Used to monitor effect of HAART

Resistance Testing
At entry into care and before initiation of or any modification in HAART
Determines viral resistance to HAART

Quantitative Immunoglobulin
As needed depending on immune function
IgG increased
IgA frequently increased

Chemistry Panel
At entry into care and every 3 to 6 months
Lactate dehydrogenase increased (all fractions)
Serum albumin decreased
Total protein increased
AST and ALT elevated
Total bilirubin elevated
Serum glucose elevated
GFR decreased

Lipid Panel
At entry into care and, if abnormal, every 6 to 12 months
Total cholesterol increased
Triglycerides increased
LDL cholesterol increased

Anergy Panel
As needed
Nonreactive (anergic) or poorly reactive to infectious agents or environmental materials (e.g., pokeweed, phytohemagglutinin mitogens and antigens, mumps, *Candida*)

Urinalysis With Urine Protein and Creatinine
At entry into care and every 6 to 12 months or as needed
To detect urinary tract infections, hematuria or dysfunction
Albuminuria or proteinuria elevated

Hepatitis A, B, and C Serology
At entry into care and as needed
To detect the presence of hepatitis antigens/antibodies

Blood Cultures
As needed
To detect septicemia

Tuberculosis (TB) Screening
At entry into care, every 12 months and as needed

Chest Radiograph
At entry into care and as needed
To detect latent or active diseases such as *Pneumocystis jiroveci (carinii)* infection or tuberculosis

HLA-B5701 Testing
If considering use of abacavir

Tropism Testing
If considering use of CCR5 antagonist

Papanicolaou Test
At entry into care, annually or as indicated by condition
To detect dysplasia or abnormal cell changes

Pregnancy Testing
If female at entry into care and as needed
To detect early pregnancy

Serum VDRL
At entry into care and as needed depending on risk factors

Gonorrhea, Chlamydia and Trichomoniasis Testing
At entry into care, every 12 months or depending on risk factors

Renal or Abdominal Ultrasound
To detect hepatosplenomegaly
To detect renal disease

Dilated Retinal Examination
At entry into care, every 12 months and as needed
To detect CMV, HIV retinopathy or other changes

Cytomegalovirus (CMV) Antibody
At entry into care and as needed
To be aware of disease risk in advanced disease or to avoid exposure

ALT, Alanine aminotransferase; *AST,* aspartate aminotransferase; *HAART,* highly active antiretroviral therapy.
Data from U.S. Department of Health and Human Services, Health Resources and Services Administration: Guide for HIV/AIDS clinical care – 2014 edition. Rockville, MD, 2014.

leukocytes, fecal calprotectin, ova and parasites, bacterial culture and sensitivity, and fecal occult blood. In addition to the aforementioned tests, it is important to measure serologic values for hepatitis A, B, and C viruses, as well as to obtain screening tests for sexually transmitted infections in light of the risk of coinfection. To screen for kidney dysfunction, urinalysis, glomerular filtration rate (GFR), and quantitation of albuminuria also need to be performed when diagnosed and periodically thereafter.

KEY POINTS

- HIV virions are attracted to cells with CD4 receptors such as T cells; microglial cells; monocyte-macrophages; follicular dendritic cells; immortalized B cells; retinal cells; Langerhans cells in the skin; bone marrow stem cells; cervical cells; bone marrow–derived circulating dendritic cells; and enterochromaffin cells in the colon, duodenum, and rectum.
- The hallmark of AIDS is a decrease in the number of $CD4^+$ cells, including T helper lymphocytes and macrophages, especially in mucosal membranes. B-cell responsiveness is decreased because of dependence on T helper cell cytokines.
- The key element in HIV infection is the high level of virion production and the high level of $CD4^+$ cell death.
- $CD4^+$ cell death occurs via several mechanisms. Cross-linking of CD4 receptors by viruses may result in T-cell death, apoptosis, or anergy. Virions may cause the linkage of infected and uninfected cells, followed by cell fusion and death. B cells may form antibodies against infected T cells. Viral budding may cause excessive loss of cell membranes.
- Laboratory testing for HIV is accomplished by using either the ELISA or the Western blot test. Usually the ELISA test is performed first. If it is positive, the Western blot test is performed to confirm the presence of specific antibodies against HIV protein antigens.
- HIV is an infectious disease that progresses to AIDS and is characterized by different clinical manifestations at each stage. Individuals move through the stages at different rates.
- Flulike symptoms and the formation of anti-HIV antibodies (seroconversion) characterize the early stage of viral seeding. Next, symptoms of early immune dysfunction are present, including lymphadenopathy, fever, and night sweats. A surge in viral production and a drop in the $CD4^+$ lymphocyte count follow this stage.
- In the later stages, $CD4^+$ counts continue to fall and the person is subject to a number of opportunistic infections and tumor formation. An HIV-positive individual is diagnosed with AIDS when the $CD4^+$ T-cell count is less than 200/μL or when a category C AIDS indicator condition is present.
- Children often have rapidly progressive disease, with onset of AIDS between ages 4 and 8 years.

BOX 12.3 Common Agents of Infection in Patients With AIDS

Viruses

Herpes Simplex Virus 1 and 2 (HSV-1 and HSV-2)
Human Herpes Virus III (Varicella zoster or herpes zoster)
Human Herpes 8 (HHV-8 or Kaposi sarcoma–associated herpesvirus [KSHV])
Cytomegalovirus
Polyomavirus – JC virus or John Cunningham virus
Epstein–Barr virus
Human papillomavirus (HPV)
Hepatitis A (HAV)
Hepatitis B (HBV)
Hepatitis C (HCV)
Adenovirus
Morbillivirus (Measles RNA virus)

Bacteria

Campylobacter spp.
Escherichia coli spp.
Shigella spp.
Neisseria spp.
Salmonella spp.
Chlamydia spp.
Staphylococcus spp.
Haemophilus influenzae spp.
Legionella spp.
Treponema pallidum spp.
Mycobacterium avium (MAC) or *kansasii* spp.
Bordetella pertussis spp.
Clostridium botulinum spp.

Fungi

Candida albicans
Cryptococcus neoformans
Histoplasma capsulatum
Coccidioides immitis
Nocardia

Protozoa

Pneumocystis jiroveci (carinii)
Toxoplasma gondii
Isospora belli
Cryptosporidium
Giardia lamblia
Entamoeba histolytica

CLINICAL MANIFESTATIONS

HIV affects all body systems, particularly the integumentary, pulmonary, GI, neurologic, and ocular systems. GI manifestations develop in nearly all persons with HIV because of the major effect of HIV infection on the GI system. Pulmonary and cutaneous symptoms develop in approximately 50% to 75% of all persons with HIV, and neurologic symptoms develop in 50% to 60%. Box 12.3 outlines the common agents of infection in patients with AIDS.

Systemic Manifestations

The course of HIV infection parallels the functioning of the immune system. As immune function declines, the number of opportunistic infections and malignancies increases and the normal functioning of organ systems declines.

One of the most significant systemic symptoms is malnutrition, or wasting syndrome. In Africa, HIV is known as "slim disease" because of the wasting. Malnutrition is defined as unintended, involuntary loss of greater than 10% body weight. The systemic symptoms attributable to HIV infection–related malnutrition include major muscle wasting; weight loss; and loss of protein, vitamins, minerals, and other nutrients. HIV malnutrition is multifactoral, including an elevated metabolic rate with increased resting energy expenditure (REE), chronic inflammation, malabsorption, anorexia, decreased intake of food, and the effect of multiple opportunistic insults. There is a 10% to 30% increase in resting metabolic rate attributable to secondary infections or elevated plasma viral load.

In addition, levels of TNF, IL-1, IL-6, and other proinflammatory cytokines are elevated in HIV infection, causing anorexia as well as

increased metabolism of fat and depletion of fat stores. There are deficiencies of vitamins and nutrients, particularly fat soluble vitamins, zinc levels, and selenium. Vitamin B_{12} and vitamin A are decreased and are associated with HIV disease progression. Malabsorption is affected by the loss of GI-associated lymphoid tissue (especially during the initial phase of HIV infection), impairing the integrity of the epithelial mucosal barrier and predisposing to secondary infections of the GI tract. Low body mass index (BMI), failure to regain weight after weight loss, and continued weight loss after the start of medication hasten disease progression and are risk factors for increased mortality. In children, growth failure is indicative of poor outcomes and mortality.

Malnutrition is a leading cause of death among AIDS patients worldwide. Prevention is key, involving assessment of nutritional parameters as well as nutritional education/counseling and exercise. A patient's BMI; weight; waist-to-hip ratio; midarm circumference; calorie and protein intake; and prealbumin, serum albumin, and triglyceride levels are frequently measured.

To prevent or delay the wasting process, some medications have been used, including anabolic steroids, growth hormone treatments, cytokine inhibitors, and appetite stimulants. Ketotifen is a TNF inhibitor and antihistamine that is used because its side effects are appetite stimulation and weight gain. Thalidomide, which is also a TNF inhibitor, appears to be effective against wasting syndrome and increases fat-free mass. Oxandrolone is an anabolic steroid designed specifically to promote weight gain, particularly lean body mass. Megestrol acetate (Megace), a progestational agent, and dronabinol, an antiemetic, are often used to decrease nausea and increase appetite. Use of human growth hormone (somatropin, Serostim) may also be given to increase lean body mass.

Vitamins A, C, D, and E; the B vitamins; zinc; selenium; sulfur-containing amino acids; and other antioxidants are also prescribed. The use of these nutrients can prevent the up-regulation of inflammatory cytokines and thereby decrease inflammation. High-protein, high-energy meals and snacks are recommended, along with nutritional supplements to meet the required amounts of energy, protein, and micronutrients needed as a result of increased metabolic rates (REE). High-fat foods should be avoided because they increase diarrhea, as can lactose-containing foods. Nutritional supplements (Ensure, Nitrofuel, Sustacal Plus, Advera, Lipisorb), which provide both protein and calories, are full of nutrients and can be formulated either with or without lactose and with or without medium-chain triglyceride oil (a more easily digested fat). Implementation of total parenteral nutrition, including gastrostomy or jejunostomy tube feedings, is reserved for those with severe malnutrition and GI manifestations

Gastrointestinal Manifestations

GI manifestations are nearly universal in persons with HIV. In fact, the GI tract is the major target organ in HIV infection. It is targeted in all phases of HIV infection, especially in the acute phase. The gut-associated lymphoid tissue is regularly exposed to antigens from microbial and dietary sources, and HIV is a significant direct pathogen in the GI tract. A wide variety of GI symptoms can occur, including chronic diarrhea (due to HIV enteropathy, chronic infectious diarrhea, medication-related diarrhea, or irritable bowel syndrome), oral candidiasis, dysphagia, mucous membrane ulcerations, (as a primary manifestation or secondary to inflammation), abdominal pain, anorectal disease, GI bleeding, GI Kaposi sarcoma, and GI non-Hodgkin lymphoma.

The most common HIV GI complication is chronic diarrhea, which increases with decreasing $CD4^+$ counts. The diarrhea, often watery or bloody, causes malabsorption and consequently severe weight loss. This complication of HIV-related malnutrition causes muscle loss leading to increased morbidity and risk of death. Antiemetics and antidiarrheals are often useful in controlling symptoms. In patients who are profoundly immunocompromised, intravenous therapy with an electrolyte replacement may be needed.

Whenever chronic diarrhea or other GI symptoms develop in a patient with HIV, it is important to determine the cause: infectious or noninfectious. Infectious causes of GI symptoms can be the result of multiple opportunistic infectious agents. Most patients are less likely to have infectious causes of chronic diarrhea due to improved therapy. For those patients who have infectious diarrhea, some of the most significant infectious agents are viruses such as CMV and herpes simplex; fungi such as *Candida;* bacteria such as *Salmonella, Shigella, Clostridium difficile, Chlamydia trachomatis, E coli,* and *Campylobacter;* and parasites such as *Giardia, Isospora, Entamoeba histolytica,* and *Cryptosporidium.* Treatment involves the use of disease-appropriate antimicrobials, depending on the offending organism.

A common cause of diarrhea is the protozoa *Cryptosporidium parvum,* which infects the intestinal epithelial lining. This organism is transmitted via water, food, animals, and other humans. The onset is generally acute and associated with explosive diarrhea within 4 to 14 days after infection. In nonimmunocompromised persons, the symptoms last up to 2 weeks, but in immunocompromised persons, the diarrhea and symptoms can persist indefinitely. Cryptosporidiosis causes nausea; vomiting; severe, watery, nonbloody diarrhea (more than 15 to 20 L); abdominal pain; cramping; electrolyte disturbances; and dehydration.

Diagnosis is made by stool examination for ova and parasites and by bacterial culture and sensitivity. Antibiotic treatment is not always effective, but antimicrobials such as the macrolides have been used. In the acute phase, some patients must be given intravenous hydration for support. Thereafter, increased oral intake along with low-residue, high-protein, high-calorie diets and loperamide (Imodium) 2 mg tablets, up to 16 to 18 tablets or 36 mg per day, are used to help control the diarrhea.

Prevention of *Cryptosporidium* infection is most important. Preventive activities include routine testing of well water, using water filters at home, avoiding ice or unfiltered tap water both at home and in restaurants, and avoiding fresh fruits or vegetables rinsed with unfiltered water. Fruits and vegetables can be washed in bottled water, filtered water, or water with 20 drops of 2% iodine per gallon.

Oropharyngeal and esophageal *Candida albicans* infections occur in most patients with HIV during the course of their disease. Most often, oral *candidiasis* is pseudomembranous in type, with white plaques that bleed when removed and leave an erythematous surface. *Candida* oropharyngeal lesions produce pain and discomfort during eating, loss of taste, and xerostomia (dry mouth). *Candida* esophageal lesions cause pain with swallowing, dysphagia, and a feeling of "throat swelling." These lesions lead to worsening wasting syndrome. Management of oropharyngeal candidiasis includes the use of topical antifungal agents (such as clotrimazole or nystatin) in a suspension or lozenges, whereas management of esophageal candidiasis includes antifungal agents by oral or intravenous routes of administration. Side effects include an unpleasant taste and GI side effects, with inconvenient dosing regimens (up to six times per day).

Pulmonary Manifestations

Pulmonary manifestations are a major source of morbidity and mortality in AIDS patients. Pulmonary diseases include opportunistic pneumonias, such as those associated with PCP, CMV, *M. tuberculosis, Histoplasma,* or *Staphylococcus,* as well as parenchymal lung diseases, including Kaposi sarcoma, lymphoma, nonspecific pneumonitis, and adult respiratory distress syndrome. Infection with *M. tuberculosis* occurs in 4% of patients with HIV and is particularly problematic in third-world countries where the tuberculosis comorbidity rates are up to 50%. Patients with HIV

have an increased risk of being infected with a multidrug-resistant type of tuberculosis organism.

PCP is a common initial opportunistic infection in HIV and is an AIDS-defining diagnosis in the United States and Europe. However, the incidence of *P. carinii* has decreased with the use of prophylaxis. *P. carinii* is a fungus (renamed *P. jiroveci*). This organism prefers alveolar environments and infects most people during early childhood. Children usually have *Pneumocystis* antibodies by 2 to 3 years of age, but the organism does not cause disease in immunocompetent persons. With immunodeficiency and $CD4^+$ T-cell counts below 200/μL, *Pneumocystis* becomes activated. The nonspecific symptoms of PCP include flulike fever, fatigue, and weight loss. The major pulmonary feature of PCP is severe hypoxemia with a PaO_2 less than 60 mm Hg. The most severe pulmonary symptoms are similar to those of adult respiratory distress syndrome (Fig. 12.11). These symptoms include decreased phospholipid (surfactant) production, early dry cough, dyspnea, tachypnea, chest discomfort, and marked pallor and cyanosis.

Diagnosis of PCP is by chest radiography, and organisms are identified in sputum with Wright–Giemsa stain. *Pneumocystis* cannot be cultured. Sputum is induced by using 3% saline via nebulizer, and patients must avoid brushing their teeth, using mouthwash, or eating for 8 hours before expectoration of a sputum sample. However, the gold standard for diagnosis of PCP is bronchoalveolar lavage with microscopy to identify organisms. Other tests that can be performed include bronchoalveolar biopsy, gallium scanning, and obtaining serum (1>3)-*B*-d-glucan (BG) testing. BG testing is being used more frequently in the diagnostic workup of HIV-positive patients with dyspnea. It is a serum test that identifies a polysaccharide present within the cell wall of *Pneumocystis.* Treatment usually includes the use of intravenous or oral trimethoprim-sulfamethoxazole (Bactrim, Septra) and parenteral and aerosolized pentamidine (NebuPent, Pentam 300).

To prevent PCP infections by prophylaxis, patients with $CD4^+$ counts of less than 200 cells/μL are given trimethoprim-sulfamethoxazole tablets either daily or three times per week to prevent the recurrence of infection. This is a highly effective suppressive therapy that prevents life-threatening PCP. However, there is growing concern about trimethoprim-sulfamethoxazole drug resistance. If patients cannot tolerate it, substitutes include aerosolized pentamidine, dapsone, or dapsone with pyrimethamine and leucovorin.

Mucocutaneous Manifestations

Mucocutaneous manifestations occur both early and late in the course of HIV infection. The early viral exanthem of HIV infection, associated with seroconversion, is an erythematous, fine maculopapular rash found on the face, trunk, and arms. It is a self-limited manifestation that occurs in 40% to 60% of all HIV-infected persons. It is generally seen within 2 to 6 weeks of exposure and lasts up to 1 to 2 weeks. It is associated with mild pruritus, fever, malaise, night sweats, fatigue, pharyngitis, weight loss, diarrhea, headache, and lymphadenopathy.

Other mucocutaneous manifestations may be allergic, infectious, or neoplastic in origin. Cutaneous symptoms depend on the cause and location. Allergic causes may be the result of drug reactions or the development of seborrheic dermatitis, psoriasis, or skin-colored papular eruptions.

Viral causes include herpes simplex virus, varicella zoster virus (Fig. 12.12), Epstein–Barr virus, and human papillomavirus. The development of genital warts (condylomata acuminata) from human papillomavirus is an early symptom of HIV disease in women. For mucocutaneous viral infections, acyclovir (Zovirax) or vidarabine is the recommended antiviral agent.

Oral hairy leukoplakia is an example of an oral mucous membrane infection first described in 1984. It occurs only in immunosuppressed individuals and is associated with Epstein–Barr virus. It occurs in up to 50% of patients with untreated HIV, especially when the $CD4^+$ count declines. Oral hairy leukoplakia is characterized by white to gray, thickened, raised lesions with vertical folds, corrugations, or "hairs" that form on the tongue and buccal mucosa. Usually they form on the

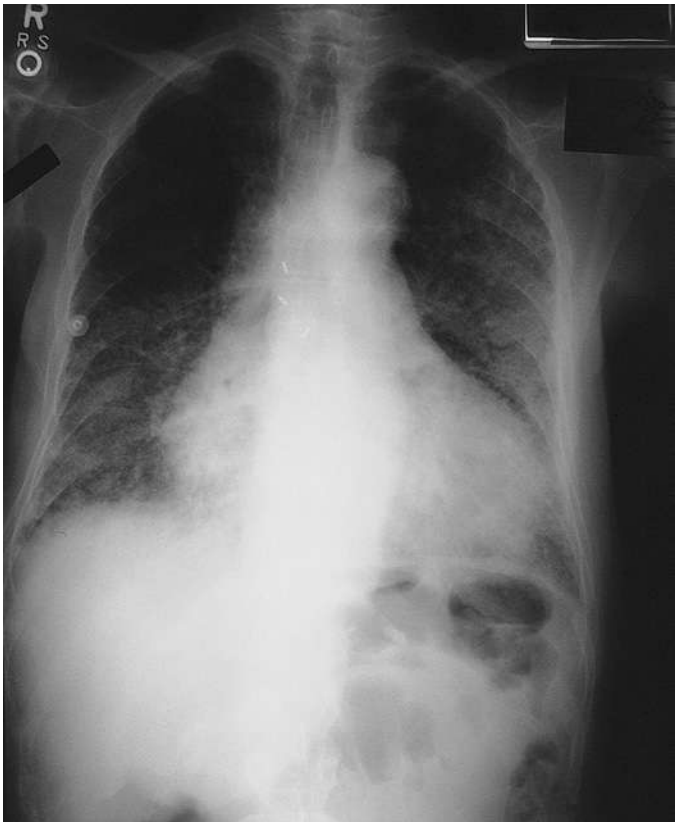


FIG 12.11 *Pneumocystis jiroveci (carinii).* A chest radiograph shows bilateral lower lobe interstitial infiltrates. (Courtesy Dr. Paula Karvalho, Veterans Administration Medical Center, Boise, Idaho.)

FIG 12.12 Herpes zoster in an HIV-infected individual. (From Gawkrodger DJ. *Dermatology: an illustrated color text,* ed 6, 2017, Elsevier Ltd.)

FIG 12.13 Oral hairy leukoplakia, a manifestation of Epstein–Barr virus infection in HIV-infected individuals. (From Hall LD et. al. Epstein–Barr virus: dermatologic associations and implications. *J Am Acad Dermatol* 2015;72(1):1–19.)

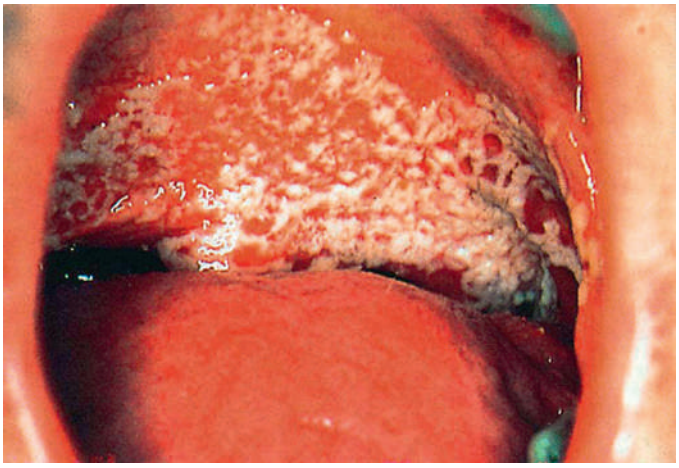


FIG 12.14 *Candida albicans* in an HIV-infected person. (From Taylor SC et al. *Treatments for skin of color,* Philadelphia, 2011, Saunders.)

FIG 12.15 HIV-associated Kaposi sarcoma in the macular stage. (From Schwartz RA et. al. Kaposi sarcoma: a continuing conundrum. *J Am Acad Dermatol* 2008;59(2):179–206.)

sides of the tongue (Fig. 12.13). These lesions cannot be removed with a tongue blade, which differentiates this infection from oral candidiasis or thrush. They are not usually painful. There is no specific treatment for oral hairy leukoplakia. It will often resolve with treatment of the HIV infection with improvement in the $CD4^+$ cell count.

In HIV-infected persons, herpes simplex viruses 1 and 2 cause the formation of large groups of painful vesicles on an erythematous base. The HSV-1/-2 vesicles then rupture, crust, and become large, ulcerative, and occasionally necrotic. They are chronic and painfully persistent and usually occur on the genitalia, digits, and perianal or perioral areas. Herpes simplex virus may produce protein that enhances the replication of HIV.

Bacterial infectious causes of mucocutaneous manifestations include *Mycobacterium avium* or *Staphylococcus aureus. Staphylococcus* is a common bacterial skin infection in patients with HIV associated with folliculitis, furuncles (boils), or bullous impetigo. Occasionally sepsis may occur. Treatment includes application of topical antibiotics, use of an antibacterial soap, and administration of systemic antibiotics, either oral or intravenous, as needed. Abscesses usually are surgically opened and drained.

Fungal skin infectious agents include *Candida* (Fig. 12.14), *Cryptococcus,* or *Histoplasma*. Vaginal candidiasis is the most common early skin symptom in HIV-positive women. Other infectious agents include parasites such as the mites that cause scabies. Treatment is with topical, oral, or intravenous antifungal or antiparasitic agents, depending on the severity of the infection.

Neoplasms can also occur, including Kaposi sarcoma, squamous cell carcinoma, basal cell carcinoma, or cutaneous lymphomas. Kaposi sarcoma is a vascular neoplasm that affects the skin and mucous membranes, lymphatics, and other internal organs. It occurs in immunosuppressed patients and is caused by human herpesvirus 8 (HHV-8). Kaposi sarcoma is one of the few neoplasms indicative of immune system dysfunction. It is the second most common tumor found in HIV-infected patients worldwide and the most common cancer in sub-Saharan Africa.

The skin lesions of Kaposi sarcoma are individual tumors that begin as flat or macular subcutaneous patches. The patches initially range from light pink to deep purple and are painless, nonblanching, and nonpruritic (Fig. 12.15). The lesions evolve from patches into thickened plaques or large nodules that may change to brown over time, especially in darkly pigmented persons. They may occur anywhere on the body and usually begin on the head—face, eyelids, conjunctivae, pinnae, scalp, or buccal membranes (Fig. 12.16). Lesions range in size from a few millimeters to coalesced patches covering large areas of the body. The lesions are highly vascular but do not bleed excessively. The lesions also may occur internally in either the lungs or the intestines in approximately 40% of patients.

Kaposi sarcoma is usually managed with a combination of chemotherapy and highly active antiretroviral therapy (HAART). Occasionally radiation therapy is also used. Surgery is rarely indicated except to remove large, uncomfortable lesions. Radiation therapy is used primarily for oral or cutaneous lesions. Mitotic inhibitors (e.g., vinblastine, vincristine), liposomal anthracyclines (bleomycin, doxorubicin), or immunomodulators (e.g., interferon-α-2b) have been useful chemotherapeutic agents in the management of Kaposi sarcoma. Smaller lesions may be managed with intralesional injections of a mitotic inhibitor. Those with the best prognosis tend to have limited disease; no other opportunistic infections; and no weight loss, fevers, or night sweats.

Gynecologic Manifestations

The primary genital barrier to HIV virus is the genital epithelial cell. When proinflammatory cytokines are produced in response to HIV

FIG 12.16 HIV-associated Kaposi sarcoma in the nodular stage. (From Schwartz RA et. al. Kaposi sarcoma: a continuing conundrum. *J Am Acad Dermatol* 2008;59(2):179–206.)

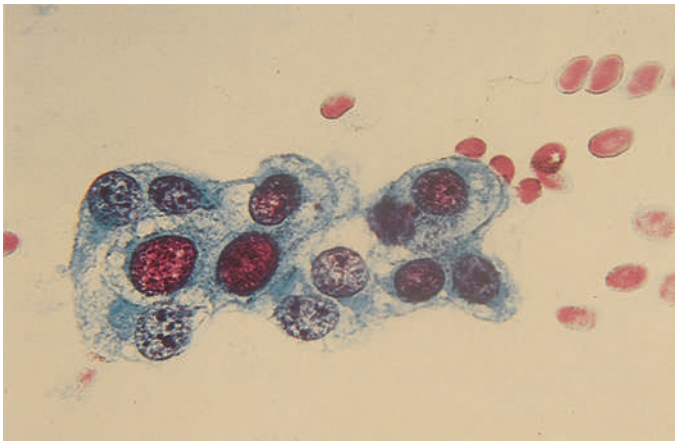


FIG 12.17 Cervical dysplasia showing epithelial cells of irregular shape and size. Women with AIDS require more frequent monitoring because cervical dysplasia commonly occurs and progresses rapidly.

virus, the usually tight epithelial junction barrier becomes impaired, allowing HIV virus to move across the epithelium. The gynecologic manifestations of HIV disease are marked by persistent monilial vaginitis secondary to *C. albicans,* cervical dysplasia, and neoplasia, as well as pelvic inflammatory disease (PID). PID is common in HIV-positive women. It is caused by a variety of organisms, including *C. trachomatis,* and is managed with antibiotics.

Cervical dysplasia affects 40% of HIV-infected women (Fig. 12.17). Cervical dysplasia has no symptoms, but the cell changes can lead to neoplasia (cancer). Therefore either Papanicolaou smears or colposcopic examinations should be performed every 6 months to detect cervical cancer early in HIV-positive women. Cervical cancer is particularly aggressive in women with HIV.

Neurologic Manifestations

Neurologic manifestations are often the reason that people with HIV seek treatment. HIV invades the central nervous system (CNS) early in the course of its infection during primary systemic viremia through migrating myeloid and lymphoid cells. HIV then infects perivascular macrophages, astrocytes, and microglial cells, which are the main cellular targets of HIV in the brain. Microglial HIV infection results in elevated levels of the enzyme *glutaminase,* which produces increased extracellular glutamate concentrations leading to increased neurotoxicity.

Central and peripheral nervous system manifestations can be caused by HIV infection directly, as a result of infectious agents, secondary to neoplasms causing space-occupying lesions, or from secondary immune-inflammatory reactions. Opportunistic infectious agents affecting the neurologic system include *Toxoplasma* and *Cryptococcus.* A variety of peripheral neuropathies can result from HIV infection directly, although some may be caused by herpes zoster infection. History of coinfection with hepatitis C, substance abuse, and other comorbidities can compound the neurologic manifestations associated with HIV infection. Socio-economic factors (such as low educational level) and psychological factors (such as depression, mood disorders, and sleep disorders) may also contribute to some neurologic dysfunction.

The most common HIV-associated neurologic manifestation is known as *HIV-associated neurocognitive abnormality (HAND).* HAND is divided into three subdisorders, also known as *Frascati criteria.* They include asymptomatic neurocognitive impairment (ANI), mild neurocognitive disorder (MND), and HIV-associated dementia (HAD). HAND is caused directly or indirectly by HIV infection or viral products, by cytokine-related cellular damage, or by the competition or interference between gp120, *tat,* or *vpr* and neuroleukin, a nerve growth factor. This disorder can affect both adults and children and is often undiagnosed in its early stages. Perinatally infected children with HIV have poorer overall cognitive function compared with non–HIV-infected children in a recent study.

In early disease with ANI, the symptoms involve an abnormality in two or more cognitive abilities with no functional impairment. Progression to MND involves cognitive impairment with mild functional impairment. HAD is characterized by marked cognitive impairment with marked functional impairment. HAD can occur when other opportunistic infections begin to appear later in the disease process and is characterized by progressive cognitive impairment or subcortical dementia. In other words, the patient is alert but demented and confused. Computed tomography shows diffuse atrophy in the cerebral cortex, widened sulci, ventricular enlargement, and shrinking of the basal ganglia. Cerebrospinal fluid analysis shows elevated protein and abnormal IgG levels.

The cognitive neurologic symptoms associated with HAND include inattentiveness, confusion, forgetfulness, loss of concentration, slower verbal response, headache, apathy, and inability to complete or perform complex tasks. Symptoms may wax and wane over the course of a day, with intermittent periods of lucidity and confusion. Symptoms can progress to global dementia associated with marked memory impairment and disorientation. Before motor strength declines, the patient may forget time, place, person, and activities, leading to safety issues such as wandering, leaving appliances on, and forgetting to take medications. The associated focal motor deficits include slower motor responses, clumsiness, weakness, loss of balance, handwriting changes, and slurred speech. With progression, motor strength declines, with subsequent large muscle weakness causing difficulty walking and moving. Associated generalized symptoms consist of fever and mild metabolic acidosis. Behavioral symptoms include personality changes, social withdrawal, depression, poor hygiene and grooming, lack of insight, apathy, agitation, and, less commonly, anxiety and hyperactivity. In children, head circumference does not increase with age. As the disorder progresses, the neurologic symptoms become more severe. Ataxia, hypertonia, tremors, and incontinence appear. The person may be alert but cognitively impaired, mute, and paraplegic. Hemianopia (partial blindness), myoclonus, and seizures also may develop. The person may become comatose and lethargic with other systemic dysfunctions.

Management of HAND includes treatment with antiretroviral agents and neuroleptics (to control agitation), as well as coordination of home

and environmental safety plans and patient/family support. Use of neurocognitive function testing is important. Simple self-administered tests can be used when patients come to a clinic for screening. More detailed screening tests such as the HIV Dementia Scale are useful for more comprehensive evaluations of patients who are increasingly symptomatic.

Ocular Manifestations

Eye diseases occur in up to 70% of patients who have HIV infection. Ocular manifestations of HIV infection may be of infectious or noninfectious origin. Noninfectious causes of ocular problems include HIV retinopathy, vascular anomalies, malignancy, and diseases induced by medication. Infections, particularly opportunistic infections, are the most common causes of eye diseases in HIV patients. Infectious causes include bacteria such as *Treponema pallidum* (syphilis) and *Staphylococcus;* fungi including *Candida, Cryptococcus,* and *Histoplasma;* protozoa such as *Pneumocystis* and *Toxoplasma;* and viruses such as herpes simplex and CMV.

The most common and severe type of ocular infection is CMV retinitis, which affects up to 30% of HIV patients usually late in the disease course when $CD4^+$ T-cell counts are low. After an insidious onset, ocular CMV causes perivascular hemorrhages, fluffy exudates, and vasculitis in the retina, leading to destruction and necrosis of the retina with resulting blindness. Symptoms include decreased visual acuity, floaters, and damage to the retina, leading to optic nerve atrophy, optic neuropathy, retinal detachment, and blindness. Treatment of CMV retinitis involves the use of anti-CMV agents (e.g., ganciclovir or foscarnet) intravenously and the use of oral anti-CMV agents for prophylaxis. An intraocular sustained-release anti-CMV implant is also available for the management of acute and chronic CMV retinitis.

HIV-associated retinopathy causes the development of cotton-wool spots and microvascular retinal changes. Cotton-wool spots are small, indistinct white spots with associated hemorrhage. These changes are not as severe as CMV retinitis and may remit spontaneously.

Cardiovascular Manifestations

Cardiovascular disease is an important treatment consideration in an aging HIV-positive population due to preexisting cardiovascular disease, age, sex, race/ethnicity, comorbidities, and hereditary factors, as well as the direct effect of HIV infection and treatment. HIV-infected patients have an increased risk for a variety of cardiovascular diseases, including dyslipidemia, earlier incidence and progression of atherosclerosis, hypertension, acute myocardial infarction, coronary artery disease, cardiac arrhythmias, and cardiomyopathy. Lipid abnormalities and cardiovascular disease are due to HIV immunosuppression, $CD4^+$ T-cell count, HIV viral load, endothelial function, HAART, or a combination of these factors. Cardiomyopathy and myocarditis are due to direct invasion of myocardial cells by HIV virus or as a result of inflammation and autoimmune mechanisms, although the exact mechanisms are not clear. Studies of autopsy findings have found 50% of AIDS patients have myocardial inflammation. Hypertriglyceridemia and lipogenesis are the result of inflammatory responses to HIV infection secondary to elevated serum IFN-α level, as well as decreased triglyceride clearance from the blood in HIV infection. Antiretroviral therapies may also cause dyslipidemias primarily because of the effects on hepatocytes. Nonucleoside reverse transcriptase inhibitors (NRTIs) affect plasma lipids less than ritonavir-boosted protease inhibitors. In general, newer antiretroviral therapies affect lipid profiles less than older therapies.

Symptoms of cardiovascular disease in HIV-infected patients vary from vague symptoms such as weakness, dyspnea, and fatigue to acute chest pain, creatine kinase (CK) elevations, ventricular and valvular dysfunction, and pulmonary edema. The treatment of HIV-associated cardiovascular disease includes education/counseling on heart-healthy diet, exercise, and smoking cessation; management of hypertension and diabetes if present; treatment of dyslipidemia with medications as needed; and use of antiplatelet agents. The treatment of dyslipidemia includes the use of fish oils, fibrates, statins, niacin, and ezetimibe. Of the statins, atorvastatin (Lipitor), pravastatin (Pravachol), rosuvastatin (Crestor), and pitavastatin (Livalo) are less likely to cause adverse interactions with HAART medications. Simvastatin is contraindicated when taking protease inhibitors because of toxic drug–drug interactions when taken together. Fibrates, such as gemfibrozil, are generally well tolerated and do not interact adversely with most HAART medications. Ezetimibe, in combination with Pravachol, appears to benefit most patients. However, ezetimibe, which blocks cholesterol absorption in the intestine, does not appear to lower lipid levels alone.

Manifestations in Other Systems

Renal impairment can also occur with HIV infection. This is due to high circulating levels of immune inflammatory markers, such as TNF receptor 2, IL-2, and others, and/or the effect of antiretroviral therapy. Increased kidney dysfunction over time, as indicated by a decline in GFR and increasing albuminuria, can lead to AIDS-associated nephropathy (AIDS-related glomerulopathy or HIV-associated nephropathy), drug-induced ischemia, and renal failure. It is recommended that frequent renal function screening be performed to monitor for renal impairment.

Hematologically, individuals with HIV have anemia, thrombocytopenia, and granulocytopenia. Liver dysfunction is also a problem in persons with HIV because there is an increased risk of concomitant hepatitis C or B (HCV/HBV) infection. This comorbidity of HIV/HCV or HIV/HBV causes increased mortality from end-stage liver disease. Also compounding liver dysfunction in patients with HIV is the effect of multiple medications on the liver itself.

HIV-related endocrine and metabolic abnormalities are due to direct HIV infection; chronic inflammation; destruction of endocrine tissue from infection, cancer, or inflammation; use of pharmacologic agents; or the effect of severe illness on hormonal function/secretion and metabolic homeostasis. Insulin resistance is a major metabolic complication of HIV. Changes in the glucocorticoid and insulin sensitivity of target tissues and altered cytokine production may be due to the effects of HIV proteins on the hypothalamic–pituitary–adrenal axis or excess renin–angiotensin–aldosterone system activation. These effects can be either direct (from insulin signal interference at the cellular level) or indirect (from antiretroviral therapy and lipid dysfunction). Insulin resistance leads to the development of abnormal glucose metabolism, which can cause metabolic syndrome changes, including obesity and diabetes mellitus. The adrenal gland is another organ affected by HIV infection because of increased adrenal secretion of cortisol with stress. Injury to the adrenal gland rarely leads to frank cortisol deficiencies. Levels of thyroid hormones (e.g., triiodothyronine [T_3], thyroxine [T_4], and thyroid-stimulating hormone) may be elevated or decreased. Thyroid hormone abnormalities, dysregulation of the autoimmune response, and development of Graves disease can occur in some HIV-infected patients. Hypogonadism can be primary in HIV infection or secondary to antiretroviral therapy.

Adipose tissue endocrine function is also affected by HIV infection and antiretroviral therapy. This dysfunction may cause the development of HIV-1 HAART-associated lipodystrophy syndrome in some persons. It is characterized by subcutaneous adipose tissue atrophy, dorsocervical fat ("buffalo hump") accumulation, and visceral adipose tissue hypertrophy. The cause of this dysfunction is multifactorial due to the effects on

systemic metabolism, as well as enhanced production of proinflammatory cytokines and excessive free fatty acid release and is not well understood. Not all people treated with antiretroviral therapy develop lipodystrophy. It is suspected that there may be a genetic component or predisposition to development, history of dyslipidemia, or long-term treatment.

Rheumatologic manifestations of HIV infection are varied, encompassing osteoporosis, osteopenia, myopathies, and musculoskeletal infections (infectious arthritis and osteomyelitis). Musculoskeletal infections are caused by a decrease in the number of $CD4^+$ T cells. The most common infectious organisms include *S. aureus, Streptococcus pneumoniae, C. albicans, Mycobacterium kansasii,* and *Mycobacterium avium-intracellulare.* Other manifestations are caused by immune-mediated arthritis (such as Reiter disease, psoriatic arthritis, and undifferentiated spondyloarthropathy syndromes) or occur as a direct result of the immune response to HIV (including polymyositis, vasculitis, and immune complex diseases related to the production of autoantibodies). The increasing incidence of osteopenia, osteoporosis, and osteonecrosis in the HIV population is significant and may be due either to HIV infection or to antiretroviral therapy. Treatment includes using vitamin D plus calcium replacement, bisphosphonate medications, and selective estrogen receptor modulators as well as risk factor modification through lifestyle changes.

Manifestations in Children

Children with HIV become symptomatic much faster than adults—usually within their first year of life. Because of the invasion of virus, children's growth and development are markedly affected, including physical growth retardation with failure to thrive, impaired intellectual development, and impaired motor functioning with decreased coordination. The infant develops normally until the virus begins its nervous system invasion. After that time, neurologic impairment is characterized by development of weakness, loss of previously accomplished developmental milestones, hypotonia, or hypertonia. Extensive candidiasis without any relationship to antibiotic therapy may be an early symptom. Respiratory problems, including the development of PCP, are common. Endocrine changes in children, including adrenal insufficiency, growth disorders, thyroid dysfunction, metabolic abnormalities, and osteopenia, can occur.

Serious bacterial and viral infections can also develop in HIV-infected children. Without treatment or vaccination, HIV-infected children develop severe and repeated bouts of communicable diseases such as chickenpox or measles, as well as opportunistic infections. Routine vaccinations for all vaccine-preventable diseases such as tetanus and measles are recommended for HIV-infected children. The long-term response to the vaccines is poor due to defective cellular immune responses and T-cell anergy. Over time, the antibodies to these diseases decline with decreasing seroprotection rates. As a result, it is recommended that all children and adolescents have booster doses of vaccines to maintain effective antigen-specific immune response.

For HIV-positive and -negative infants, the current recommendation is for HIV-positive mothers on antiretroviral therapy to breast feed their infants. Breast milk of HIV-positive mothers on antiretroviral therapy has been found to inhibit HIV transmission with low risk of breast-feeding transmission. This is particularly helpful in resource-limited settings where access to infant formula is limited. Because lactose intolerance is common in HIV-infected children, soy formulas are often used when formula feeding. Dairy products may be introduced into the diet gradually, as tolerated. Children need particular attention to their diet for adequate growth and development with increased intake of calories and protein, as well as nutrient-rich snacks such as raisins, fruits, and nuts.

KEY POINTS

- All body systems are affected by HIV.
- Early HIV infection is characterized by fever, chills, headaches, nausea, vomiting, diarrhea, fatigue, weakness, arthralgia, sore throat, stiff neck, photophobia, irritability, and rash.
- The most significant systemic symptom is malnutrition or wasting, which is due to a combination of factors, including an elevated metabolic rate, chronic inflammation, malabsorption, anorexia, and the effect of multiple opportunistic insults.
- GI symptoms occur frequently in patients with HIV. Symptoms include diarrhea caused by *Cryptosporidium* or other agents, ulceration, and candidiasis, as well as multiple opportunistic infections.
- Pulmonary symptoms include opportunistic pneumonias (particularly PCP), tuberculosis, and adult respiratory distress syndrome.
- Mucocutaneous symptoms occur both early and late in the course of HIV infection. One of the first symptoms is the viral exanthem that occurs during the primary infection. Other manifestations may be allergic; infectious, such as candidiasis or herpes, human papillomavirus, and Epstein–Barr virus infections; or neoplastic, such as Kaposi sarcoma.
- Neurologic manifestations include peripheral neuropathy, encephalopathy with dementia, headache, apathy, and focal deficits.
- Gynecologic manifestations include persistent monilial vaginitis, cervical dysplasia, and PID.
- Ocular manifestations include HIV-associated retinopathy, CMV retinitis, malignancy, and a variety of infectious causes.
- Children with HIV have growth and development problems, including impaired physical growth, intellectual development, and motor functioning.

TREATMENT

Antiretroviral Therapy Recommendations

The WHO issued new guidelines for the treatment of HIV-1 infection in 2015. WHO recommends lifelong treatment for all children, all pregnant women, and all adults and adolescents regardless of $CD4^+$ count with priority for those with $CD4^+$ counts below 350 cells/mm^3 and those with AIDS-defining illness. They also recommend preexposure prophylaxis as additional prevention for people at substantial risk of HIV infection. These guidelines were developed as a result of the TEMPRANO and START trials, which showed starting early treatment ($CD4^+$ cell count above 500 cells/mm^3) results in less serious illness and fewer AIDS-related deaths. Early antiretroviral treatment seems to be better tolerated by patients and helps to prolong life, but it is imperative that patients starting antiretroviral therapy be committed to lifelong treatment and adherence to therapy with knowledge of the risks and benefits.

The goal of medication and therapeutic management of HIV/AIDS is to delay disease progression, restore or preserve immunologic function, suppress plasma HIV viral load, minimize clinical manifestations, reduce morbidity, prolong survival, and prevent HIV transmission. Drug management of HIV infection has evolved from monotherapy, or therapy with one agent, to the use of multiple medications, called *HAART*. This polydrug therapy approach involves the administration of multiple antiretroviral agents. It provides better viral suppression, thereby decreasing viral load, increasing $CD4^+$ counts, and decreasing resistance for a longer period. The objective of HAART is to provide the greatest viral suppression for the longest time to prevent viral mutations. This approach makes good sense. If one drug blocks 90% of viral replication, the others may eliminate the rest of the resistant virions. However, even with HAART, complete viral eradication is not possible with current treatment strategies. It has also been found that continuous HIV therapy

is more beneficial than treatment interruption with prolonged drug treatment holidays.

Antiretroviral therapy in HIV infection is more successful in some patients than in others, depending on both internal and external factors. Some important factors include prescribing the appropriate polydrug therapy and ensuring maximal convenience of therapy, as well as aiding patient adherence to the treatment regimen. Other factors include early identification of HIV infection and starting early HAART. However, there are also unknown factors involving the response of individual patients to the therapy, including the rate of reduction of viremia in response to treatment. People who present for treatment later in HIV infection with higher viremia and lower $CD4^+$ counts tend to have increased morbidity, increased progression, and decreased immune system responses, as well as poor adherence to therapy.

HIV drug treatment failure is the result of HIV resistance, which is a widespread concern and is usually caused by poor adherence to HAART, poor toleration of the drugs, prior exposure to single or multiple antiretroviral drug therapy, or counteracting interactions among the drugs used. Virus resistance increases with treatment protocol noncompliance. It is easy to see why persons taking the multidrug regimen may fail to comply with nutritional and drug therapy because of the sheer volume of drugs to be taken in a day. A person infected with HIV may take up to 13 to 30 pills per day; he or she must also remember which medications should be taken with food or on an empty stomach and which medications cannot be taken together simultaneously. In addition, some persons with HIV infection may be demented because of the disease, homeless, or addicted to intravenous drugs—any of which limits the person's ability to adhere to the strict treatment regimens. Finally, some of the treatments for opportunistic infection or cancer involve many other drugs: intravenous, oral, and intracavital. Patient compliance with complicated drug therapy that has many side effects and high cost is often variable. Therefore the key appears to be encouraging patient compliance by tailoring medication regimens that improve tolerability and convenience.

Some of the current medications available to adults with HIV have not been recommended for children younger than 13 years. In the past, treatment for infants and children was often delayed weeks or months. This is because virion levels are usually not detectable for up to 2 weeks after birth. With the new guidelines, treatment is started early.

The major classes of antiretroviral medications are nucleotide and nucleoside reverse transcriptase inhibitors (NRTIs), nonnucleoside reverse transcriptase inhibitors (NNRTIs), protease inhibitors (PIs), fusion inhibitors, integrase strand transfer inhibitors (INSTIs), CCR5 inhibitors, and antiretroviral boosters. The current optimal combination of medications for treatment of HIV-1 involves using at least three different drugs from two different classes for initial treatment. Usually there are two nucleoside analogs (NRTIs), such as tenofovir and emtricitabine or abacavir and lamivudine, and one active drug from one of the following classes: PIs boosted with ritonavir, an NNRTI (which is called *a protease-sparing regimen*), an INSTI, or a CCR5 antagonist. Viral resistance to treatment is reduced by the complete suppression of the virus with multiple drugs (Fig. 12.18). There are now a variety of antiretroviral medication choices available to meet patient needs, tolerance, and cost concerns (Table 12.4). Medication treatment options for HIV-2–infected patients are more limited. HIV-2 is resistant to NNRTIs, although it may be sensitive to NRTIs.

Nucleoside Reverse Transcriptase Inhibitors

NRTIs resemble the natural substances used by the virus to build HIV DNA. NRTIs prevent HIV replication by inhibiting HIV reverse transcriptase, thereby preventing HIV DNA synthesis. They have been found to slow progression of the disease. Nucleoside analogs include zidovudine, didanosine, zalcitabine, stavudine, lamivudine, and abacavir. All must be converted to an active state intracellularly.

Nucleotide Reverse Transcriptase Inhibitors

Nucleotide RTIs act like the nucleoside RTIs noted earlier, but they use a different enzyme to become activated intracellularly. They are able to be activated in more cells due to this difference. Tenofovir is the major nucleotide RTI.

Nonnucleoside Reverse Transcriptase Inhibitors

NNRTIs also inhibit reverse transcriptase, but by a different mechanism. Drugs in this class include efavirenz (Sustiva), nevirapine (Viramune), and delavirdine (Rescriptor). These medications need not be converted intracellularly to be activated. The greatest benefit is that they are potent antiretrovirals. The biggest problem is that they affect the cytochrome P450 system, which increases drug interactions and means they must

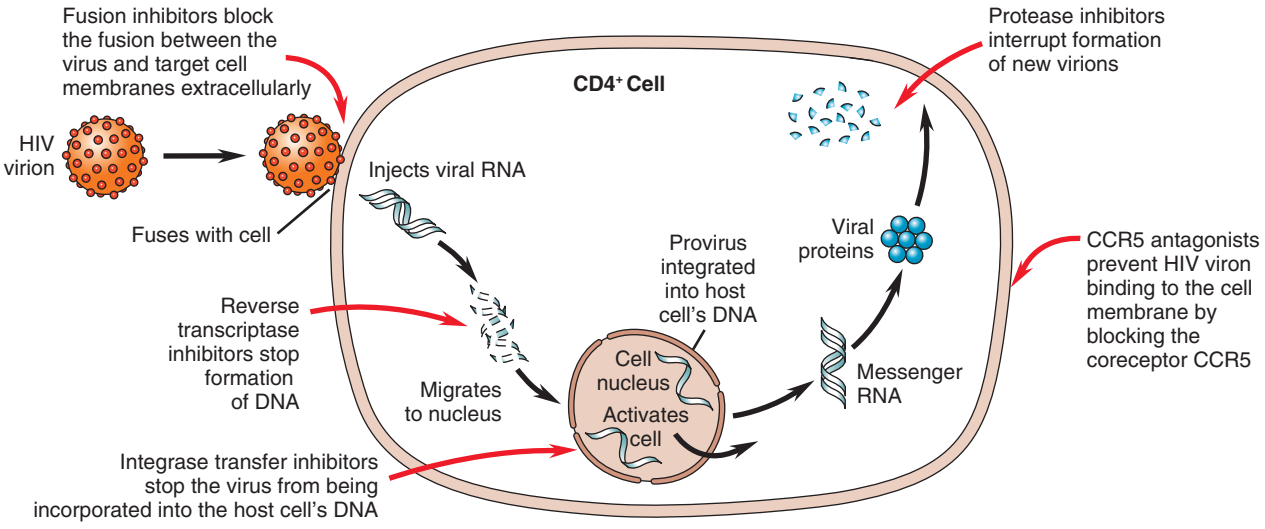


FIG 12.18 Antiretroviral therapy. The stages in the life cycle of HIV in which antiretroviral therapy is effective are shown.

TABLE 12.4 Antiretroviral Therapy

Class	Generic Name (Trade Name)	Use in Pregnancy or Children	Side Effects	Monitoring
Nucleoside reverse transcriptase inhibitors	Zidovudine (AZT, Retrovir) Lamivudine (3TC, Epivir) Didanosine (DDI, Videx) Stavudine Emtricitabine (Emtriva, FTC) Abacavir (Ziagen, ABC)	Pregnancy Class B or C: approved and used in pregnancy as early as 10 weeks' gestation Can be given to infants and children. Unlikely to interact with drugs metabolized by cytochrome P450 enzyme system.	Pancreatitis, bone marrow toxicity, anemia/neutropenia, peripheral neuropathy, dyslipidemia, insulin resistance, hepatic toxicity, headache, nausea, vomiting, diarrhea, insomnia, malaise, myalgia, hypersensitivity, lactic acidosis, osteopenia, confusion Abacavir hypersensitivity reaction in patients with positive HLA-B'5701.	Liver function studies, CBC, metabolic panel studies (chem 14)
Nucleotide reverse transcriptase inhibitors	Tenofovir (Viread, TDF)	Pregnancy Class B. Avoid concomitant use with NSAIDS. Necessary to stay hydrated.	Renal dysfunction, bone loss, dehydration	Liver function studies, CBC, metabolic panel studies (chem 14), serum phosphate
Nonnucleoside reverse transcriptase inhibitors	Nevirapine (Viramune) Delavirdine (Rescriptor) Efavirenz (Sustiva, EFV) Etravirine (ETR, Intelence) Rilpivirine (Edurant)	Pregnancy Class D: not approved d/t teratogenic and CNS defects Avoid during lactation Approved in children over 3 yr Should be taken with a high-fat meal. Significant potential for effect on cytochrome P450 enzyme system.	CNS effects, dizziness, drowsiness, concentration problems, insomnia, vivid dreams, depression, headache, maculopapular rash, nausea, vomiting, diarrhea, myalgia, hyperlipidemia, hyperglycemia, liver enzyme elevations, Stevens–Johnson syndrome, hepatitis, vitamin D deficiency	Liver function studies, CBC, chem 14, lipid function; take between 6 and 9 PM to sleep through side effects
Protease inhibitors	Saquinavir (Invirase, SQV) Ritonavir (Norvir, RTV) Atazanavir (Reyataz, ATV) Nelfinavir (Viracept) Tipranavir (Aptivus, TPV) Fosamprenavir (Lexiva, FPV) Darunavir (Prezista, DRV)	Pregnancy Class B or C: approved in pregnancy and in children Avoid during lactation. Must take with food. Have many serious drug interactions especially with statins. Often combined with other antivirals to overcome viral resistance and lower drug dosages.	Hyperlipidemia, lipid abnormalities, lipodystrophy, elevated LFTs/uric acid, dizziness, anxiety, bleeding, pancreatitis, MI, stroke, weakness, asthenia, headache, nausea, diarrhea, vomiting, anorexia, abdominal pain, taste perversion, paresthesias, spontaneous bleeding in people with hemophilia	Liver function studies, CBC, chem 14, lipid function, uric acid levels
Fusion inhibitor	Enfuvirtide (ENF, Fuzeon)	Pregnancy Class B: approved. Approved in children older than 6 yr Must be given twice daily by subcutaneous injection.	Fatigue, insomnia, diarrhea, nausea, abdominal pain, anorexia, elevated CPK, myalgia, cough, bacterial pneumonia, flulike syndrome, thrombocytopenia, hyperlipidemia, elevated LFTs	Liver function studies, chem 14, lipid function
CCR5 antagonist	Maraviroc (Selzentry)	Pregnancy Class B: approved. Approved in adolescents 16 yr and older Not recommended in lactation.	Hepatotoxicity, hypersensitivity reactions, fever, cough, vascular hypertensive disorder, dizziness, insomnia, pruritus, lipodystrophy, elevated LFTs, peripheral neuropathy, myalgia	Liver function studies, CBC, chem 14, lipid function
Integrase transfer inhibitor	Raltegravir (Isentress, RAL) Dolutegravir (Tivicay, DTG)	Pregnancy Class B or C: approved. Approved in adolescents 16 yr and older	Myopathy, rhabdomyolysis, hyperlipidemia, hypertension, fatigue, dizziness, elevated glucose/lipase/LFTs/creatinine, anemia, thrombocytopenia, diarrhea, nausea	Liver function studies, CBC, chem 14, lipid function, uric acid levels

CBC, Complete blood cell count; *CNS,* central nervous system; *CPK,* creatine phosphokinase; *d/t,* due to; *LFTs,* liver function tests; *MI,* myocardial infarction; *yr,* year.

be given cautiously. They can only be administered in combination with other antiretrovirals.

Protease Inhibitors

PIs attack at another phase in the viral life cycle. These medications inhibit the enzyme protease, whose action is to clip the viral protein precursors to the appropriate size and is essential for HIV maturation, infection, and replication. As a result, immature, defective, and noninfectious viral particles are released. Drugs in this class include saquinavir, ritonavir, indinavir, nelfinavir, fosamprenavir, atazanavir, ritonavir-boosted tipranavir, and ritonavir-boosted darunavir. They also affect the cytochrome P450 system. Protease inhibitor therapy is extremely expensive, with a retail price of approximately $6000 to $8000 per year. Protease inhibitors are never used as single agents because of the potential for a patient to develop resistance. PIs have poor CNS penetration. The advantages of ritonavir-boosted drugs include overcoming viral resistance, lower drug dosages with fewer pills, and less frequent dosing.

Fusion Inhibitors

Fusion inhibitors (e.g., enfuvirtide [Fuzeon, ENF]) work extracellularly by blocking fusion between viral and target cell membranes. They are always used with other treatment regimens to decrease viral load, never as monotherapy. They increase the effects of protease inhibitors. They are given only by subcutaneous injection. The pediatric dosage is weight based and given twice daily. The adult dose is usually 90 mg twice daily. They are also very expensive—up to $3250 for a 30-day supply.

CCR5 Inhibitors

CCR5 inhibitors are best suited for earlier infection when the CCR5 tropic virus (T tropic phase) predominates (see Fig. 12.5). The major drug in this category is maraviroc. This medication acts to prevent infection by blocking the coreceptor CCR5 and preventing HIV binding to the cell membrane. Tropism testing is necessary before treatment with this medication. It is always given with other antiretroviral medications. It also interacts with the P450 system.

Integrase Strand Transfer Inhibitors

INSTIs are the newest class of antiretroviral medications. These medications target and inhibit *integrase* encoded by the viral *pol* gene. This prevents the insertion of viral DNA in the human chromosome. The first INSTI approved by the FDA was raltegravir (Isentress). They are generally well tolerated.

Other Treatments and Vaccines

Antibiotics, antivirals, antifungals, antiparasitics, and antimycobacterial medications are administered as needed to treat the many opportunistic infections that are secondary to HIV infection. Human granulocyte colony–stimulating factor (filgrastim [Neupogen]) may be used to improve innate immunity by increasing the number of neutrophils in persons with neutropenia. This agent is helpful in decreasing medication-induced neutropenia. It is given daily in a subcutaneous dose calculated according to the patient's weight. However, it is of questionable value because it may act as a growth factor for cancers. Epoetin alfa (erythropoietin) is used to manage medication-induced anemia. An associated side effect is hypertension. Intravenous immunoglobulin can sometimes be used in HIV-infected children with T-cell counts greater than 200/μL. It helps to decrease the incidence of serious bacterial, minor bacterial, viral, and opportunistic infections. Interferon alfa-2b (Intron A) is used to treat AIDS-related Kaposi sarcoma in adults. Other studies are being conducted in an effort to find ways to rebuild the immune system. Researchers want to make sure that the T cells cloned after initiation of antiretroviral therapy will respond to both new and old infections.

Prevention of HIV infection by vaccine-conferred active immunity is the ultimate goal of current research. This task is extremely difficult because of HIV strain variability and HIV mutation frequency. To date, none of the vaccine trials have been successful, and several have been stopped in phase II. Ongoing research is being carried out in an effort to quantify the effectiveness of other vaccines in stimulating both cellular and humoral responses to HIV. At this time, there is no effective vaccine available for prevention of HIV.

KEY POINTS

- Management of HIV and AIDS includes the use of antiretroviral medications, including NRTIs, NNRTIs, PIs, fusion inhibitors, INSTIs, and CCR5 antagonists.
- The current optimal combination of medications is multidrug therapy with two nucleoside analogs (NRTIs) and one active drug from one of the following classes: PIs boosted with ritonavir, an NNRTI, an INSTI, or a CCR5 antagonist.
- Efforts to stimulate immune function with peptide growth factors and the development of vaccines are under investigation.
- Aggressive treatment of opportunistic infections with appropriate antibiotics and antivirals is a large part of the treatment regimen.

SUMMARY

HIV is an RNA virus that primarily infects and destroys the immune system. In so doing, it destroys one of the basic foundations of human regulation and protection. HIV decreases the body's ability to fight organisms, opens the door to opportunistic infections, and allows neoplasms to emerge. HIV can infect anyone of any age. The ultimate intracellular pathogen, it slowly destroys the host while manufacturing billions of copies of itself. Study of this virus has improved our understanding of the immune system as well as of cellular function.

RESOURCES

HIV Epidemiology

Considerations for antiretroviral use in special patient populations – HIV-2 infection. *AIDS info*. April 8, 2015.

Division of HIV/AIDS Prevention: *HIV in the United States: at a glance*, Atlanta, GA, July, 2014, Centers for Disease Control and Prevention (CDC).

Division of HIV/AIDS Prevention: *HIV and AIDS in the United States by geographic distribution*, Atlanta, GA, May 11, 2015, Centers for Disease Control and Prevention (CDC).

Estrada-Aguirre JA, et al: Protective effect of CCR5 delta-32 allele against HIV-1 in Mexican women. *Curr HIV Res* 11(6):506–510, 2013.

Kuhar DT, et al: Updated US Public Health Service Guidelines for the Management of Occupational Exposures to Human Immunodeficiency

Virus and Recommendations for Postexposure Prophylaxis. *Infect Control Hosp Epidemiol* 34(9):2013.
Kwong J, Gabler S: Counseling, screening, and therapy for newly-diagnosed HIV patients. *Nurse Pract* 40(10):34–43, 2015.
Swanstrom R, Coffin J: HIV-1 pathogenesis: the virus. *Cold Spring Harb Perspect Med* 2(12):a007443, 2012.
Taborda-Vanegas N, et al: Genetic and immunological factors involved in natural resistance to HIV-1 infection. *Open Virol J* 5:35–43, 2011.
UNAIDS: Press Release: *UNAIDS report shows that 19 million of the 25 million people living with HIV today do not know that they have the virus and Core Epidemiology Slides*, Geneva, July 16, 2014, World Health Organization (WHO).
Zapata W, et al: Identification of innate immune antiretroviral factors during in vivo and in vitro exposure to HIV-1. *Microbes Infect* November 6, 2015 [e pub ahead of print].

HIV Biology

Freed EO: HIV-1 assembly, release and maturation. *Nat Rev Microbiol* 13:484–496, published online June 29, 2015.
Guo H, et al: HIV-1 infection induces interleukin-1*B* production via TLR8 protein-dependent and NLRP3 inflammasome mechanisms in human monocytes. *J Biol Chem* 289(31):21716–21726, 2014.
Passos DF, et al: Purinergic signaling and human immunodeficiency virus/acquired immune deficiency syndrome: from viral entry to therapy. *World J Virol* 4(3):285–294, 2015.
Shen R, et al: Interactions between HIV-1 and mucosal cells in the female reproductive tract. *Am J Reprod Immunol* 71(6):2014.
Tamamis P, Floudas CA: Molecular recognition of CCR5 by an HIV-1 gp120 V3 loop. *PLoS ONE* 9(4):e95767, 2014.

Pathogenesis, Clinical Manifestations, and Management

Abraham AG, et al: Kidney dysfunction and markers of inflammation in the Multicenter AIDS Cohort Study. *J Infect Dis* 212(7):1100–1110, 2015.
Bhutia E, et al: Lipodystrophy syndrome among HIV infected children on highly active antiretroviral therapy in northern India. *Afr Health Sci* 14(2):408–413, 2014.
Buechler MD, et al: T cell anergy and activation are associated with suboptimal humoral responses to measles revaccination in HIV-infected children on anti-retroviral therapy in Nairobi, Kenya. *Clin Exp Immunol* 181(3):451–456, 2015.
CDC: Recommended laboratory HIV testing algorithm for serum or plasma specimens. *Quick Reference Guide – Laboratory Testing for the Diagnosis of HIV Infection: updated recommendations*, June 27, 2014. cdc.gov/view/cdc/23447.
Chahar HS, et al: Exosomes and their role in the life cycle and pathogenesis of RNA viruses. *Viruses* 7(6):3204–3225, 2015.
Chiotan C, et al: Cytomegalovirus retinitis in HIV/AIDS patients. *J Med Life* 7(2):237–240, 2014.
Chrousos GP, Zapanti ED: Hypothalamic-pituitary-adrenal axis in HIV infection and disease. *Endocrinol Metab Clin North Am* 43(3):791–806, 2014.
Cohen S, et al: Poorer cognitive performance in perinatally HIV-infected children versus healthy socioeconomically matched controls. *Clin Infect Dis* 60(7):1111–1119, 2015.
Cubas R, et al: Reversible reprogramming of circulating memory T follicular helper cell function during chronic HIV infection. *J Immunol* 195(12):5625–5636, 2015.
deBree GJ, Lynch RM: B cells in HIV pathogenesis. *Curr Opin Infect Dis* 29(1):23–30, 2016.
Dikman AE, et al: Human immunodeficiency virus-associated diarrhea: still an issue in the era of antiretroviral therapy. *Dig Dis Sci* 60(8):2236–2245, 2015.
Doores KJ: The HIV glycan shield as a target for broadly neutralizing antibodies. *FEBS J* 282(24):4679–4691, 2015.
Elbirt D, et al: HIV-associated neurocognitive disorders (HAND). *Isr Med Assoc J* 17:54–59, 2015.
Feinstein MJ, et al: A systematic review of the usefulness of statin therapy in HIV-infected patients. *Am J Cardiol* 115(12):1760–1766, 2015.
Gbabe OF, et al: Treatment of severe or progressive Kaposi's sarcoma in HIV-infected adults. *Cochrane Database Syst Rev* (8):CD003256, 2014.
Kibirige D, Ssekitoleko R: Endocrine and metabolic abnormalities among HIV-infected patients: a current review. *Int J STD AIDS* 24(8):603–611, 2013.
Kiselnik D, et al: Acute myocarditis and myopathy as presenting manifestations of human immunodeficiency virus infection. *Isr Med Assoc J* 17:524–525, 2015.
Koethe JR, Heimburger DC: Nutritional aspects of HIV-associated wasting in sub-Saharan Africa. *Am J Clin Nutr* 91(4):S1138–S1142, 2010.
Loomba-Albrecht LA, et al: Endocrinopathies in children infected with human immunodeficiency virus. *Endocrinol Metab Clin North Am* 43(3):807–828, 2014.
Lucas GM, et al: Clinical practice guideline for the management of chronic kidney disease in patients infected with HIV: 2014 update by the HIV Medicine Association of the Infectious Diseases Society of America. *Clin Infect Dis* 59(9):e96–e138, 2014.
Mir F, et al: *HIV and nutrition*. March 4, 2014.
Nightingale S, et al: Controversies in HIV-associated neurocognitive disorders. *Lancet Neurol* 13(11):1139–1151, 2014.
Nkenfou CN, et al: Distribution of CCR5-Delta32, CCR5 promoter 59029 A/G, CCR2-641 and SDF1-3'A genetic polymorphisms in HIV-1 infected and uninfected patients in the west region of Cameroon. *BMC Res Notes* 6:288, 2013.
Pham TV, Torres M: Human immunodeficiency virus infection-related heart disease. *Emerg Med Clin North Am* 33(3):613–622, 2015.
Price RW, et al: Evolving character of chronic central nervous system HIV infection. *Semin Neurol* 34(1):7–13, 2014.
Salerno D, et al: Serum and bal beta-D-glucan for the diagnosis of Pneumocystis pneumonia in HIV positive patients. *Respir Med* 108(11):1688–1695, 2014.
Schuetz A, et al: Initiation of ART during early acute HIV infection preserves mucosal Th17 function and reverses HIV-related immune activation. *PLoS Pathog* 10(12):e1004543, 2014.
Selik RM, et al: Revised surveillance case definition for HIV infection – United States, 2014. *MMWR Recomm Rep* 63(RR03):1–10, 2014.
Sension M, Deckx H: Lipid metabolism and lipodystrophy in HIV-1-infected patients: the role played by nonnucleoside reverse transcriptase inhibitors. *AIDS Rev* 17(1):21–38, 2015.
Shen X, Tomaras GD: Alterations of the B-cell response by HIV-1 replication. *Curr HIV/AIDS Rep* 8(1):23–30, 2011.
Srinivasa S, et al: RAAS activation is associated with visceral adiposity and insulin resistance among HIV-infected patients. *J Clin Endocrinol Metab* 100(8):2873–2882, 2015.
Sticchi L, et al: Seroprevalence and vaccination coverage of vaccine-preventable diseases in perinatally HIV-1-infected patients. *Hum Vaccin Immunother* 11(1):263–269, 2015.
Subramaniam P, et al: Levels of salivary immunoglobulin A (SigA) in HIV infected children. *J Clin Pediatr Dent* 39(4):377–381, 2015.
Thorarinsdottir K, et al: $CD21^{-/low}$ B cells: A snapshot of a unique B cell subset in health and disease. *Scand J Immunol* 82(3):254–261, 2015.
Wahl A, et al: Breast milk of HIV-positive mothers has potent and species-specific in vivo HIV-inhibitory activity. *J Virol* 89(21):10868–10878, 2015.
Weetman AP: Thyroid abnormalities. *Endocrinol Metab Clin North Am* 43(3):781–790, 2014.
World Health Organization: *Guidelines on when to start antiretroviral therapy and on pre-exposure prophylaxis for HIV*. September, 2015.

13

Alterations in Oxygen Transport

Susan G. Trevithick

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- What factors are necessary for normal red blood cell production?
- How are oxygen and carbon dioxide transported in the circulation?
- How are laboratory tests used to detect anemia and polycythemia?
- What are the general effects of anemia on body systems?
- How are history, clinical manifestations, and laboratory studies used to differentiate the various forms of anemia?
- How are history, clinical manifestations, and laboratory studies used to differentiate the various forms of polycythemia?
- What are the appropriate treatment measures for each of the common types of anemia and polycythemia?

CHAPTER OUTLINE

Composition of Blood, 260
Organic and Inorganic Components, 261
Cellular Components, 261
Erythrocytes, 262
Leukocytes, 262
Platelets, 262
Structure and Function of Red Blood Cells, 262
Hematopoiesis, 263
Hemoglobin Synthesis, 265
Hemoglobin Synthesis in Infants, 266
Nutritional Requirements for Erythropoiesis, 266
Energy and Maintenance of Erythrocytes, 267
Red Cell Production, 267
Red Cell Destruction, 268
Gas Transport and Acid–Base Balance, 268
Oxygen Transport, 268
Carbon Dioxide Transport, 270
Alterations in Oxygen Transport, 270
Anemia, 272
General Effects of Anemia, 272
Anemia Related to Decreased Red Cell Production, 277
Aplastic Anemia, 277
Anemia of Chronic Renal Failure, 277
Anemia Related to Vitamin B_{12} (Cobalamin) or Folate Deficiency, 278
Iron Deficiency Anemia, 279
Anemia Related to Inherited Disorders of the Red Cell, 279
Thalassemia, 279
Sickle Cell Anemia, 281
Hereditary Spherocytosis, 282
Glucose-6-Phosphate Dehydrogenase Deficiency, 282
Anemia Related to Extrinsic Red Cell Destruction or Loss, 284
Hemolytic Disease of the Newborn, 284
Antibody-Mediated Drug Reactions, 284
Acute Blood Loss, 285
Other Extrinsic Abnormalities, 285
Transfusion Therapy, 286
Polycythemia, 286
Polycythemia Vera, 286
Secondary Polycythemia, 292
Relative Polycythemia, 293

Blood is a critical body fluid composed of formed elements and cells suspended in plasma that circulates through the cardiovascular system. As the primary transport system of the body, blood is involved in the physiologic and pathologic activities of all organs. The red blood cell (RBC), or erythrocyte, is essential to oxygen transport within the circulatory system. RBCs contain large numbers of hemoglobin molecules, which are designed to move oxygen efficiently from the lungs to other body tissues. Hemoglobin also aids in acid–base balance. In addition, RBCs carry carbon dioxide wastes away from the cells and back to the lungs for expiration.

COMPOSITION OF BLOOD

The total blood volume averages 75.5 mL/kg in men and 66.5 mL/kg in women, which is 5 to 6 L or 7% to 8% of body weight. The blood cells comprise approximately 45% and the blood plasma 55% of the

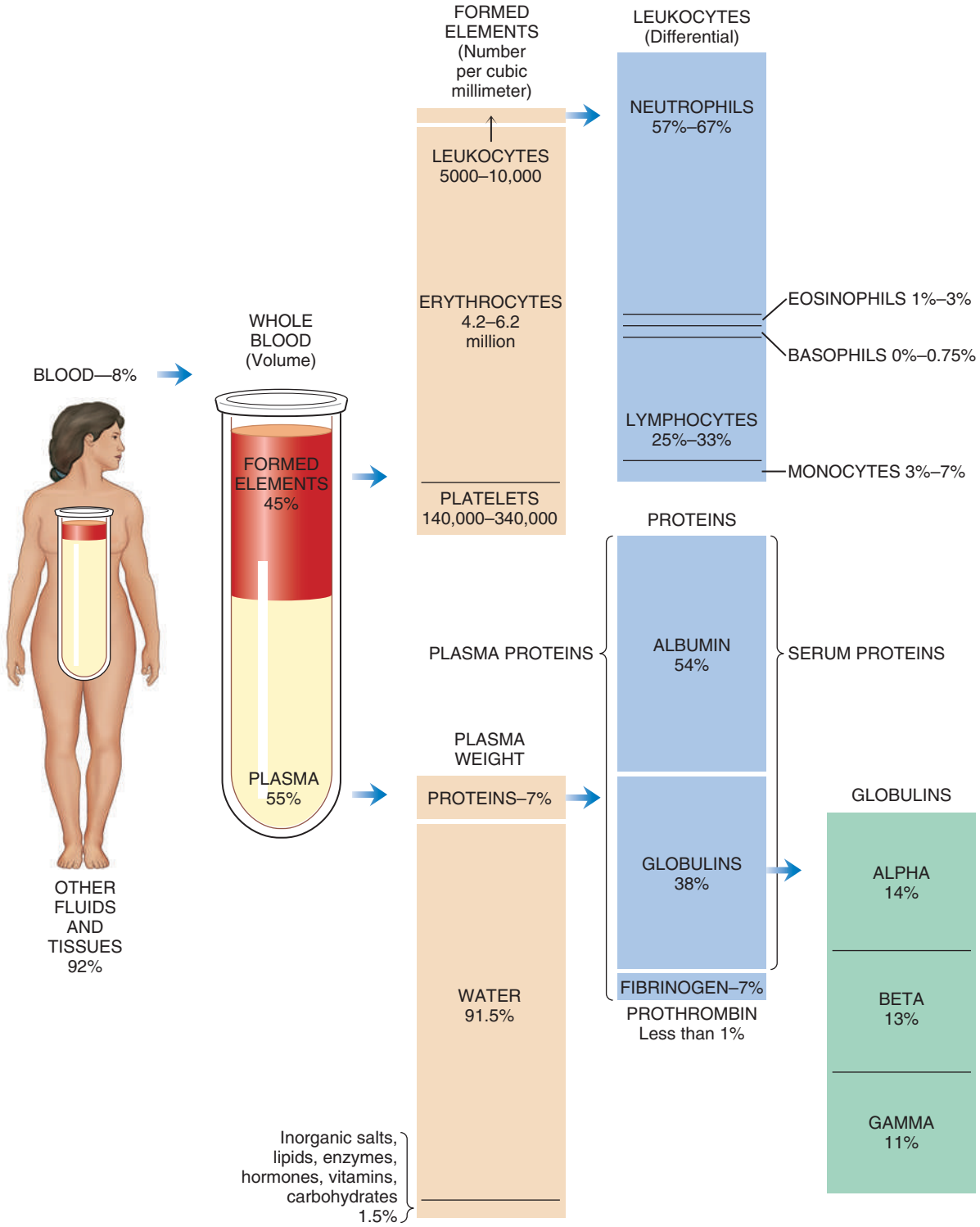


FIG 13.1 Composition of blood in the normal adult.

blood volume. Blood plasma is composed of about 92% water and 7% plasma proteins (Fig. 13.1). The arterial pH of normal blood is 7.35 to 7.45.

Organic and Inorganic Components

The plasma proteins are formed mainly in the liver. They are unable to leave the vascular space under normal circumstances and assist in regulating blood volume and the body's fluid balance. Plasma proteins contribute to colloid osmotic pressure, which is important in maintaining blood pressure. There are three general types of plasma proteins. The first is serum albumin, which is an essential factor in maintaining blood volume and pressure. The second is serum globulin, which is composed of three general fractions: the α fraction is associated with the transport of bilirubin, lipids, and steroids; the β fraction is associated with the transport of iron and copper in plasma; and the γ fraction contains the antibody molecules. Fibrinogen is the third major type of plasma protein. It is the inactive precursor of fibrin, which forms the framework of blood clots. Regulatory proteins, such as hormones and enzymes, are also present in the plasma. Diffusible nonorganic substances, such as sodium chloride, calcium, potassium, iodine, and iron, are used by body cells and constitute 0.9% of plasma. Diffusible organic constituents, such as urea, uric acid, xanthine, creatine, creatinine, and ammonia, are products of tissue metabolism that are transported from the tissues to the kidneys and skin for excretion. Also included in this category are nutritive organic materials, such as amino acids, glucose, fats, and cholesterol, which are foodstuffs in solution absorbed from the gastrointestinal (GI) tract. They are transported to other body tissues for utilization and storage (Table 13.1).

Cellular Components

The amounts of the different cellular components in the blood vary with age. Table 13.2 gives normal values from birth to 21 years.

TABLE 13.1 Organic and Inorganic Components of Blood

Constituent	Amount/Concentration	Major Functions
Water	92% of plasma weight	Medium for carrying all other constituents
Electrolytes	Total <1% of plasma weight	Keep H_2O in extracellular compartment; act as buffers; function in membrane excitability
Na^+	136–145 mEq/L (142 mM)	
K^+	3.5–5 mEq/L (4 mM)	
Ca^{2+}	4.5–5.5 mEq/L (2.5 mM)	
Mg^{2+}	1.5–2.5 mEq/L (1.5 mM)	
Cl^-	100–106 mEq/L (103 mM)	
HCO_3^-	27 mEq/L (27 mM)	
Phosphate (mostly HPO_4^{2-})	3–4.5 mEq/L (1 mM)	
SO_4^{2-}	0.5–1.5 mEq/L (0.5 mM)	
Proteins	6–8 g/dL (2.5 mM)	
Albumin	3.5–5.5 g/dL	Provides colloid osmotic pressure of plasma; acts as buffers; binds other plasma constituents (e.g., lipids, hormones, vitamins, metals)
Globulins	1.5–0.3 g/dL	Enzymes; enzyme precursors; antibodies (immune globulins); hormones
Fibrinogen	0.2–0.4 g/dL	Clotting factor
Gases, arterial plasma		
CO_2 content	22–30 mmol/L of plasma	By-product of metabolism; most CO_2 content is from HCO_3^- and acts as buffer
O_2	PaO_2, 80 mm Hg or greater (arterial); $P\bar{v}O_2$, 30–40 mm Hg (venous)	Oxygenation
N_2	0.9 mL/dL	By-product of protein catabolism
Nutrients		Provide nutrition and substances for tissue repair
Glucose and other carbohydrates	70–105 mg/dL (5.6 mM)	
Total amino acids	40 mg/dL (2 mM)	
Total lipids	450 mg/dL (7.5 mM)	
Cholesterol	150–250 mg/dL (4–7 mM)	
Individual vitamins	0.0001–2.5 mg/dL	
Individual trace elements	0.001–0.3 mg/dL	
Waste products		
Urea (BUN)	10–20 mg/dL (5.7 mM)	End product of protein catabolism
Creatinine	0.7–1.5 mg/dL (0.09 mM)	End product of energy metabolism
Uric acid	2.5–8 mg/dL (0.3 mM)	End product of protein metabolism
Bilirubin	0.3–1.1 mg/dL	End product of red blood cell destruction
Direct conjugated	0.1–0.5 mg/dL	
Indirect unconjugated	0.1–0.7 mg/dL	
Individual hormones	0.000001–0.05 mg/dL	Functions specific to target tissue

Adapted with permission from Vander AJ et al: *Human physiology: the mechanisms of body function,* ed 7, New York, 1998, McGraw-Hill.

TABLE 13.2 Age-Related Changes in Hematologic Values

Age	Hemoglobin (g)	Hematocrit (%)	RBC Count (millions/mm^3)	Platelets (thousands/mm^3)	Reticulocytes (%)
Birth	17.6	55	5.5	350.0	5.0
24 hr	18.0	56	5.3	400.0	5.2
1 wk	17.0	54	5.0	300.0	1.0
2 mo	12.4	30	4.3	260.0	0.5
6 mo	11.5	34	4.6	250.0	0.8
2 yr	12.9	40	4.8	250.0	1.0
6 yr	14.1	42	4.8	250.0	1.0
14 yr	15.0	M: 45; F: 42	5.1	250.0	1.0
21 yr	15.0	M: 45; F: 42	5.1	250.0	1.0

Data from Platt W: *Color atlas and textbook of hematology,* ed 2, Philadelphia, 1979, Lippincott, p 4. Reproduced by permission of William R. Platt, MD.

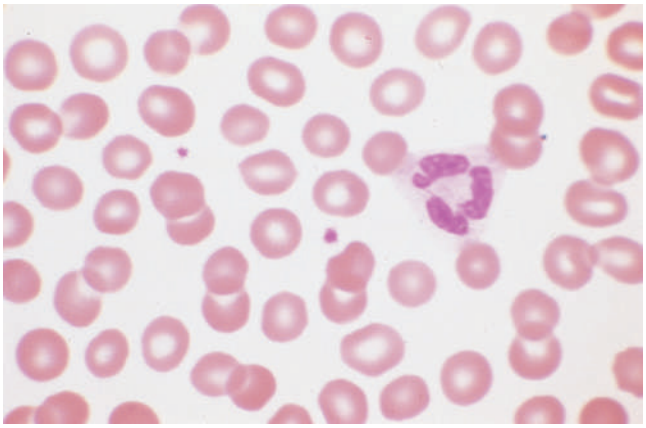


FIG 13.2 Mature erythrocytes. A mature neutrophil is also shown. (Courtesy Beth Payne, Sacred Heart Medical Center, Spokane, WA.)

Erythrocytes

Of the cellular elements of blood (Table 13.3), RBCs, or erythrocytes, are the most numerous, with normal concentrations ranging from 4.2 to 6.2 million cells/mm^3. RBCs are responsible for transporting oxygen to the tissues and participate in both removing carbon dioxide from the tissues and buffering blood pH. They have no cytoplasmic organelles, nucleus, mitochondria, or ribosomes. Therefore RBCs cannot synthesize protein or carry out oxidative reactions. Instead the erythrocyte's cytoplasm consists of a solution containing proteins, hemoglobin, and electrolytes that regulates diffusion through the cellular membrane. RBCs live for 80 to 120 days in the circulation; then they die and are replaced. Hemoglobin is the main functional constituent of the red cell. It is a protein that enables the blood to transport 100 times more oxygen than could be transported in plasma alone. An enzyme inside RBCs, carbonic anhydrase, is responsible for the buffering mechanism of red cells.

The erythrocyte's size and shape also contribute to its function as a gas carrier (Fig. 13.2). It is a small, biconcave disk (about 7.2 µm in diameter) that must circulate through splenic sinusoids and capillaries, which are only 2 µm in diameter. This remarkable feat is accomplished through a property called *reversible deformability,* which allows the RBC to assume a torpedo-like conformation and then return to a biconcave disk shape.

Leukocytes

White blood cells (WBCs), or leukocytes, protect the body by phagocytosis of microorganisms and other debris and participate in immune antibody formation. Leukocytes act primarily in the tissues, but are also transported in the circulatory and lymphatic systems. The average adult has approximately 5000 to 10,000 leukocytes per cubic millimeter of blood. Monocytes and granulocytes are WBCs that share a common lineage with RBCs and platelets. Because of the interrelationship of RBCs, WBCs, and platelets, which are all derived from the myeloid stem cell (Fig. 13.3), abnormalities in these cells are seen in some red cell diseases. Leukocyte structure and function are discussed in detail in Chapter 9.

Platelets

Platelets are essential in the formation of blood clots and in the control of bleeding. They are not cells but are circulating cytoplasmic fragments of megakaryocytes and are incapable of mitotic division. They contain cytoplasmic granules that release biochemical mediators involved in the hemostatic process. Normally, 150,000 to 400,000 platelets/mm^3 circulate freely in the blood. An additional one-third of the body's platelets are in a reserve pool in the spleen. The average life span of platelets in the peripheral blood is approximately 4 to 5 days.

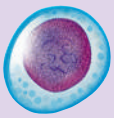
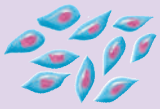
KEY POINTS

- Of the 4 to 6 L of blood in the circulatory system, approximately 45% is blood cells and 55% is plasma. The plasma fraction contains dissolved substances, including nutrients, ions, plasma proteins, metabolic wastes, hormones, and enzymes.
- Red cells function to carry oxygen and carbon dioxide in the blood. They have a limited life span of 80 to 120 days because they contain no cytoplasmic organelles and thus are incapable of replacing lost or damaged cellular components. The normal red cell concentration is 4.2 to 6.2 million cells/mm^3.
- Leukocytes, or WBCs, are the other cell type present in blood. Leukocytes circulate in much lower numbers than RBCs (5000 to 10,000/mm^3). Leukocytes are important mediators of immunity.
- Platelets are not cells but are small fragments of megakaryocytes. The normal platelet count is 150,000 to 400,000 cells/mm^3.

STRUCTURE AND FUNCTION OF RED BLOOD CELLS

The cellular components of blood originate in the yolk sac mesenchyme, move to the liver and spleen during fetal life, and finally are limited to the marrow of the body skeleton (Fig. 13.4). Bone marrow provides a special environment for hematopoietic cell proliferation and maturation. Developing cells are held in a fine reticular meshwork, which provides free access to plasma nutrients but retains developing cells until their maturity allows penetration of the endothelial barrier. In times of need,

TABLE 13.3 Characteristics of Blood Cells

Cell	Structural Characteristics	Normal Amounts in Circulating Blood*	Function	Life Span
Erythrocyte (red blood cell)	Nonnucleated biconcave disk containing hemoglobin	Males: $4.7–6.2 \times 10^{12}$/L Females: $4.2–5.4 \times 10^{12}$/L	Gas transport to and from tissue cells and lungs	80–120 days
Leukocyte (white blood cell)	Nucleated cell	$5.0–10 \times 10^{9}$/L	Body defense mechanisms	See later
Lymphocyte	Mononuclear immunocyte	$1.2–3.4 \times 10^{9}$/L; 20%–44% leukocyte differential	Humoral and cell-mediated immunity	Days or years, depending on type
Neutrophil	Segmented polymorphonuclear granulocyte with neutrophilic granules	$1.4–6.5 \times 10^{9}$/L; 50%–70% leukocyte differential	Phagocytosis, particularly during early phase of inflammation	5 days
Eosinophil	Segmented polymorphonuclear granulocyte with eosinophilic granules	$0–0.7 \times 10^{9}$/L; 0%–4% leukocyte differential	Phagocytosis, antibody-mediated defense against parasites; participates in mucosal immune response	Unknown
Basophil 	Segmented polymorphonuclear granulocyte with basophilic granules	$0–0.2 \times 10^{9}$/L; 0%–2% leukocyte differential	Transport and release of heparin and histamine; involved in immune and inflammatory responses	Unknown
Monocyte-macrophage	Large mononuclear phagocyte	$0.11–0.59 \times 10^{9}$/L; 2%–9% leukocyte differential	Phagocytosis; process and present antigens	Months to years
Platelet 	Discoid cytoplasmic fragment derived from megakaryocytes	$150–400 \times 10^{9}$/L	Hemostasis after vascular injury; forms hemostatic plug, provides cofactors, maintains vascular endothelium	4–5 days

Illustrations from Patton KT, Thibodeau GA: *Anatomy & physiology,* ed 8, St Louis, MO, 2013, Mosby, p. 606.
*Given in SI units.

immature cells (reticulocytes and nucleated red blood cells, or NRBCs) are released early into the circulation; their presence in increased numbers is a sign that the hematopoietic system is stressed or is experiencing disease.

Hematopoiesis

Hematopoiesis is the developmental process leading from pluripotential stem cells to mature, differentiated red cells, neutrophils, eosinophils, basophils, monocytes, and platelets. Lymphopoiesis describes this process for lymphocytes. Both hematopoietic and lymphopoietic stem cells probably derive from a single totipotent stem cell pool in fetal development, but it is uncertain if this is the functioning stem cell after birth (Fig. 13.5). Research suggests that a pluripotential stem cell that is stimulated by erythropoietin and other poietins to cause further differentiation into separate cell lines may be the primary stem cell in adults.

Hematopoiesis is a two-stage process that involves mitotic division (proliferation) and maturation (differentiation). Each type of blood cell has stem cells that undergo mitosis when stimulated by a specific biochemical signal, indicating that the number of circulating cells

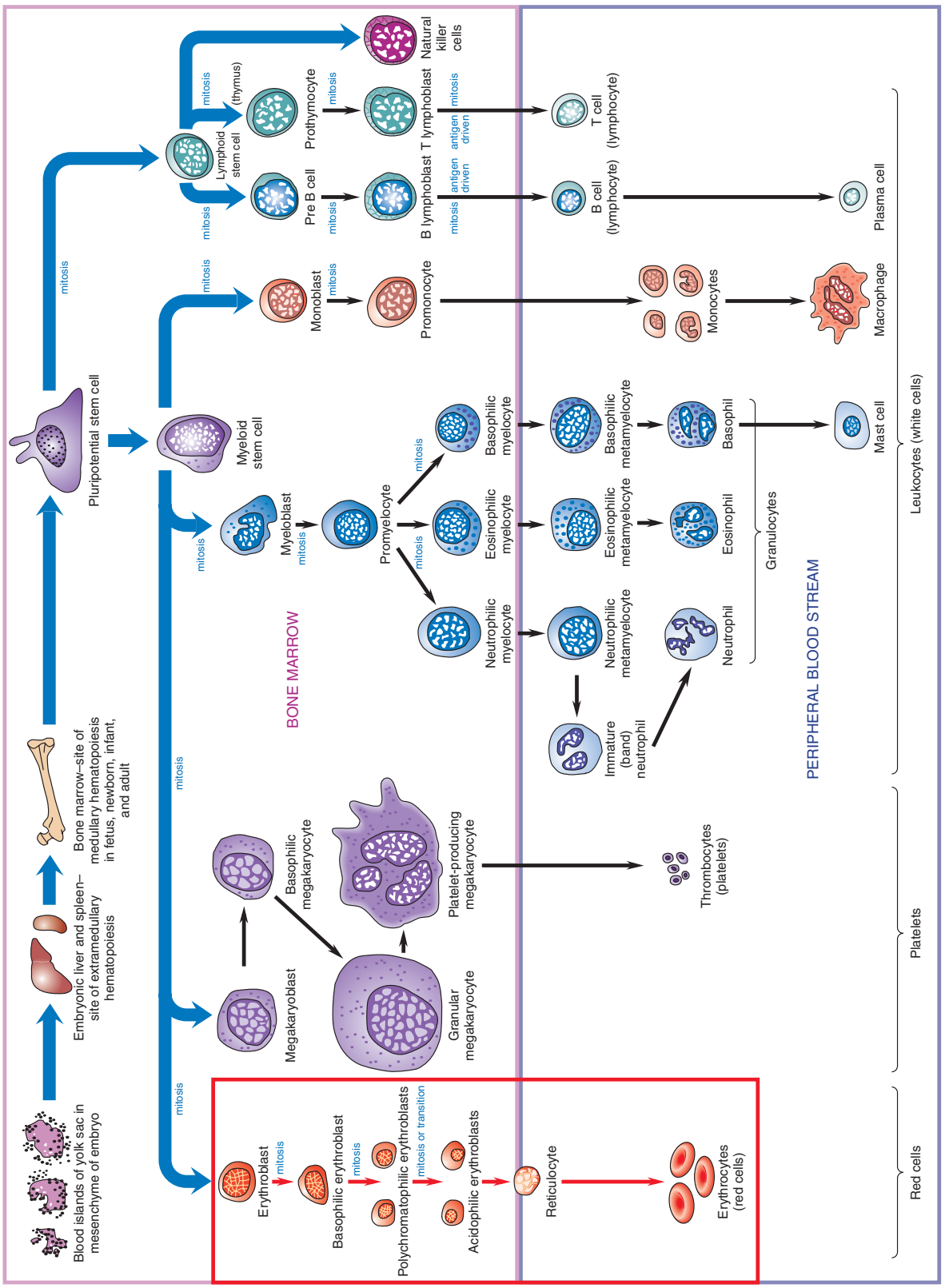


FIG 13.3 Maturation of human blood cells. Probable pathways of blood cell differentiation from the pluripotential stem cell to mature leukocytes, erythrocytes, and platelets. Production of cells begins in embryo blood islands of the yolk sac. As the embryo matures, production shifts to the liver and spleen (extramedullary hematopoiesis) and progresses to bone marrow (medullary hematopoiesis). In an adult, all production is in the bone marrow. Current thinking is that all cell production begins with a pluripotential stem cell, which differentiates into either a myeloid stem cell or a lymphoid stem cell, which then differentiates into a specific blast cell. For example, red cell differentiation begins with the proerythroblast, which matures into a basophilic erythroblast, to a polychromatophilic erythroblast, and to an acidophilic erythroblast, all of which are found in the bone marrow. Red cell differentiation concludes with production of reticulocytes and mature red cells (erythrocytes), which normally are found only in the peripheral blood.

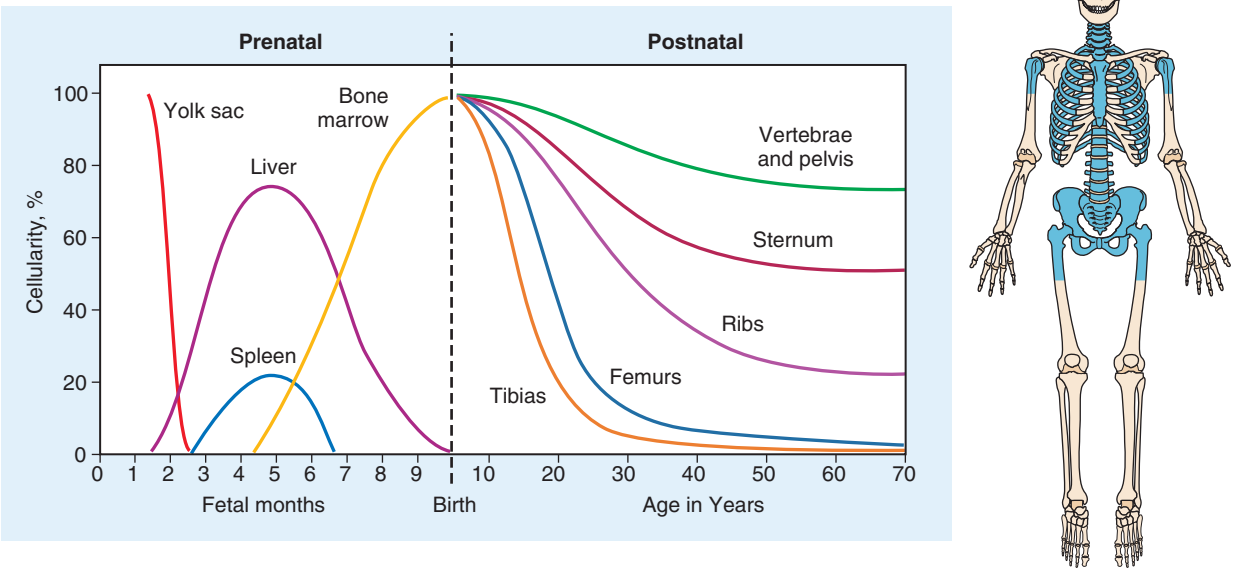


FIG 13.4 Location of active marrow growth in the fetus and adult. During fetal development, hematopoiesis is first established in the yolk sac mesenchyme, later moves to the liver and spleen, and finally is limited to the bony skeleton. From infancy to adulthood, there is progressive restriction of productive marrow to the axial skeleton and proximal ends of the long bones, which appear as shaded areas on the drawing of the skeleton.

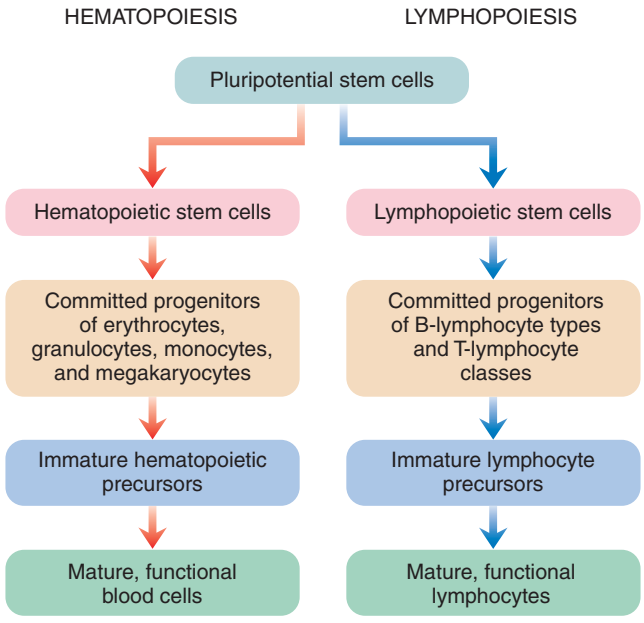


FIG 13.5 Stem cells and normal hematopoiesis.

has decreased. Medullary or bone marrow hematopoiesis continues throughout life and can be accelerated by several mechanisms, including (1) an increase in differentiation of daughter cells, (2) an increase in number of stem cells, and (3) conversion of yellow (fatty) bone marrow (which does not produce cells) to red marrow (which does produce cells). Marrow conversion is stimulated by erythropoietin, which is the hormone from the kidney that stimulates erythrocyte production. In adults, extramedullary hematopoiesis, or production of blood cells in tissue other than bone, is usually due to disease.

Erythrocyte development is shown in detail in Fig. 13.3. During this process, the cell changes from a large nucleated cell, rich in ribosomes,

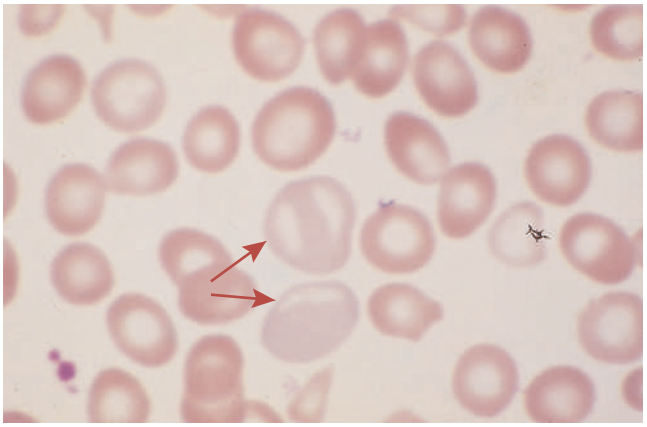


FIG 13.6 Reticulocytes seen on peripheral blood smear. The two reticulocytes in the center *(arrows)* still contain remnants of intracellular organelles. (Courtesy Beth Payne, Sacred Heart Medical Center, Spokane, WA.)

to a reticulocyte, which is a small disk that has lost its nucleus. The reticulocyte (Fig. 13.6) leaves the marrow, enters the bloodstream, and matures into an erythrocyte in 24 to 48 hours. During this period, mitochondria and ribosomes disappear; the cell can no longer synthesize hemoglobin, and it relies on glycolysis for adenosine triphosphate (ATP) production. The normal reticulocyte count is 1% of the total RBC count. This makes it a useful test to determine effective erythropoietic activity because erythropoietin stimulates uncommitted stem cells to differentiate into proerythroblasts.

Hemoglobin Synthesis

The immature red cell can be viewed as a factory for hemoglobin synthesis. In a mature red cell, hemoglobin, the oxygen-carrying protein, composes about 90% of the cell's dry weight in the form of approximately 300 hemoglobin molecules. Hemoglobin that is carrying oxygen is called

oxyhemoglobin. Hemoglobin is composed of two pairs of polypeptide chains—the globins. Each globin has an attached heme molecule that is composed of iron plus a protoporphyrin molecule (Fig. 13.7).

After dietary iron is absorbed in the duodenum and proximal jejunum, it is transported through the plasma by the protein transferrin to transferrin iron receptors on the RBC membrane. The transferrin-receptor complex is engulfed by the cell into an invagination of the cell surface. The invagination becomes sealed off and forms an intracytoplasmic vacuole. Iron is then released and either stored as ferritin or used to synthesize heme (Fig. 13.8). About 67% of total body iron is bound to heme in erythrocytes and muscle cells, and 30% is stored bound to ferritin or hemosiderin-containing macrophages and hepatic parenchymal cells. The remaining 3% is lost daily in urine, sweat, bile, and epithelial cells that are shed in the intestines. The mitochondria are responsible for the synthesis of protoporphyrin. The final heme molecule consists of four porphyrin moieties assembled in a ring structure around a central iron molecule.

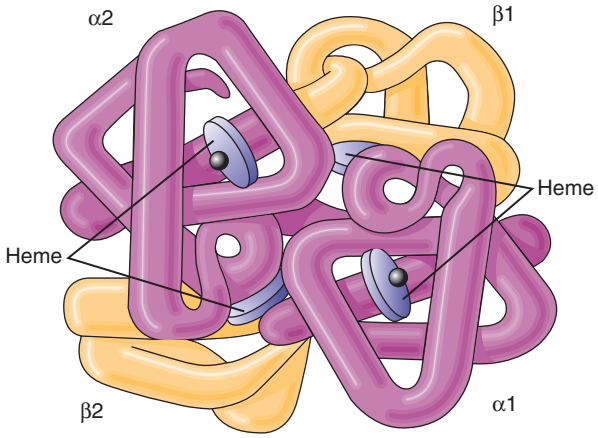


FIG 13.7 Molecular structure of hemoglobin. The molecule is a spherical tetramer weighing approximately 64,500 daltons. It contains two α- and two β-polypeptide chains and four heme groups.

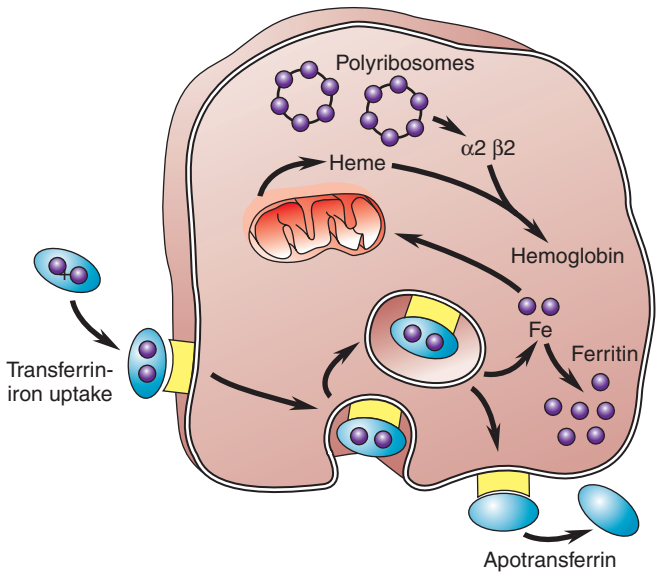


FIG 13.8 Intracellular pathways for iron uptake and incorporation into hemoglobin in erythroblasts in the bone marrow. The iron–transferrin complex is picked up by a membrane-associated receptor and brought into the cell by invagination and formation of an intracytoplasmic vacuole. The iron is then released and stored as intracytoplasmic ferritin or used to synthesize heme, the precursor of hemoglobin. The transferrin–receptor complex is returned to the cell membrane, where the apotransferrin is expelled back into the circulation. (Redrawn from Hillman RS, Finch CA, editors: *Red cell manual,* ed 6, Philadelphia, 1992, FA Davis, p 8.)

Hemoglobin Synthesis in Infants

When an infant is born, many mechanisms occur to decrease the infant's hemoglobin level. At birth, erythropoietin, a hormone that stimulates RBC production, disappears from blood plasma; there is an increase in arterial oxygen saturation; and the infant is born with immature bone marrow. These mechanisms cause a slow rate of RBC production at birth. In addition, infants primarily have fetal hemoglobin, accounting for about 70% of their total hemoglobin. Although fetal hemoglobin is a more efficient oxygen carrier, it has a shorter life span than adult hemoglobin. This causes RBCs to be turned over every 70 to 90 days, instead of 120 days for adult RBCs. Fetal hemoglobin may suppress production of erythropoietin.

Hemoglobin levels gradually decrease in the infant over the first 2 to 3 months because of the rapid destruction of fetal hemoglobin, decreased RBC production and depressed erythropoietin production. Additionally, the infant experiences rapid growth during this time, which creates quick expansion of blood volume that further dilutes the supply of hemoglobin. Maternal iron stores are rapidly depleting at this time, and the baby's iron stores will gradually diminish by 6 months.

As fetal hemoglobin is metabolized, the iron is released and stored. The body has enough iron to synthesize hemoglobin, but there is no stimulation to create hemoglobin at this time. Hemoglobin levels will continue to decrease until the oxygen needs of the tissues in the body are depleted enough to stimulate erythropoietin production. Release of erythropoietin causes erythropoiesis to resume. Adult hemoglobin is made at this time with the iron stored in the body. Hemoglobin levels will steadily increase in the infant starting about 6 months of age as fetal hemoglobin is replaced by adult hemoglobin.

Globin is assembled from two pairs of polypeptide chains produced on specific ribosomes. The protein chain produced in fetal life is altered after birth by sequential gene suppression and activation. At birth, red cells contain mainly fetal hemoglobin (hemoglobin F), which is composed of two α chains and two γ chains. Hemoglobin F is a more efficient gas carrier under decreased oxygen tension than hemoglobin A and releases CO_2 more readily. Within 120 days, fetal hemoglobin disappears and is replaced by adult hemoglobin (hemoglobin A) (Fig. 13.9). This switch is the result of globin genes and is not well understood. Hemoglobin A is composed of two α chains and two β chains and constitutes 97% of the hemoglobin found in adults. Hemoglobin A_2 comprises 2% to 3% of hemoglobin found in adults and is composed of $\alpha_2\delta_2$.

Several hundred hemoglobinopathies have been described that have changes in the two α chains or two β chains. Most are characterized by the substitution of only one amino acid and are classified by the polypeptide chain in which the substitution occurs.

Nutritional Requirements for Erythropoiesis

In addition to iron, which is required for hemoglobin synthesis, the normal development of erythrocytes requires adequate supplies of protein, vitamins, and minerals. Erythropoiesis cannot proceed in the absence of vitamins, especially B_{12}, folate, B_6, riboflavin, pantothenic acid, niacin, ascorbic acid, and vitamin E. Folates and vitamin B_{12} (cobalamin) are absorbed from food by the ileal mucosa. Folate deficiencies or vitamin B_{12} deficiencies lead to impaired DNA synthesis in erythroid cells because the vitamins are coenzymes in a large number of key reactions in cellular metabolism. Absorption of vitamin B_{12}

requires intrinsic factor in the gastric juice. Intrinsic factor is secreted by the stomach parietal cells and binds to vitamin B_{12}. The complex then moves down the GI tract to the ileum, where it attaches to specific receptor sites on the ileum mucosal cell. It is absorbed into the cell, released, and transported in the blood to the tissues and liver.

Energy and Maintenance of Erythrocytes

For the RBC to perform efficiently and survive in the circulation for the full 120-day life span, it must have a source of energy. Without an energy source, ion pumps fail and the RBC becomes sodium logged and potassium depleted. The shape changes from a biconcave disk to a sphere, and it is quickly removed from the circulation by the filtering action of the spleen and the mononuclear phagocyte system. The metabolism of the RBC is limited because of the absence of a nucleus, mitochondria, and other subcellular organelles. Although the binding, transport, and release of O_2 and CO_2 is a passive process that does not require energy, other energy-dependent metabolic processes occur that are essential to RBC viability. The chief metabolic pathway, accounting for about 90% of the glucose used, is the anaerobic or Embden–Meyerhof pathway. The Embden–Meyerhof pathway provides ATP for regulation of intracellular Na^+, K^+, Ca^{2+}, and Mg^{2+} concentrations via cation pumps. About 10% of the glucose undergoes aerobic glycolysis in the hexose monophosphate shunt. The hexose monophosphate shunt provides nicotinamide adenine dinucleotide phosphate (NADPH) and glutathione (GSH) to reduce cellular oxidants. This protects the cell from permanent oxidant injury. The methemoglobin reductase pathway protects hemoglobin from oxidation via nicotinamide adenine dinucleotide (NADH) and methemoglobin reductase. Last, the Rapoport–Luebering pathway forms 2,3-diphosphoglycerate (2,3-DPG), which facilitates oxygen release to the tissues. These pathways contribute energy for maintaining (1) high intracellular K^+, low intracellular Na^+, and very low intracellular Ca^{2+} levels (cation pumps); (2) reduced hemoglobin concentration; (3) high levels of reduced GSH; and (4) membrane integrity and deformability. Deficiencies of enzymes that regulate these pathways can be due to natural causes, such as the normal aging process, or to an inherited deficiency of an enzyme.

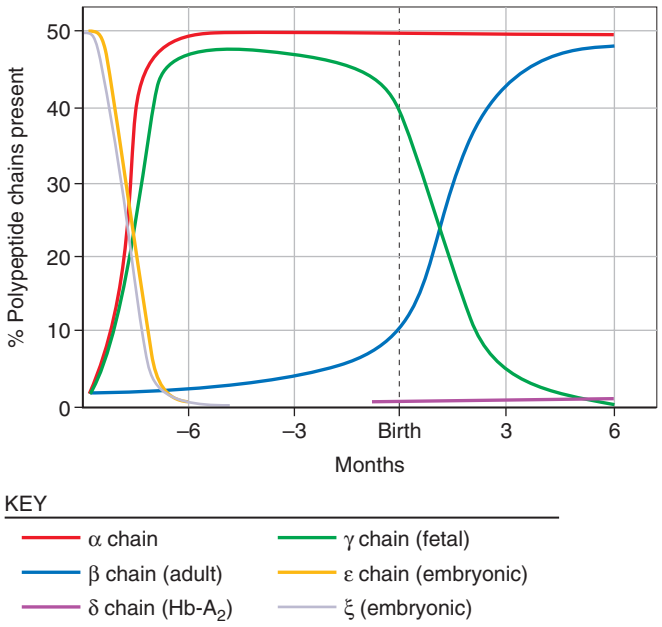


FIG 13.9 Changes in hemoglobin with development. Sequential suppression and activation of individual globin genes in the immediate postnatal period result in a switch from fetal hemoglobin (hemoglobin F: two α chains and two γ chains) to adult hemoglobin (hemoglobin A: two α chains and two β chains). A small amount of hemoglobin A2 (two α chains and two δ chains) is also present in the adult. (From Young NS et al, editors: *Clinical hematology,* Philadelphia, 2006, Mosby, p 24.)

Red cell membrane structures are matrices formed from a double layer of phospholipids. In the red cell membrane, the globular proteins floating on the "sea of lipids" form a protein network on the cytoplasmic surface of the membrane. Half of the mass of the membrane is lipid, which is partially responsible for many of its physical characteristics. Both passive cation permeability and mechanical flexibility can be significantly influenced by changing the lipid composition of the membrane. Maintenance and renewal of membrane lipids in well-developed RBCs is important, and problems in these pathways result in premature cell death.

Red Cell Production

When blood is described as a single body system, it is called the erythron (Fig. 13.10). The erythron includes the blood cells and their bone marrow

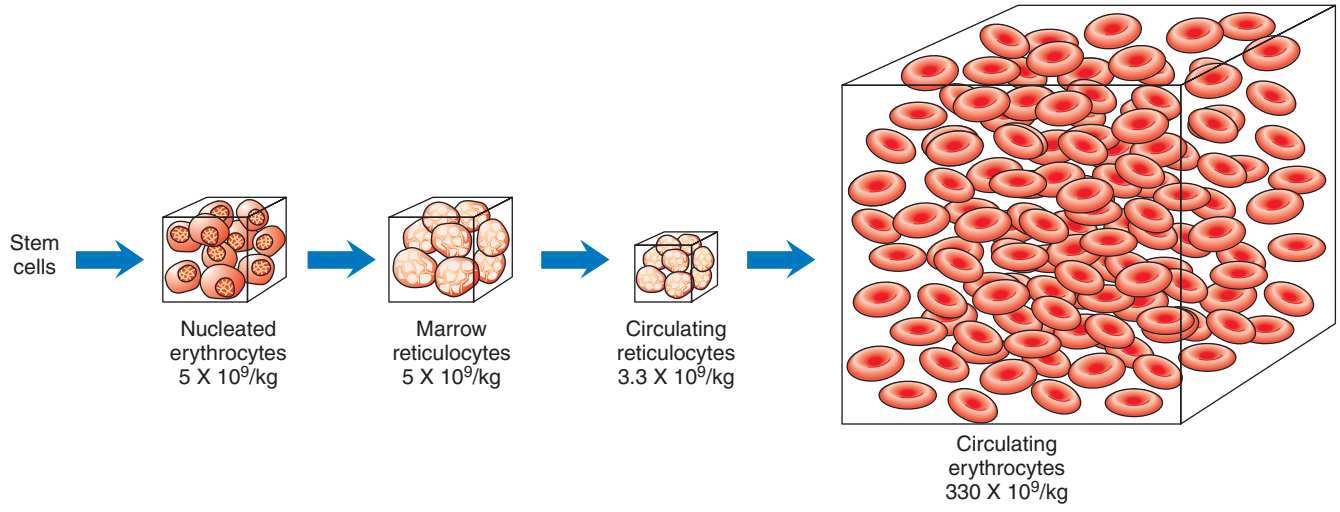


FIG 13.10 Scale model of the erythron, showing the relative proportions of each component. The numbers below each box indicate the average number of cells per kilogram of body weight. (Redrawn from Wintrobe M et al, editors: *Clinical hematology,* ed 8, Philadelphia, 1981, Lea & Febiger, p 109.)

precursors. The size of the erythron increases or decreases based on the erythropoietic process and the pathologic changes in red cells seen in anemia. Erythropoiesis is controlled by a system sensitive to alterations in the concentration of hemoglobin in the blood. A decrease in hemoglobin level decreases the tissue oxygen tension in the kidney. In response to this hypoxia, the kidney secretes a hormone, erythropoietin, that stimulates primitive stem cells in the bone marrow to differentiate into proerythroblasts or pronormoblasts, thereby increasing the erythron (Fig. 13.11). Hypoxia from a low hemoglobin saturation level and other causes, such as high altitudes, can also initiate this response.

Red Cell Destruction

As the red cell ages, the various enzyme activities decrease, amounts of membrane lipids decrease, levels of hemoglobin A_2 and methemoglobin increase, and changes in cell size occur. Methemoglobin is formed when the iron of the hemoglobin molecule is oxidized to the ferric state (Fe^{3+}). The cell loses its ability to deform and becomes increasingly fragile. These aging red cells are then removed by the mononuclear phagocytic system. The red cells are digested by proteolytic and lipolytic enzymes in phagolysosomes of macrophages. Almost 80% to 90% of this process occurs in macrophages of the spleen and liver. Only 10% to 20% of normal destruction occurs intravascularly.

Globin is broken down into amino acids, and the iron is recycled. Porphyrin is reduced to bilirubin, which is transported to the liver and conjugated by the enzyme glucuronyl transferase. Finally, conjugated bilirubin is excreted in the bile as glucuronide. Bacteria in the intestine convert conjugated bilirubin into urobilinogen, which is excreted primarily in the stool but also in the urine (Fig. 13.12). Any condition causing increased red cell destruction increases the load of bilirubin to be cleared, which leads to increased serum levels of unconjugated bilirubin and increased excretion of urobilinogen. Increased levels of circulating bilirubin give the skin a yellowish tone, which is called *jaundice*. In newborns, the albumin levels for bilirubin transport are low and the amount of liver glucuronidase available for bilirubin conjugation is low, which may cause an accumulation of toxic unconjugated bilirubin. Unconjugated bilirubin is toxic because in this form it is lipid soluble and can easily cross cell membranes. This form of bilirubin has a high affinity for basal ganglia of the central nervous system. The conjugated form of bilirubin is water soluble but lipid insoluble, so it cannot cross cell membranes.

KEY POINTS

- Red cell development from pluripotential stem cells in the bone marrow is stimulated by a hormone growth factor called *erythropoietin*. Erythropoietin is secreted into the bloodstream by kidney cells in response to low oxygen tension in the blood.
- During development, red cells lose their nuclei and other cytoplasmic organelles. A reticulocyte is an immature red cell that still retains some cellular organelles. An increased blood reticulocyte count is a useful indicator of increased red cell production.
- Hemoglobin is the major component of red cells. It is composed of two pairs of polypeptide chains, each of which has a heme molecule attached. Oxygen can bind reversibly to an iron molecule at the center of each heme. When fully saturated, a hemoglobin molecule carries four oxygen molecules and is referred to as *oxyhemoglobin*.
- Red cell production requires adequate amounts of several nutrients, particularly iron, vitamin B_{12}, and folate. Lack of intrinsic factor inhibits absorption of B_{12} from the small intestine and is a risk factor for anemia.
- Red cells rely on glycolysis for energy production because they do not contain mitochondria. As energy production declines because of red cell aging and loss of essential glycolytic enzymes, the cell swells, is trapped in the spleen, and is removed from the circulation. Red cell degradation releases bilirubin, a toxic substance that is conjugated in the liver and excreted in urine and bile.

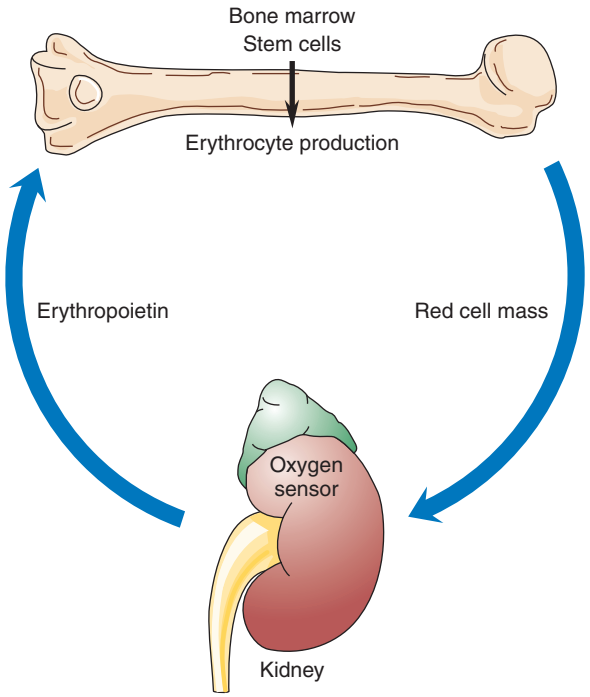


FIG 13.11 Feedback circuit illustrating the role of erythropoietin in the regulation of red cell mass.

GAS TRANSPORT AND ACID–BASE BALANCE

RBCs have many important functions in the body related to gas transport and acid–base balance. RBCs contain hemoglobin, which is responsible for oxygen transport to the body tissues. Oxygen combines with the heme portion of hemoglobin to form oxyhemoglobin in a reversible bond in the pulmonary capillary attributable to a high partial pressure of oxygen (P_{O_2}) and is carried to the tissues, where it is released. Large quantities of carbonic anhydrase in RBCs catalyze the reaction between CO_2 (produced by cellular metabolism in the tissues) and water to form carbonic acid, which dissociates into hydrogen and bicarbonate ions for elimination by the lungs and kidneys. Approximately 90% of the CO_2 in the arterial blood and 60% of the CO_2 in the venous blood are transported as bicarbonate. Finally, the hemoglobin protein directly binds with the remaining CO_2 to form carbaminohemoglobin for CO_2 transport. Carbamino compounds are acid–base buffers responsible for as much as 50% of the whole blood–buffering power.

Oxygen Transport

Transport of oxygen to the body tissues and removal of carbon dioxide is a complex process involving interdependent functioning of the lungs, heart, and blood (Fig. 13.13). Approximately 97% of oxygen in the blood is transported on red cells reversibly combined with hemoglobin (oxyhemoglobin), and 3% is dissolved in plasma. Each hemoglobin molecule can bind four atoms of oxygen. Despite a combining potential of 1.39 mL of oxygen per gram of hemoglobin in pure hemoglobin, a maximum of about 1.34 mL of oxygen per gram of hemoglobin is

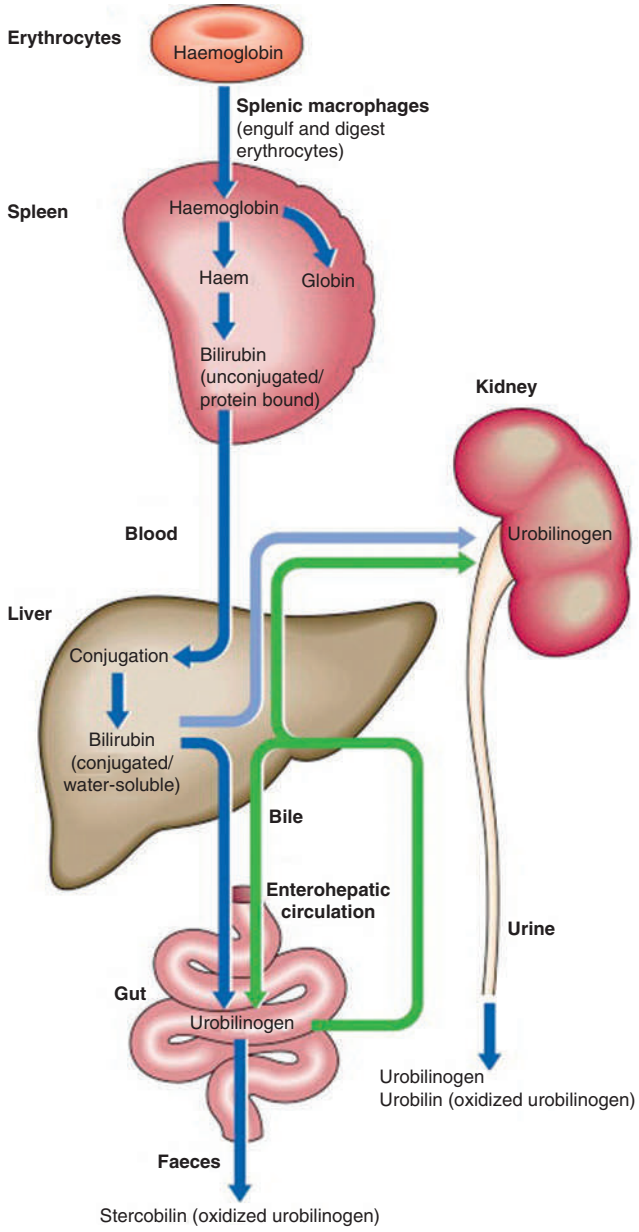


FIG 13.12 Most hemoglobin degradation occurs in the macrophages of the spleen. The globin and iron portions are conserved and reused. Heme is reduced to bilirubin, eventually degraded to urobilinogen, and excreted in the feces. Thus indirect indicators of erythrocyte or erythrocyte destruction include the blood bilirubin level and urobilinogen concentration in the feces. (From Cross SS: *Underwood's pathology: a clinical approach*, ed 6, Churchill Livingstone, 2013, Elsevier Ltd.)

available due to a reduction of about 4% by impurities such as methemoglobin. The blood of an average person contains approximately 15 g of hemoglobin per 100 mL of blood. Therefore in the average person, the hemoglobin in 100 mL of blood can combine with approximately 20 mL of oxygen if the hemoglobin is 100% saturated. This value is expressed as 20 vol%.

The partial pressure of oxygen (PO_2) reflects the pressure or tension that oxygen exerts when it is dissolved in blood. Partial pressure is measured in millimeters of mercury (mm Hg). In the pulmonary capillaries, where PO_2 is high, oxygen binds efficiently with hemoglobin, but in the tissue capillaries, where PO_2 is low, oxygen is released from hemoglobin. The partial pressure affects the tendency of oxygen to bind with hemoglobin. The partial pressure of oxygen in arterial blood (PaO_2) is usually 80 to 100 mm Hg, whereas the partial pressure of oxygen in venous blood (PvO_2) is usually 35 to 40 mm Hg. The amount of hemoglobin bound to oxygen relative to the total amount of hemoglobin is expressed as the oxygen saturation and is given in a percentage. Saturation of arterial blood with oxygen (SaO_2) is normally 95% to 100%, whereas that of venous blood (SvO_2) is 60% to 80%.

The oxygen–hemoglobin dissociation curve (Fig. 13.14) describes the relationship between PO_2 and SO_2. The upper part of the curve represents oxygen uptake in the lungs and demonstrates that significant changes in PO_2 result in only small changes in SO_2 to help ensure adequate oxygen delivery to the tissues. On the steep lower portion of the curve, reflecting the venous blood, small changes in venous PO_2 result in large changes in SvO_2. Therefore the tissues are protected with an available oxygen reserve as large quantities of oxygen are released from the blood for relatively small decreases in PO_2. Normally, tissue PO_2 does not rise above 40 mm Hg to enhance diffusion of oxygen from the blood to the tissues. The strength of the bond between hemoglobin and oxygen is called the *oxygen–hemoglobin affinity.* For any given PO_2, hemoglobin saturation will be higher when affinity is increased and saturation will be lower when affinity is decreased. Changes in hemoglobin affinity are represented by shifts in the oxyhemoglobin dissociation curve (see Fig. 13.14). Shifts in the oxyhemoglobin curve affect the ability of hemoglobin to bind O_2 in the lungs and release it in the tissues. The ability of hemoglobin to release oxygen to the tissues is commonly assessed at point P50 on the oxygen–hemoglobin dissociation curve. The P_{50} is the PO_2 at which 50% of the hemoglobin is saturated. A decrease in oxygen affinity (shift to the right on the oxyhemoglobin dissociation curve) or an increase in oxygen affinity (shift to the left) can be caused by the conditions listed in Fig. 13.14.

A shift of the oxyhemoglobin dissociation curve attributable to changes in the blood levels of PCO_2 and the H^+ concentration is important to enhance oxygen uptake by the blood in the lungs and the release of oxygen from the hemoglobin to the body tissues. This is called the *Bohr effect.*

A shift of the oxyhemoglobin dissociation curve to the right enhances oxygen release to the cell. The shift provides the increase in oxygen delivery that is needed during exercise and other types of stress, as well as in chronic disease states. A shift of the oxyhemoglobin dissociation curve to the left is seen with a decrease in H^+ ion concentration, with a decrease in PCO_2, with an increase in pH, with a decrease in temperature, with a decrease in 2,3-DPG concentration, in some congenital hemoglobinopathies, and with increased carboxyhemoglobin concentration.

Another important factor affecting O_2 delivery to tissues is the arterial oxygen content (CaO_2). Arterial blood oxygen content (CaO_2) and venous blood oxygen content (CvO_2) can be calculated by adding the amount of oxygen combined with hemoglobin and the amount of oxygen dissolved in plasma (Table 13.4).

Oxygen delivery, or ($\dot{D}O_2$), is the amount of oxygen (in milliliters) delivered per minute to the tissues. It is calculated by multiplying the arterial oxygen content (CaO_2) by the cardiac output (CO). CO is usually between 4 and 8 L/min. Therefore oxygen delivery is approximately 1000 mL/min.

Oxygen consumption ($\dot{V}O_2$) is the amount of oxygen consumed by the tissues and is measured in milliliters of oxygen per minute. Once the oxygen reaches the tissues, oxygen consumption is controlled by the rate of energy expenditure within the cells, or the rate at which adenosine diphosphate (ADP) is formed from ATP to provide energy. The increasing concentration of ADP enhances the metabolic utilization of oxygen. Oxygen consumption can be determined by subtracting the oxygen remaining in the venous blood (CvO_2) from the oxygen delivered

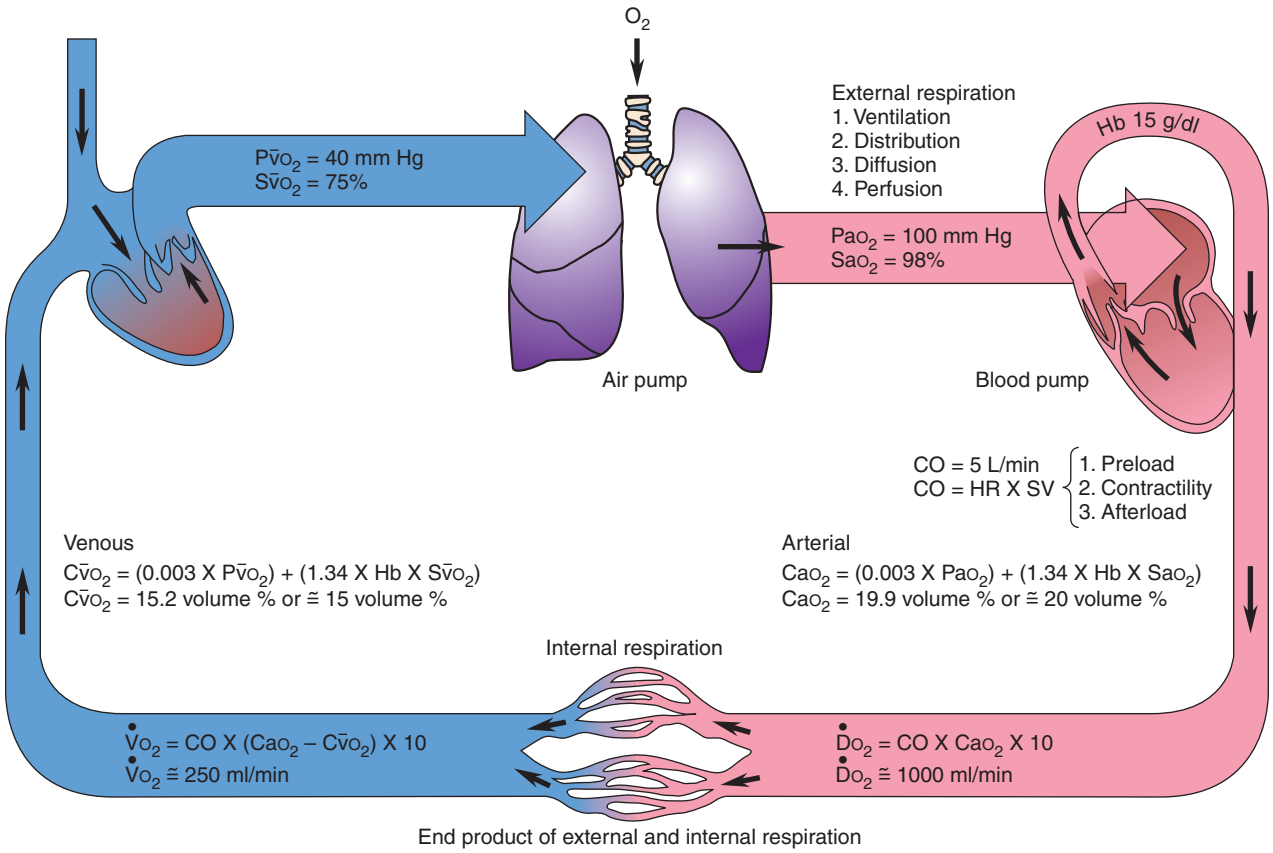


FIG 13.13 Oxygen transport. Diffusion of oxygen into the pulmonary capillaries occurs when alveolar PO_2 exceeds capillary PO_2. Maintenance of this gradient is dependent on adequate alveolar ventilation and perfusion. Delivery of oxygenated blood to the tissues *($\dot{D}O_2$)* is determined by the content of oxygen in the blood (CaO_2) and the cardiac output (CO). The difference between arterial and venous oxygen is a reflection of oxygen consumption by tissues *($\dot{V}O_2$). Hb,* Hemoglobin; *HR,* heart rate; *SV,* stroke volume.

to the tissues by the arteries (CaO_2), and is known as the *Fick equation.*

Gas values (pressure and content) relative to the oxygenation of blood are summarized in Table 13.4.

Carbon Dioxide Transport

RBCs are also important in the transport of carbon dioxide in the blood. Carbon dioxide, a by-product of cellular metabolism, is transported in three forms in the blood: (1) as dissolved gas, (2) as bicarbonate ion (HCO_3^-), and (3) in association with hemoglobin (Fig. 13.15). The partial pressure of carbon dioxide (PCO_2) reflects the pressure or tension that carbon dioxide exerts when it is dissolved in the blood. Partial pressure is measured in millimeters of mercury (mm Hg). In the pulmonary capillaries, carbon dioxide easily dissociates from hemoglobin and then diffuses across the alveolar membrane into the alveolar sacs. In the body tissues, the carbon dioxide inside the cells diffuses into the blood and attaches to the hemoglobin as oxygen is released to the tissues. The partial pressure of carbon dioxide in the arterial blood ($PaCO_2$) is usually 40 mm Hg and in the venous blood ($PvCO_2$) it is usually 45 mm Hg. These differences are much smaller than those of oxygen, but carbon dioxide diffuses much more readily. Dissolved carbon dioxide combines slowly with water in the plasma to form carbonic acid (H_2CO_3), but in the red cell the presence of carbonic anhydrase acting as a catalyst significantly accelerates this reaction. Carbonic acid rapidly dissociates into hydrogen ions (H^+) and bicarbonate ions (HCO_3^-). As the concentration of HCO_3^- in the red cell increases, it diffuses into the plasma, whereas the H^+ remains. This increase in intracellular cation concentration causes the anion chloride to diffuse from the plasma into the red cell to maintain electrical neutrality. This is referred to as the chloride shift.

Hemoglobin provides an excellent acid–base buffer by reacting with the free hydrogen ions and directly with carbon dioxide to form carbaminohemoglobin ($HbCO_2$), which is easily dissociated in the lungs to yield free carbon dioxide for exhalation. Unloading of oxygen in the tissue facilitates the loading of carbon dioxide and is referred to as the *Haldane effect.*

Alterations in Oxygen Transport

There must be sufficient circulating hemoglobin mass to meet the metabolic needs of the body. A feedback mechanism ensures that when the amount of oxygen reaching the tissues decreases, a compensatory increase occurs in the production of red cells. The feedback mechanism regulating RBC production is under the control of erythropoietin (Fig. 13.16).

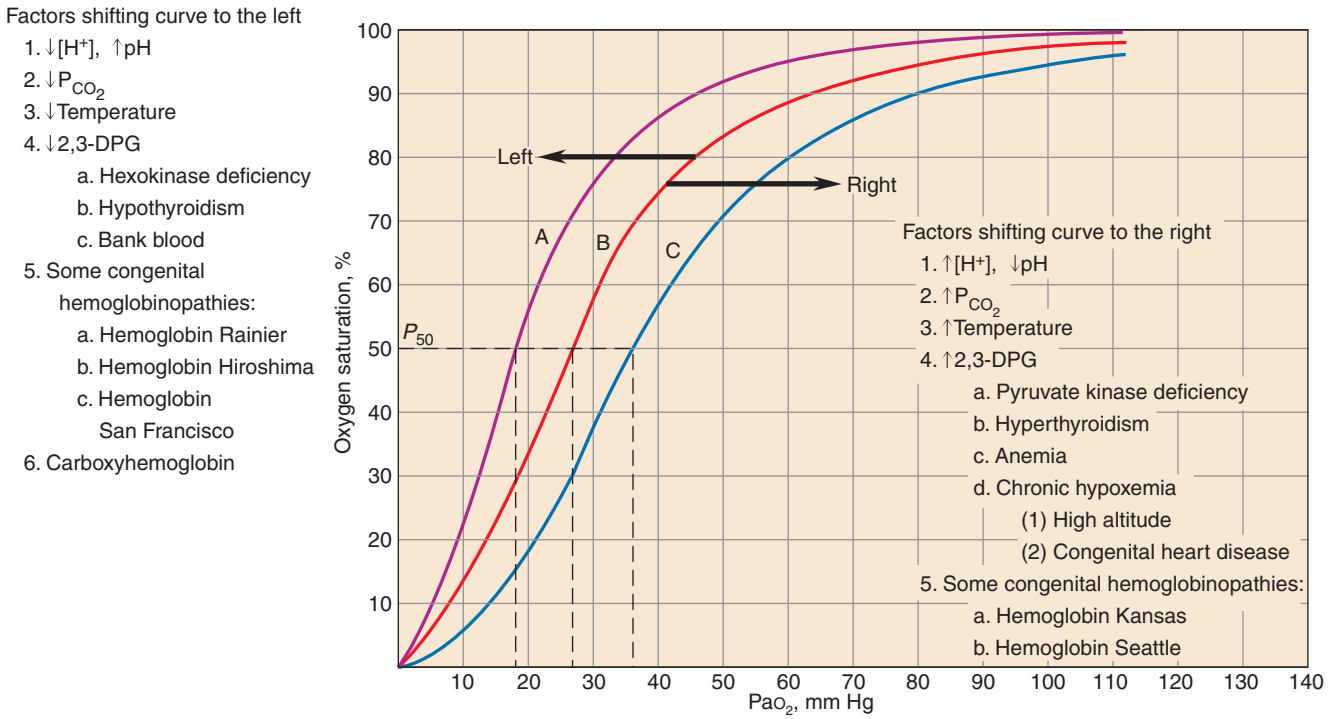


FIG 13.14 Oxygen–hemoglobin dissociation curve: factors affecting hemoglobin's affinity for oxygen. Curve B is the standard oxyhemoglobin dissociation curve. Factors that shift the curve to the left are represented in curve A; factors that shift the curve to the right are represented in curve C. *2,3-DPG,* 2,3-Diphosphoglycerate. (Redrawn from Gottlieb JE: Breathing and gas exchange. In Kinney MR et al, editors: *AACN's clinical reference for critical care nursing,* ed 3, New York, 1993, McGraw-Hill, p 672.)

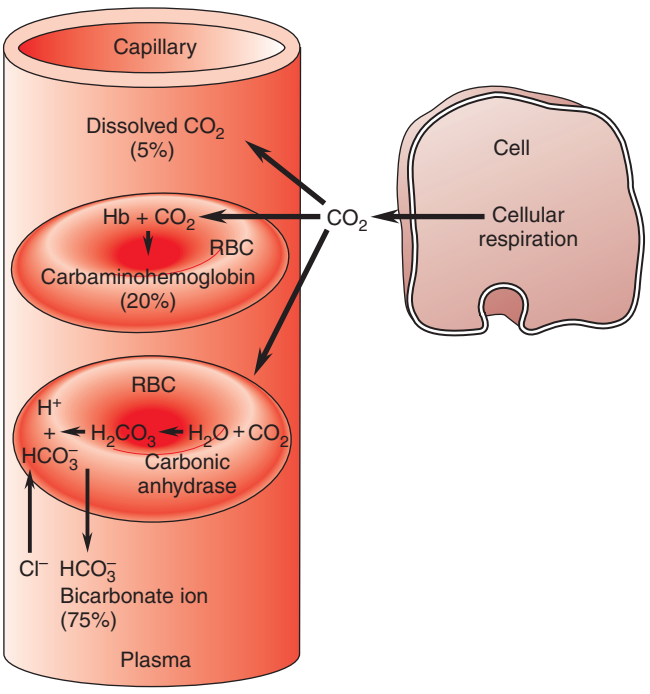


FIG 13.15 Carbon dioxide is transported in three forms in the blood. Transportation of CO_2 (1) as dissolved gas, (2) as bicarbonate ion (HCO_3^-), and (3) in association with hemoglobin (Hb).

KEY POINTS

- Nearly all (97%) of the oxygen transported in blood is bound to hemoglobin within the red cells. Only 3% is dissolved in plasma. It is this 3% that is measured as PaO_2. At a normal PaO_2, hemoglobin is 95% to 100% saturated with oxygen. About 25% of the bound oxygen is unloaded to the tissues, resulting in a venous hemoglobin saturation of about 75%.
- The oxyhemoglobin dissociation curve describes the relationship between the partial pressure of oxygen and hemoglobin saturation. In the lung, where PO_2 is high (100 mm Hg), oxygen is loaded onto hemoglobin. In the tissues, where PO_2 is low (40 mm Hg), oxygen is unloaded from hemoglobin to tissues.
- The affinity of hemoglobin for oxygen is affected by temperature, acid–base status, 2,3-DPG levels, and carbon dioxide concentration. Affinity decreases at the tissue level because of increased levels of acid, 2,3-DPG, and carbon dioxide. This shift to the right of the oxyhemoglobin dissociation curve enhances unloading of oxygen to the tissue. A shift to the left occurs in the lungs, where blood is more alkalotic and carbon dioxide levels are lower. The increased affinity of hemoglobin for oxygen at the lung facilitates oxygen binding.
- The oxygen content of arterial blood is calculated by adding the amount bound to hemoglobin (Hb) plus the amount dissolved in plasma: $CaO_2 = (Hb \times 1.34 \times SaO_2) + (PaO_2 \times 0.003)$. Oxygen delivery to the body tissues is calculated by multiplying CaO_2 by cardiac output (CO): $\dot{D}O_2 = CaO_2 \times CO \times 10$.
- The consumption of oxygen by tissues can be estimated using the Fick equation: $\dot{V}O_2 = CO \times (CaO_2 - CvO_2) \times 10$. Oxygen consumption increases with increased tissue metabolism.

- Hemoglobin is an important factor in carbon dioxide transport in the blood. In the tissues, hemoglobin binds carbon dioxide to form carbaminohemoglobin, which then releases carbon dioxide in the lungs. RBCs contain the enzyme carbonic anhydrase, which greatly increases the conversion of carbon dioxide and water into HCO_3^- and H^+ at the tissue level. In the lungs, the reaction proceeds in reverse, producing carbon dioxide, which is eliminated by the lungs.
- Factors that decrease hemoglobin mass (such as anemia) or decrease arterial saturation (such as hypoxia from either cardiac or pulmonary conditions) impair oxygen delivery to the body tissues. This stimulates an increased release of erythropoietin and the production of RBCs.

ANEMIA

Erythrocyte disorders are divided into two groups: (1) anemia, defined as a deficit of red cells, and (2) polycythemia, defined as an excess of red cells (Box 13.1). An anemic patient has tissue hypoxia attributable to the low oxygen-carrying capacity of the blood. In contrast, a patient with polycythemia has increased blood viscosity and volume attributable to the increase in the number of RBCs. (Polycythemia is discussed in greater detail later in this chapter.) Table 13.5 presents the laboratory findings for erythrocyte disorders in terms of relative anemia, absolute anemia caused by decreased RBC production, and absolute anemia caused by increased RBC destruction.

Relative anemia is characterized by normal total red cell mass with disturbances in the regulation of plasma volume. For example, in pregnant women the average plasma volume is 43% greater than in nonpregnant women, which causes a "dilutional anemia."

Absolute anemia includes those types of anemia with an actual decrease in the numbers of red cells. This can be caused by decreased production of red cells or increased destruction of red cells.

General Effects of Anemia

The clinical manifestations of anemia include a reduction in oxygen-carrying capacity, tissue hypoxia, and compensatory mechanisms to restore tissue oxygenation. Increased pulmonary and cardiac function increases the oxygen supply, and an increase in oxygen extraction occurs to protect tissues. Specific adaptations to anemia to increase oxygenated blood flow include an increase in the heart rate, cardiac output, and circulatory rate and a preferential increase in blood flow to vital organs. Specific adaptations to anemia to increase oxygen utilization by tissues include an increase in 2,3-DPG concentration in erythrocytes and a decreased oxygen affinity of hemoglobin in tissues. Selective tissue

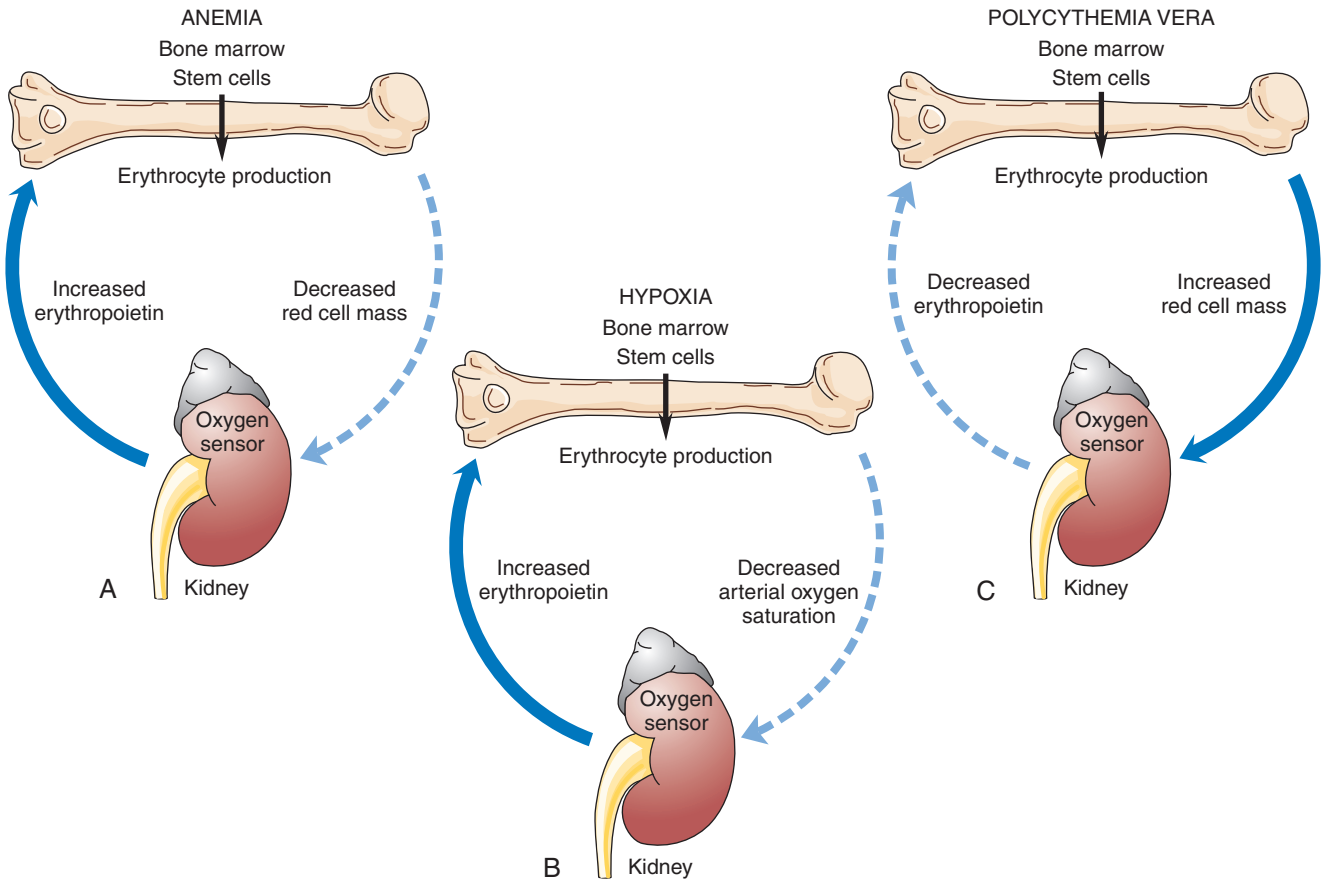


FIG 13.16 Alterations in the erythropoietin feedback circuit. Any factor decreasing oxygen delivery to the oxygen sensor cells results in increased secretion of erythropoietin and a compensatory increase in erythrocyte production as illustrated in **(A)** for anemia, with a decrease in erythrocyte mass; and in **(B)** for hypoxia, with a decrease in arterial oxygen saturation. An increase in erythrocyte mass, as occurs with polycythemia vera **(C),** decreases erythropoietin production.

TABLE 13.4 Gas Values Significant to the Oxygenation of Blood

Gas Values	Description	Measurement/Reflection	Calculation/Formula	Normal Value/Formula
CaO_2	Arterial blood oxygen content is amount of oxygen carried in arterial blood	Measured in milliliters of oxygen per deciliter of blood (mL/dL), or vol%	Sum of oxyhemoglobin (15 Hb g/100 mL × 1.34 mL O_2/g Hb × 97.5% arterial saturation = 19.6 vol%) + amount of oxygen dissolved in plasma (PaO_2 = 100 mm Hg × 0.003 vol%/mm Hg = 0.3 vol%)	~20 vol% Formula: 19.6 vol% oxyhemoglobin + 0.3 vol% dissolved in plasma = 19.9 vol%
CvO_2	Venous blood oxygen content is amount of oxygen carried in venous blood	Measured in milliliters of oxygen per deciliter of blood (mL/dL), or vol%	Sum of oxyhemoglobin (15 Hb/100 mL × 1.34 mL O_2/g Hb × 75% venous saturation = 15.0 vol%) + amount of oxygen dissolved in plasma CvO_2 = 40 mm Hg × 0.003 vol%/mm Hg = 0.12 vol%)	~15 vol% Formula: 15 vol% oxyhemoglobin + 0.12 vol% dissolved in plasma = approximately 15 vol%
$\dot{D}O_2$	Oxygen delivery or transport is amount of oxygen delivered to tissues	Measured in milliliters of oxygen per minute (mL/min)	*Normal arterial:* $\dot{D}O_2$ = cardiac output (L/min) × CaO_2 × 10 *Normal venous:* $\dot{D}O_2 = \text{cardiac output (L/min)} \times CvO_2 \times 10$	*Normal arterial:* ~1000 mL of O_2/min *Normal venous:* ~750 mL of O_2/min
PaO_2	Partial pressure of oxygen in arterial blood	Measured in millimeters of mercury (mm Hg) Reflects tension or pressure that is exerted by oxygen when it is dissolved in plasma		Normal PaO_2 is 80–100 mm Hg
$PaCO_2$	Partial pressure of carbon dioxide in arterial blood	Measured in millimeters of mercury (mm Hg) Reflects tension or pressure that is exerted by carbon dioxide when it is dissolved in plasma		Normal $PaCO_2$ is 35–45 mm Hg
PvO_2	Partial pressure of oxygen in venous blood	Measured in millimeters of mercury (mm Hg) Reflects tension or pressure that is exerted by oxygen when it is dissolved in plasma		Normal PvO_2 is 35–40 mm Hg
$PvCO_2$	Partial pressure of carbon dioxide in venous blood	Measured in millimeters of mercury (mm Hg) Reflects tension or pressure that is exerted by carbon dioxide when it is dissolved in plasma		Normal $PvCO_2$ is 41–51 mm Hg
SaO_2	Amount of hemoglobin bound to oxygen relative to total amount of hemoglobin, both reduced and bound, in arterial blood	Expressed as percentage		Normal SaO_2 is 95%–100%
SvO_2	Amount of hemoglobin bound to oxygen relative to total amount of hemoglobin, both reduced and bound, in venous blood	Expressed as percentage		Normal SvO_2 is 60%–80%
V̇O2	Oxygen consumption is amount of oxygen consumed by tissues	Measured in milliliters of oxygen per minute (mL/min) Oxygen consumption is derived from difference between arterial oxygen transport and venous oxygen transport	$\dot{V}O_2 = \text{cardiac output} \times (CaO_2 - CvO_2) \times 10$	Normal $\dot{V}O_2$ is 200–250 mL of O_2/min

BOX 13.1 Classification of Anemia and Polycythemia

I. Anemia

A. Absolute

1. Decreased red cell production
   - (a) Acquired
     - 1) Pluripotent stem cell failure
       - a) Aplastic anemia
         - {1} Radiation induced
         - {2} Drugs and chemicals
         - {3} Viruses
         - {4} Idiopathic
       - b) Anemia of leukemia and myelodysplastic syndrome
       - c) Anemia associated with marrow infiltration
       - d) Anemia associated with chemotherapy
     - 2) Erythroid progenitor cell failure
       - a) Pure red cell aplasia
       - b) Endocrine disorders
       - c) Acquired sideroblastic anemia
     - 3) Functional impairment of erythroid progenitors
       - {1} Megaloblastic anemias
         - a) B_{12} deficiency
         - b) Folate deficiency
         - c) Acute megaloblastic anemia due to nitrous oxide
         - d) Drug-induced megaloblastic anemia
       - {2} Iron deficiency anemia
       - {3} Anemia from other nutritional deficiencies
       - {4} Anemia of chronic disease
       - {5} Anemia of renal failure
       - {6} Anemia caused by chemical agents
       - {7} Anemia caused by thalassemias
       - {8} Erythropoietin antibodies
   - (b) Hereditary
     - 1) Pluripotent stem cell failure
       - {1} Fanconi anemia
       - {2} Shwachman syndrome
       - {3} Dyskeratosis congenita
     - 2) Erythroid progenitor cell failure
       - {1} Diamond–Blackfan syndrome
       - {2} Congenital dyserythropoietic syndrome
     - 3) Hereditary sideroblastic anemia
     - 4) Thalassemias
2. Increased red cell destruction or loss
   - (a) Acquired
     - 1) Mechanical
       - a) Macroangiopathic (artificial heart valves)
       - b) Microangiopathic (disseminated intravascular coagulation [DIC])
       - c) Parasites and microorganisms
     - 2) Antibody mediated
     - 3) Hypersplenism
     - 4) Chemical and physical injury
     - 5) Acute blood loss
   - (b) Hereditary
     - 1) Hemoglobinopathies (sickle cell)
     - 2) Red cell membrane disorders
     - 3) Red cell enzyme defects
     - 4) Porphyrias

B. Relative (increased plasma volume)

1. Macroglobulinemia
2. Pregnancy
3. Athletes
4. Postflight astronauts

II. Polycythemia (Erythrocytosis)

A. Relative (decreased plasma volume)

1. Dehydration
2. Diuretics
3. Stress or smoker's erythrocytosis

B. Absolute (increased red cell volume)

1. Primary polycythemia
   - (a) Acquired (polycythemia vera)
   - (b) Hereditary congenital polycythemia
2. Secondary polycythemia
   - (a) Appropriate
     - 1) Altitude
     - 2) Cardiopulmonary disorders
     - 3) Increased hemoglobin affinity for oxygen
   - (b) Inappropriate
     - 1) Renal cysts and tumors
     - 2) Hepatoma
     - 3) Cerebellar hemangioblastoma
     - 4) Essential

Adapted from Prchal JT: Clinical manifestations and classification of erythrocyte disorders. In Lichtman MA, Kipps TJ, Seligsohn U, Kaushansky K, Prchal JT, editors: *Williams hematology,* ed 8, New York, 2010, McGraw-Hill. Accessed at www.accessmedicine.com/content.aspx?aID=6108487.

perfusion provides shunting to vital organs in short-term compensation, and increased erythropoietic activity is stimulated to provide long-term compensation. The extent of the physiologic adaptations is influenced by (1) the severity of the anemia; (2) the competency of the pulmonary and cardiac systems; (3) the oxygen requirements of the individual, which are dependent on physical and metabolic activity; (4) the duration of the anemia; (5) the underlying disease or condition; and (6) the presence and severity of coexisting disease. Mild anemia is usually associated with no clinical symptoms; however, early symptoms in patients with mild to moderate anemia include fatigue, generalized weakness, and loss of stamina, followed by tachycardia and exertional dyspnea. Healthy young patients may present with very few symptoms at hemoglobin levels of 7 to 8 g/dL; however, elderly patients and patients with cardiovascular or pulmonary disease may have symptoms with even modest degrees of anemia and hemoglobin levels of 9 to 11 g/dL. Specific symptoms related to moderate to severe anemia are orthostatic and nonorthostatic hypotension, vasoconstriction, pallor, tachypnea, dyspnea, tachycardia, transient murmurs, angina pectoris, heart failure, intermittent claudication, night cramps in muscles, headache, lightheadedness, tinnitus, roaring in the ears, and faintness.

TABLE 13.5 Laboratory Findings for Erythrocyte Disorders

Disease	HCT	HB	MCV	MCH	MCHC	Retic	RBC	WBC	PLT	Blood Smear	Laboratory Tests	Diagnostic Characteristics
Relative anemia	Low	Low	Normal	Low	Low	Normal	Low	Low	Low	Normal	Plasma volume increased, causing relative decrease in number of cells	Increased volume can be caused by pregnancy, splenomegaly, IV infusions
Absolute Anemia												
Aplastic anemia	Low	Low	Normal	Low	Low	Low	Low	Low	Low	Normocytic, hypochromic RBCs; lack of neutrophils; increased lymphocytes	HbF may be increased; erythropoietin increased; bone marrow aplastic	Specific cause should be identified and removed from environment
Chronic renal failure	Low	Low	Normal	Normal	Normal	Low	Low	Normal	Normal	Normocytic, normochromic RBCs; RBCs often have spicules	Erythropoietin decreased; bone marrow production suppressed	Kidney tests abnormal
Pernicious anemia	Low	Low	High	High	Normal	Low	Low	Low	Low	Oval macrocytes; hypersegmented segs	Decreased B_{12} level; positive Schilling test	Neurologic symptoms; increased bilirubin
Folate deficiency	Low	Low	High	High	Normal	Low	Low	Low	Low	Oval macrocytes; hypersegmented segs	Decreased folic acid level; negative Schilling test	No neurologic symptoms
Iron deficiency	Low	Low	Low	Low	Low	Normal or low	Low	Normal	Normal	Microcytic, hypochromic RBCs	Serum iron decreased; iron-binding	Bone marrow iron decreased
Thalassemia	Low	Low	Low	Low	Low	High	Low	Normal	Normal	Microcytic, hypochromic RBCs	Decreased osmotic fragility; hemoglobin electrophoresis diagnostic; serum iron, TIBC, and ferritin normal	Hereditary disease
Absolute Anemia Caused by Increased Destruction												
Intrinsic Abnormality												
Sickle cell	Low	Low	Normal	Normal	Normal	High	Low	Normal	Normal	Normocytic, normochromic RBC target cells; sickle cells; NRBCs	HbS present on electrophoresis	Hereditary disease

Continued

TABLE 13.5 Laboratory Findings for Erythrocyte Disorders—cont'd

Disease	HCT	HB	MCV	MCH	MCHC	Retic	RBC	WBC	PLT	Blood Smear	Laboratory Tests	Diagnostic Characteristics
Hereditary spherocytosis	Low	Low	Normal	Normal	Normal to high	High	Low	Normal	Normal	Spherocytes present	Bilirubin elevated; haptoglobins reduced; abnormal RBC fragility	Hereditary disease
G6PD deficiency	Low	Low	Normal	Normal	Normal	High	Low	High	Normal	Heinz body smear positive	Tests only abnormal in hemolytic episodes	Hereditary disease
Extrinsic Abnormality												
HDNB	Low	Low	Normal	Normal	Normal	High	Low	High	Normal	Spherocytes, NRBCs	Bilirubin elevated; Coombs test positive; urinary urobilinogen increased	Jaundice; edema; hepatosplenomegaly
Antibody-mediated drug reactions	Low	Low	Normal	Normal	Normal	High	Low	Normal	Normal	Polychromatic RBCs due to increased reticulocytes	Bilirubin elevated; Coombs test positive; urinary urobilinogen increased	Jaundice
Acute blood loss	Normal to low	Normal to low	Normal	Normal	Normal	High	Normal to low	Normal to low	Normal to low	Appears normal until reticulocytes increase	Values depend on severity of hemorrhage and when blood is drawn	
Polycythemias												
Relative polycythemia	High	High	Normal	Normal	Normal	Normal	High	High	High	Normal	Plasma volume decreased, causing relative increase in number of blood cells	Decrease in volume
Absolute polycythemia vera	High	High	Normal	Normal	Normal	High	High	High	High	Teardrops, macrocytes, and NRBCs may be present; shift to left on differential	O_2 saturation normal; bone marrow hypercellular; all three cell lines increased	
Secondary polycythemia	High	High	Normal	Normal	Normal	High	High	Normal	Normal	Normal	Hypoxemia may be evident; serum erythropoietin elevated	Lung disease may be present

G6PD, Glucose-6-phosphate dehydrogenase; *Hb,* hemoglobin; *Hct,* hematocrit; *HDNB,* hemolytic disease of the newborn; *MCH,* mean corpuscular hemoglobin; *MCHC,* mean corpuscular hemoglobin concentration; *MCV,* mean corpuscular volume; *NRBC,* nucleated red blood cell; *PLT,* platelet; *RBC,* red blood cell; *RETIC,* reticulocytosis; *Segs,* segmented neutrophils; *TIBC,* total iron binding capacity; *WBC,* white blood cell.

ANEMIA RELATED TO DECREASED RED CELL PRODUCTION

Aplastic Anemia

Etiology and pathogenesis. Aplastic anemia is a stem cell disorder that is characterized by a reduction of hematopoietic tissue in the bone marrow, fatty marrow replacement, and pancytopenia. The decrease in functional bone marrow mass is usually caused by toxic, radiant, or immunologic injury to the bone marrow stem cells, which causes a decrease in the levels of red cells, white cells, and platelets, or pancytopenia.

Aplastic anemia can be classified as acquired or familial. Acquired aplastic anemia can be caused by chemical and physical agents, such as those listed in Table 13.6. Other causes include certain viral infections (e.g., hepatitis, Epstein–Barr virus, HIV, dengue), some mycobacterial infections, diffuse eosinophilic fasciitis, pregnancy, Simmonds disease, and sclerosis of the thyroid. Familial aplastic anemia is associated with Fanconi constitutional pancytopenia, pancreatic deficiency in children, and putative hereditary defect in cellular uptake of folate.

Laboratory features. Pancytopenia is characterized by low red cell, white cell, and platelet counts. The magnitude of the granulocytopenia is very important for the immediate prognosis. An absolute granulocyte count of less than 200/mm^3 results in immediate susceptibility to infectious complications. Coagulation tests are generally normal except for the bleeding time, which reflects the low platelet count. The ultimate diagnosis of aplastic anemia rests on the interpretation of an adequate bone marrow biopsy specimen, although important clues to the cause of pancytopenia can be obtained from the history, physical examination, and laboratory data. Pancytopenia that is not primarily hematologic in origin but secondary to other disease processes is usually an obvious diagnosis.

Clinical manifestations. Aplastic anemia is a disease of the young, with most patients presenting between 15 and 25 years of age. Another age group likely to present with aplastic anemia are those ≥60 years old. The most common form of aplastic anemia is iatrogenic—resulting from a transient marrow failure after treatment with cytotoxic chemotherapeutic drugs or irradiation. Certain chemical or physical agents directly injure proliferating and quiescent hematopoietic cells. The onset is usually insidious, and patients often present only after the late manifestations of pancytopenia are evident. The symptoms attributable to the gradual decrease in the number of RBCs include weakness, fatigue, lethargy, pallor, dyspnea, palpitations, onset of transient murmurs, and tachycardia of anemia. Fever, chills, and bacterial infections (particularly in the mouth or perirectal area) are seen secondary to neutropenia. Petechiae, bruising, nosebleeds, retinal hemorrhage, and increased menstrual flow are manifestations of aplastic anemia.

Treatment. Treatment for aplastic anemia is multifaceted and dependent on the etiology and severity of the disease. Treatment includes (1) identification and avoidance of further toxin exposure; (2) human leukocyte antigen (HLA) and ABO typing of family members to identify serologically defined loci and potential bone marrow transplant donors; (3) maintenance of minimally essential levels of hemoglobin and platelets; (4) prevention and management of infection; (5) determination of efficacy of bone marrow transplantation; and (6) implementation of other forms of therapy, such as immunosuppressive therapy or stimulation of hematopoiesis and bone marrow regeneration in patients not suited for transplantation. In patients with severe disease, the major curative approach is allogeneic bone marrow transplantation; however, only one-third of all patients have compatible donors. Preparative regimens using cyclophosphamide and antithymocyte globulin followed by posttransplant immunosuppression with cyclosporine and methyltrexate as prophylaxis against graft-versus-host disease (GVHD) have resulted in a 90% disease-free survival rate at 2 years for patients with bone marrow transplants derived from an HLA-matched sibling donor.

Course and prognosis. Bone marrow transplantation is highly successful and curative for 80% to 85% of untransfused patients and 55% to 60% of patients with multiple previous transfusions. Approximately 20% to 30% of transplantation survivors with multiple previous transfusions experience severe GVHD, which can be significantly improved by immunosuppressive therapy in 50% to 70% of patients. The risk of graft failure in patients who have not been transfused is less than 5%. Prognosis is related to the absolute neutrophil count and the platelet count. Children respond better than adults with both bone marrow transplantation and immunosuppression therapy, especially in patients with mild to moderate disease. Aplastic anemia is fatal unless successfully managed with bone marrow transplantation.

Anemia of Chronic Renal Failure

Etiology and pathogenesis. The anemia of chronic renal failure occurs primarily from failure of the renal endocrine function, which causes impaired erythropoietin production and secondarily from failure of the renal excretory function, leading to hemolysis, bone marrow cell depression, and blood loss.

Laboratory features. This anemia is characterized by a decreased red cell count and low hemoglobin and hematocrit values. Some red cells appear grossly deformed with a few large spicules (Fig. 13.17). The total leukocyte differential cell count, leukocyte counts, and platelet count are usually normal. The red cell indices—mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), and mean corpuscular hemoglobin concentration (MCHC)—are also normal.

Clinical manifestations. Any of the clinical manifestations described earlier (see "General Effects of Anemia") may be evident in chronic renal failure. The hematocrit falls in proportion to the degree of renal insufficiency, and uremia occurs as the glomerular filtration rate drops below 40 mL/min. Signs and symptoms of anemia usually manifest when the hematocrit decreases to ≤20%.

Treatment. Therapy consists of dialysis when the glomerular filtration rate is less than 15 mL/min, and erythropoietin is administered to achieve the target hematocrit concentration of 33% to 36% and hemoglobin level of 11 to 12 g/dL. However, excessive correction of the hemoglobin level to greater than 12 g/dL may be associated with an increased incidence of cardiovascular and thromboembolic events. Therefore the

TABLE 13.6 Drugs Associated With Aplastic Anemia

Dose Dependent Effects	Patient Dependent (Idiosyncratic) Effects
Alkylating agents (e.g., nitrogen mustard)	Anti-inflammatory (e.g., phenylbutazone, penicillamine)
Antimetabolites (e.g., fluorouracil, methotrexate)	Anti-seizure (e.g., hydantoin, carbamezepine)
Antibiotics (e.g., chloramphenicol)	Gold salts

Modified from Kumar V: *Robbins and Cotran pathologic basis of disease,* ed 9, Philadelphia, 2015, Elsevier.

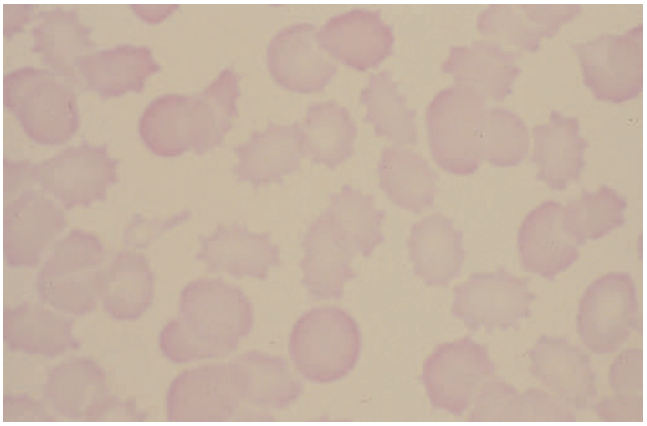


FIG 13.17 Burr cells found in acute kidney disease. (Courtesy Beth Payne, Sacred Heart Medical Center, Spokane, WA.)

use of erythropoiesis-stimulating agents to increase hemoglobin values to greater than 12 g/dL is not routinely recommended. The hematocrit and hemoglobin values are monitored at least every 2 weeks to ensure adequate oxygen-carrying capacity and minimize an increased respiratory rate and heart rate. When the target hematocrit is achieved, adult patients can be maintained by administering 50 to 100 units per kilogram per week in divided doses. Pediatric patients younger than 5 years usually require higher initial and maintenance doses. Patients with chronic renal failure are also prone to nutritional anemias (lack of iron, folate, and B_{12}) because of dietary restrictions and anorexia. Patient replacement of iron, folate, and B_{12} to adequate levels should be considered a treatment goal. Some patients may fail to respond or may be resistant to the effects of erythropoietin. The most common cause for a failure in response is iron deficiency.

Course and prognosis. More than 95% of patients respond to erythropoietin therapy. Patients who do not respond or first respond when larger doses are given should be evaluated for an adequate iron supply, infection, or excessive splenic hemolysis.

Anemia Related to Vitamin B_{12} (Cobalamin) or Folate Deficiency

Etiology and pathogenesis. The anemia resulting from a deficiency of either vitamin B_{12} (cobalamin) or folate is caused by a disruption in DNA synthesis of the blast cells in the bone marrow. This disruption produces very large abnormal bone marrow cells called megaloblasts. In the peripheral blood, the red cells are larger than normal (macrocytic); the granulocytes are hypersegmented; and the numbers of red cells, white cells, and platelets are decreased. All of these signs can be seen on the peripheral blood smear.

The classic anemia in this classification is pernicious anemia. The fundamental defect causing pernicious anemia is the lack of intrinsic factor. Without it, vitamin B_{12} cannot be absorbed, thus leading to vitamin B_{12} deficiency. This deficiency results in disordered nucleic acid metabolism, which causes megaloblastic dysplasia, a condition involving abnormal production and maturation of red cell, white cell, and platelet systems. There is strong evidence that pernicious anemia develops as a result of genetically determined autoimmune disease, which is manifested by serum and gastric juice antibodies against intrinsic factor and parietal cells. The biochemical basis of the neurologic lesions in pernicious anemia is not known. There can be peripheral nerve degeneration, degeneration of the posterior columns of the spinal cord, or both. There is some evidence of abnormal fatty acid metabolism in the peripheral nerves and degeneration of the white matter in the spinal cord in animals.

Folate deficiencies resemble vitamin B_{12} deficiencies except for the neurologic disease, which is more characteristic of vitamin B_{12} deficiency. Folate deficiencies are usually the result of dietary deficiencies, alcoholism and cirrhosis, pregnancy, or infancy.

Laboratory features. The peripheral blood shows low RBC counts of 500,000 to 750,000 cells/mm^3, low WBC counts of 4000 to 5000 cells/mm^3, and low platelet counts of 50,000 cells/mm^3. These counts are usually not as low as those seen in aplastic anemia. The bone marrow shows megaloblastic dysplasia, which results in a peripheral blood picture of macrocytosis and hypersegmented neutrophils. The red cell indices reveal normal MCH and MCHC and increased MCV. The Schilling test, which measures excretion of radioactive vitamin B_{12}, indicates low levels, and the serum level of vitamin B_{12} is low. Gastric analysis indicates a lack of free hydrochloric acid in the gastric juice (achlorhydria).

Clinical manifestations. The clinical features of vitamin B_{12} deficiency include paranoid ideation, dementia, cognitive dysfunction, delusions, and hallucinations, often referred to as *megaloblastic madness*. The neurologic abnormalities include symmetric paresthesias of the feet and hands with vibratory sense and proprioception disturbances. The paresthesias progress to spastic ataxia as a result of degenerative changes of the dorsal and lateral columns of the spinal cord. Cerebral signs include irritability; somnolence; memory impairment; and perversion of taste, smell, and vision. Manifestations of pure folate deficiency include a blunted affect in general demeanor with evidence of depression, sleep deprivation, and irritability. History of circumstances likely to result in folic acid deficiency includes poor or fad diet, frank malabsorption, or alcoholism. In folate deficiency, cerebral symptoms, such as irritability, memory loss, and personality changes, are seen. Clinical manifestations that are seen in both vitamin B_{12} and folate deficiencies include pedal edema, nocturia, tachypnea, dyspnea, and tachycardia associated with heart congestion; glossitis, weight loss, malabsorption, and episodic or chronic diarrhea with steatorrhea are GI manifestations. Musculoskeletal symptoms of arthralgia and frank arthritis are seen in autoimmune diseases; nocturnal pain and upper and/or lower extremity cramps often indicate spinothalamic tract involvement. Dermatologic symptoms include blotchy brown skin pigmentation, especially in nail beds and skin creases. When this is associated with vitiligo, autoimmune processes should be suspected.

Research has reported an association between low folate levels and the risk of neural tube defects and abnormalities of the heart, urinary tract, and limbs in neonates. These data support the routine supplementation before pregnancy of all women who might become pregnant with 1.0 mg/day of folic acid. This is the largest dose that will not mask vitamin B_{12} deficiency.

Treatment. Routine treatment with full doses of parenteral vitamin B_{12} (1 mg/day) and oral folate (1 to 5 mg/day) before the cause of the deficiency is identified should only be used in critically ill patients. In managing the anemia related to vitamin B_{12} or folate deficiency, it is important to (1) recognize that megaloblastic anemia is present; (2) ascertain if vitamin B_{12}, folate, or a combined deficiency is the cause; and (3) diagnose the underlying disease and mechanism responsible for the deficiency. In vitamin B_{12} and folic acid deficiency anemia, replacement therapy for vitamin B_{12} is oral cobalamin and folic acid. Transfusion therapy may be indicated in elderly or critically ill patients. Hypokalemia should be managed with potassium supplements to prevent sudden death, reportedly associated with a sharp drop in serum potassium level seen in vitamin B_{12} therapy.

Course and prognosis. The majority of patients respond well to replacement therapy; however, continued assessment and monitoring

of these patients are essential to prevent hematologic or neurologic relapse secondary to inadequate therapy. In patients with neurologic signs and symptoms, the reversibility of the neurologic damage is slow, with a maximal response requiring up to 6 months. Further substantial increases in recovery are unlikely after 12 months. In 90% of patients with subacute combined degeneration, major improvement is seen. The degree of functional recovery is inversely related to the extent of the disease and duration of the signs and symptoms. Patients with signs and symptoms of less than 3 months' duration may have complete reversal.

Iron Deficiency Anemia

Etiology and pathogenesis. Iron deficiency, the most common nutritional deficiency in the world, is the most common cause of anemia. Iron deficiency results in the unavailability of iron for hemoglobin synthesis. This may be due to low intake, diminished absorption (such as from chronic disease), physiologic increase in requirements (such as during pregnancy), excessive iron loss (such as from acute or chronic hemorrhage), chronic renal failure, hemodialysis, or idiopathic iron loss. The most common cause in men is occult GI bleeding and in women is menorrhagia. Iron is one of the most carefully conserved body substances, and under normal conditions very little is lost except as a result of bleeding. Normal dietary requirements, if 10% is absorbed, are as follows: adult men, 12 mg/day; adult women ages 14 to 30 years, 15 mg/day; and adult women ages 60 years or more, 10 mg/day. Pregnant women require up to 30 mg/day, and children require 10 mg/day. A normal diet supplies the adult with about 10 to 15 mg/day.

Laboratory features. In latent iron deficiency there may be no anemia; however, after patients receive iron, they respond with a significant increase in blood hemoglobin level. In a typical case caused by chronic bleeding, the reduction in hemoglobin concentration is proportionately greater than the reduction in the red cell count. The red cells are smaller and paler than normal RBCs because of the decreased amount of hemoglobin and are described as hypochromic, microcytic red cells. Therefore the red cell indices MCV, MCH, and MCHC are decreased. The white cell counts are usually normal. The platelet count varies, depending on the cause of the deficiency. In severely anemic children and infants, thrombocytopenia may be present. In patients who are bleeding, thrombocytosis may be present. The serum ferritin level is decreased to less than 10 ng/mL, the serum iron level is decreased, total iron binding capacity (TIBC) is increased, and tissue iron stores are decreased.

Clinical manifestations. Patients with iron deficiency may present with (1) no signs or symptoms, only seeking medical attention because of abnormalities noted on laboratory tests; (2) features of the underlying disorder responsible for the development of iron deficiency; or (3) manifestations common to all anemias, such as pallor, weakness, fatigue, dyspnea, palpitations, new and transient heart murmurs, irritability, headaches, or lightheadedness. Patients may also present with (4) one or more of the few signs and symptoms considered highly specific for iron deficiency, including pagophagia or pica (craving for nonfood substances such as dirt, clay, ice, laundry starch, cardboard, or hair), koilonychias (spoon-shaped nails), and blue sclerae. In addition, a high prevalence of iron deficiency with or without anemia has been reported among patients with restless legs syndrome (Ekbom syndrome), especially in the elderly. In severe cases, GI symptoms are seen, such as glossitis, dysphagia, erosions at the corners of the mouth, esophageal webbing, and atrophic gastritis, as well as changes in the fingernails, conjunctival pallor, and splenomegaly.

Treatment. To maintain a normal iron balance in the body, men need to absorb 1.0 to 1.5 mg/day and women need to absorb 2 to 3 mg/day because of iron losses with menstruation. The goal of therapy for iron deficiency anemia is to supply sufficient iron to repair the hemoglobin deficit and to replenish iron stores. Oral iron is the treatment of choice for almost all patients because of its effectiveness, safety, and economy and should always be given preference over parenteral iron for initial treatment. Iron deficiency anemia is managed with oral administration of ferrous sulfate taken separately from meals in three or four divided doses and supplying a daily total of 150 to 200 mg of elemental iron in adults or 3 mg of iron per kilogram of body weight in children until hematologic normality is reached. Infants may be given 50 to 100 mg daily in divided doses. Thereafter it is important to continue the treatment for 4 to 6 months to build iron stores. Urgent treatment may be accomplished with the administration of parenteral ferric gluconate after a test dose to determine possible hypersensitivity. Patients undergoing dialysis should have a serum ferritin level greater than 100 μg/L to optimize their response to erythropoietin administration. Although iron therapy remediates the iron deficiency anemia, the underlying cause must be determined and corrected.

Course and prognosis. The symptoms may be alleviated in the first few days of treatment. The reticulocyte count is an index of erythropoiesis. The reticulocyte count increases as the RBC production increases and usually reaches maximal levels in 7 to 12 days, and the hemoglobin level is usually normal by 2 months after initiation of therapy. Failure to obtain a complete and characteristic response to iron therapy should cause the clinician to review the findings and reevaluate the patient. One possible problem is an incorrect diagnosis, when the anemia of chronic disease is mistaken for the anemia of iron deficiency. The prognosis is excellent if the underlying cause is benign; however, even in patients with incurable disease states, management of iron deficiency anemia with iron therapy can increase the comfort level.

ANEMIA RELATED TO INHERITED DISORDERS OF THE RED CELL

A number of inherited genetic defects can result in anemia. Abnormalities of hemoglobin synthesis causing slow production (thalessemia) or increased RBC hemolysis and destruction (sickle cell disease) can cause anemia. Inherited enzyme deficiencies (G6PD deficiency) or structural abnormalities (spherocytosis) can reduce the lifespan of RBCs resulting in anemia. Common inherited forms of anemia are discussed.

Thalassemia

The thalassemias are a group of diseases associated with the presence of mutant genes that suppress the rate of synthesis of globin chains. Thalassemias are classified according to the polypeptide chain or chains with deficient synthesis, such as α-thalassemia or β-thalassemia.

Etiology and pathogenesis. There are two main classes of thalassemia, α and β, in which the α- and β-globin genes are involved, and rarer forms caused by abnormalities of other globin genes. These conditions all have in common an imbalanced rate of production of the globin chains of adult hemoglobin—excess α chains in β-thalassemia and excess β chains in α-thalassemia. Several hundred different mutations at the α- and β-globin loci have been defined as the cause of the reduced or absent output of α or β chains. The high frequency and genetic diversity of the thalassemias are related to past or present heterozygote resistance to malaria. A deficiency in one or more polypeptide chains causes decreased hemoglobin synthesis and an imbalance between α-chain and non–α-chain production. Because of the lack of hemoglobin, the anemia is severe, and the peripheral cells are microcytic and hypochromic. The disruption of the globin balance causes the abundant chains to

accumulate and precipitate within the cytoplasm. This damages the cell membranes, which leads to premature cell destruction. The most clinically severe form of the thalassemias is *thalassemia major,* which occurs in homozygous patients. *Thalassemia minor* is the term used to describe the heterozygous carrier state. For example, in homozygous β-thalassemia, the deficiency of β-chain synthesis results in the accumulation of α chains, which aggregate to form insoluble inclusions in bone marrow erythroid precursors (Fig. 13.18). These inclusions cause early destruction of 70% to 85% of marrow erythroblasts. In response to this massive destruction, erythroid cell proliferation in homozygous β-thalassemia is significant.

Patients who are significantly anemic have an increased intestinal iron absorption that is related to the degree of expansion of the RBC precursor population. This can be decreased with blood transfusions. The iron accumulates in the Kupffer cells of the liver, the macrophages in the spleen, and the parenchymal cells of the liver.

Laboratory features. Laboratory values vary, depending on the severity of the imbalance, which is determined by the genetic pattern. Because of the decrease in hemoglobin level, the red cells are hypochromic and microcytic, and red cell indices—MCV, MCH, and MCHC—are decreased. Many target cells are present. In homozygous or major syndromes, the hemoglobin concentration is often less than 7 g/dL, and there are nucleated red cells in the peripheral blood. The leukocyte number is usually increased but the platelet number is normal. The bone marrow is hypercellular, with profound erythroblastic hyperplasia. There is evidence of hemolysis with increased unconjugated bilirubin levels and increased excretion of urobilin and urobilinogen. Hemoglobin electrophoresis is performed to determine the type of abnormal hemoglobin. An increased level of fetal hemoglobin ranging from 10% to 90% is characteristic of homozygous α-thalassemia. No hemoglobin α is produced. Excess γ chains form γ4 homotetramers, or Bart hemoglobin. Excess β chains form β4 homotetramers, or hemoglobin H.

Clinical manifestations. Patients may have any of the clinical manifestations described earlier (see "General Effects of Anemia"). The clinical findings are the result of deficient α-globin production in α-thalassemia, or α-globin chain excess and persistent hemoglobin F production in β-thalassemia.

α-Thalassemia is found primarily in Asian individuals; however, it has also been documented in increasing numbers in individuals of Mediterranean or African descent. Usually patients with α-thalassemia minor are silent carriers or present with mild to moderate anemia. They are recognized during familial studies after the identification of a family member with Bart hemoglobin hydrops fetalis or hemoglobin H disease (α-thalassemia major). Infants with Bart hemoglobin hydrops fetalis are pale and edematous and have hepatomegaly, splenomegaly, and ascites. Individuals with hemoglobin H disease have typical facies and bone changes seen in β-thalassemia, splenomegaly, and hepatomegaly.

β-Thalassemia occurs mainly in individuals of Mediterranean descent and presents as thalassemia major, intermedia, or minor. It is also seen in the Middle East, parts of India and Pakistan, and throughout Southeast Asia. Untreated patients with thalassemia major have skull bone deformities from intramedullary and extramedullary bone marrow expansion, mongoloid facies, bowing and rarefaction of long bones, extension of bone marrow into paraspinal or intraabdominal tumors, icterus, hepatomegaly, splenomegaly, and cardiac failure or endocrinopathies such as diabetes mellitus and hypogonadism from excessive intestinal iron absorption. Patients with thalassemia intermedia show fewer effects of iron overload, growth retardation, marrow expansion, and splenomegaly; however, deforming bone and joint disease, chronic leg ulceration, and infection are common in this form of thalassemia. Thalassemia minor is usually relatively asymptomatic.

Treatment. Because the carrier states for the thalassemias can be identified and affected fetuses can be diagnosed by DNA analysis after the ninth to tenth week of gestation, these conditions are widely amenable to prenatal diagnosis. Currently, bone marrow transplantation is the only way in which they can be cured. Symptomatic management is based on regular blood transfusions, iron chelation therapy, and the judicious use of splenectomy. Children with thalassemia are treated with blood transfusion therapy to maintain a hemoglobin level of 11 to 13 g/dL to ensure normal growth and development and to avoid skeletal deformities. Patients should be tested for the presence of hepatitis B antibodies and immunized if they test negative. Splenectomy is recommended in children 6 to 7 years of age or in adolescents when their transfusion requirements exceed 1.5 times normal (>200 mL/kg/yr). Before splenectomy, children should be immunized with polyvalent pneumococcal vaccine, *Haemophilus influenzae,* and *Neisseria meningitidis.* Iron supplements are avoided, and chelation therapy is started when the serum ferritin levels reach 1000 μg/dL. Vitamin C is required for management of iron overload. Bone marrow transplantation has been used with success in severe β-thalassemia. The best candidates are younger children, because older children have high rejection and mortality rates. Two experimental approaches are currently being pursued in the search for more effective therapy of the thalassemias: (1) reactivation or augmentation of fetal hemoglobin production and (2) somatic gene therapy.

Course and prognosis. Infants with Bart hemoglobin hydrops fetalis inherit a defective α-thalassemia gene from both parents who only have

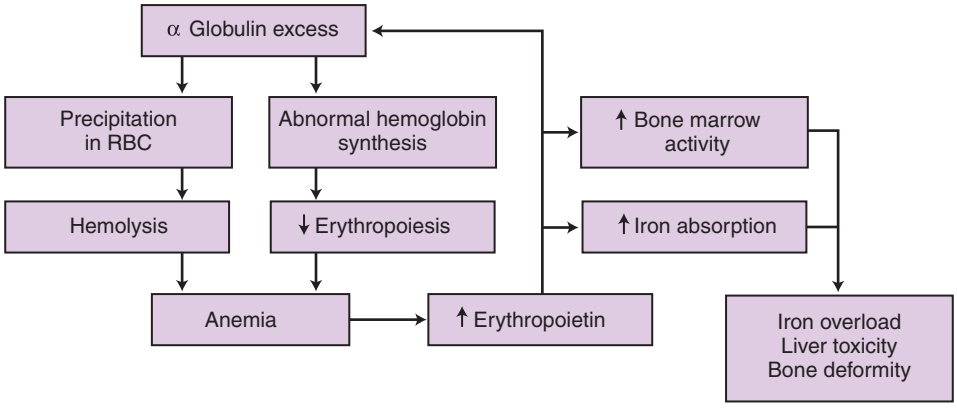


FIG 13.18 Pathophysiology of β-thalassemia. The deficiency of β globulin results in an excess of unincorporated α globulin.

the α-thalassemia trait. These infants are usually stillborn or die within hours to days of birth. Some patients with hemoglobin H disease live a full life. Patients with β-thalassemia intermedia can expect to live until middle age; however, iron loading and crippling bone disease occur in the third and fourth decades. Children with adequate treatment with iron chelation before bone marrow transplantation have disease-free survival rates up to 95%, whereas older patients and those exhibiting more than one risk factor have a rejection-free survival rate of less than 75%. Treatments under investigation include manipulation of globin gene expression with drugs such as 5-azacytidine, hydroxyurea, erythropoietin, or butyrate analogs and gene therapy directed at replacing or compensating for the defective β-globin alleles. Because this is a genetically transmitted disease, it is important for patients and parents to receive appropriate genetic counseling.

Sickle Cell Anemia

Etiology and pathogenesis. Sickle cell anemia is a genetically determined defect of hemoglobin synthesis. Sickle cell disease is a disorder in which patients inherit specific mutated variants of the β-globin gene that lead to hemoglobin polymerization. The sickle mutation of the β-globin gene results in the production of an abnormal hemoglobin called *sickle hemoglobin S (HbS)*. In hemoglobin S, valine is substituted for glutamic acid in the sixth position of the β chain, rather than the normal configuration. This apparently minor change in the molecular structure causes profound changes in hemoglobin stability and solubility. Under decreased oxygen tension, hemoglobin S undergoes polymerization, which causes the red cell to assume a sickled shape (Fig. 13.19). Patients who are homozygous produce only hemoglobin S. No hemoglobin A is synthesized because all the β chains are S chains, which combine with normal α chains to form hemoglobin S. In heterozygous patients with sickle cell trait, both normal and S chains are formed. Because fewer abnormal chains are produced than normal ones, the amount of hemoglobin A usually exceeds that of hemoglobin S. The sickle mutation has undergone positive selection during human evolution because individuals with one copy of the sickle gene and one normal β-globin gene (sickle cell trait) have a survival advantage in malaria-endemic regions. The preferential sickling of cells with malarial parasites reduces the number of parasites and allows children with sickle cell trait who are infected with these parasites to reach reproductive age. This has provided a selective advantage to the hemoglobin S trait, thereby preventing S from being genetically eliminated.

The pathogenetic signs and symptoms of sickle cell disease all relate to the red cell sickling. Sickled red cells have a decreased survival time, which causes anemia, and sickled cells cause vascular occlusion, which results in capillary stasis, venous thrombosis, and arterial emboli. The most dangerous feature of sickle cell anemia is the occurrence of acute episodes of "crisis," which can be hemolytic or vascular (Box 13.2).

Laboratory features. The laboratory features in sickle cell anemia are distinctive. The anemia is usually severe, with red cells of different shapes and sizes. Target red cells are present, and occasionally sickled cells can be seen on smears (Fig. 13.20). Red cell breakdown products are increased, which increases serum bilirubin, urobilinogen, and urobilin levels. Acute hemolytic crisis is characterized by hemoglobinuria, leukocytosis, and normoblastosis; diffuse intravascular coagulation may develop.

Clinical manifestations. Chronic hemolytic anemia, recurrent painful episodes, and acute and chronic organ dysfunction particularly of the spleen, bones, brain, kidneys, lungs, skin, and heart are the cardinal features of sickle cell anemia. Sickle cell anemia and sickle cell trait are found almost entirely in the black race. Hemolysis of the sickle cells occurs in the spleen or vascular space, and vasoocclusive events occur in the small capillaries and venules caused by sickle cells. The red cell life span is already shortened by the sickling and may precipitate a hemolytic crisis with jaundice. Sudden massive pooling of red cells, particularly in the spleen, can create a sequestration crisis, which is thought to result in the deaths that occur in the first years of life. Infarctive crises or painful episodes are a result of obstruction of blood vessels, tissue hypoxia, and tissue death and may occur throughout the body. Vasoocclusive events are described in Table 13.7. Children with sickle cell anemia are shorter and experience delayed puberty, but they attain normal height with late adolescent growth. Bony abnormalities, "hand–foot" syndrome with periostitis of the metacarpal and metatarsal bones, splenomegaly, inability to concentrate urine, priapism with subsequent impotence, underdeveloped genitalia and hypogonadism, hepatomegaly, jaundice, gallstones, tachycardia, acute chest syndrome (fever, chest pain, increasing WBC count, and pulmonary infiltrates), retinal vessel obstruction, cerebrovascular accidents, leg ulcers, and infections are all seen in sickle cell disease patients. Pregnant women may exhibit signs of pyelonephritis, pulmonary infarction, pneumonia, antepartum hemorrhage, premature fetal delivery, and fetal death.

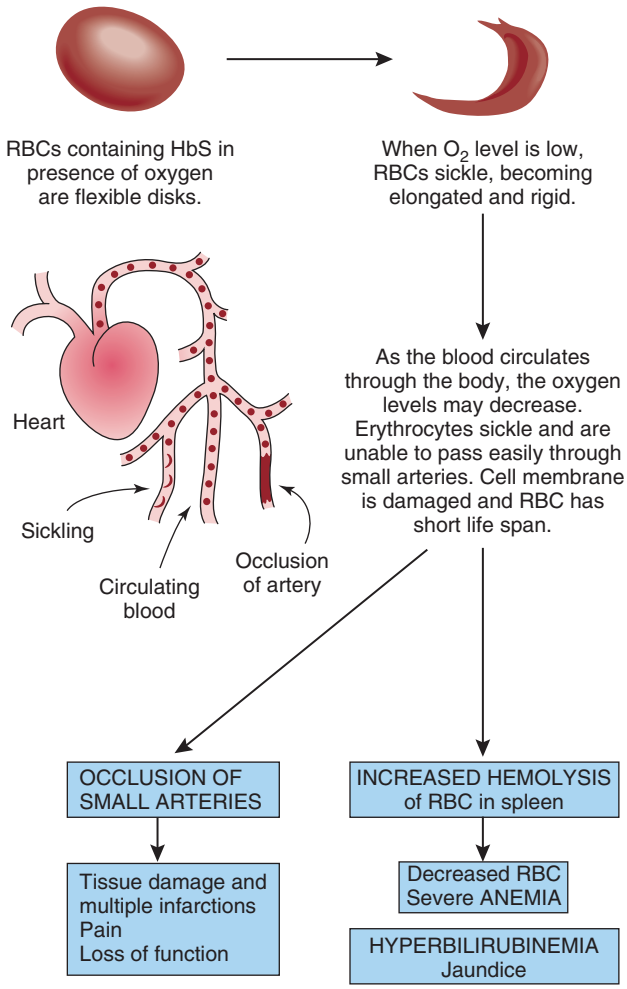


FIG 13.19 Sickle cell anemia: effects of sickling. *HbS,* Hemoglobin S; *RBC,* red blood cell. (From Gould BE: *Pathophysiology for the health professions,* ed 5, Philadelphia, 2014, Saunders, p 199.)

Treatment. Stem cell transplantation is curative and the treatment of choice. Currently there are no safe, effective antisickling agents, and treatment is primarily supportive. To avoid precipitation of a vasoocclusive crisis, it is important to prevent dehydration, infection, fever, acidosis, hypoxemia, and cold exposure. Because salicylates impose

BOX 13.2 Complications of Sickle Cell Anemia

Decreased RBC Survival
Anemia
Reticulocytosis
Hyperbilirubinemia
Increased pigment excretion
Cholelithiasis
Hyperplastic bone marrow
Osteoporosis
Osteosclerosis
Siderosis

Acute Hemolytic Crisis
Leukocytosis
Reticulocytosis
Hyperbilirubinemia
Hemoglobinuria
Normoblastosis
Diffuse intravascular coagulation (consumption coagulopathy)

Vascular Occlusion (Capillary Stasis, Venous Thrombosis, Arterial Emboli)
Splenomegaly
Splenic infarction
Splenic atrophy
Hepatomegaly
Cirrhosis
Hematuria
Sickle cell dactylitis
Aseptic necrosis of bones
Infarction of bone marrow
Infarction of various organs (brain)
Priapism
Skin ulcers
Pulmonary embolism

Painful Crisis (Occlusive Vascular Crisis)
Fever
Pain
Sudden death

Data from Miale J, editor: *Laboratory medicine hematology,* ed 6, St Louis, 1982, Mosby, p 637.

an acid load, acetaminophen is the preferred antipyretic. Vaccination for pneumococcal pneumonia should be done before 2 years of age in patients with sickle cell anemia and booster vaccinations given 3 to 5 years later. Penicillin prophylaxis is important to prevent streptococcal pneumonia and pneumococcal septicemia. Other vaccinations include *Haemophilus influenzae* type B and hepatitis B. Transfusions are used to restore normal hematocrit levels, and splenectomy is performed in children with sequestration syndrome. Treatment with oral hydroxyurea reduces leukocyte, polymorphonuclear, reticulocyte, and sickle cell counts while increasing the hemoglobin and hematocrit levels and resulting in fewer acute painful episodes. Appropriate treatment of pain is important. During pregnancy, folic acid should be given to prevent neural tube defects. If iron deficiency is present, iron supplements should also be administered. Transfusion should be used only when clinical and hematologic indicators are present.

Course and prognosis. Successful bone marrow stem cell transplantation cures sickle cell anemia. Bone marrow transplantation from a sibling-matched donor has a 94% survival rate and an 84% event-free survival rate. Sickle cell anemia is a serious disorder, and without stem cell transplantation many patients die in childhood, especially in sequestration crisis. In young children with sickle cell anemia, there is a 30% incidence of splenic sequestration crisis with a 15% death rate. Functional hyposplenia predisposes individuals to infections, such as pneumonia and chronic pyelonephritis with renal failure. Heart failure, bone marrow and fat emboli, shock, and organ failure are common causes of death. In developed countries, patients who have not undergone stem cell transplantation may survive into the third and fourth decades, whereas survival past childhood in underdeveloped countries is unusual. The survival rates have increased dramatically because of stem cell transplantation, newborn screening, early diagnosis, preventive measures to avoid sequestration crisis, and patient education.

Hereditary Spherocytosis

Etiology and pathogenesis. In hereditary spherocytosis, the red cells have defective red cell membrane skeletons, altered membrane properties, and altered cell metabolism. This causes them to have a decreased survival time in patients with an intact spleen. The disease is inherited as an autosomal-dominant trait and is characterized by red cells that are fragile microspherocytes. In addition, there is increased destruction of spherocytes (abnormal spherical erythrocytes) in the spleen. Patients have anemia, intermittent jaundice, splenomegaly, and uniform responsiveness to splenectomy. The principal cellular defect is a loss of membrane surface area attributable to defects of several membrane proteins, including ankyrin, band 3, α-spectrin, and β-spectrin.

Laboratory features. The concentration of hemoglobin within the red cells is increased. Reticulocytosis is present, and microspherocytes are seen on the blood smear. Osmotic fragility is increased, and serum unconjugated bilirubin level is increased. After splenectomy, the hemoglobin level is in the high-normal range.

Clinical manifestations. Hereditary spherocytosis is the most common hereditary hemolytic anemia and is most common in people with a northern European background. The major clinical manifestations are anemia, jaundice, splenomegaly, bile pigment gallstones, and chronic leg ulcers. The anemia is usually mild because of compensation by the erythropoietic bone marrow cells. Aplastic crisis precipitated by an infection may be seen with associated fever, abdominal discomfort, nausea, vomiting, rapidly increasing weakness, pallor, tachycardia, low blood pressure, and shock.

Treatment. Treatment usually consists of splenectomy in children with severe cases (hemoglobin concentration <8 g/dL and reticulocyte count >10%) and children with moderate disease reflected (hemoglobin concentration 8 to 11 g/dL and reticulocyte count of 8% to 10%) when the anemia compromises physical activity. All children undergoing splenectomy should be immunized with polyvalent pneumococcal vaccine, *Haemophilus influenzae,* and *Neisseria meningitidis.* Oral penicillin should be given for several years after splenectomy to prevent pneumococcal sepsis. Folic acid therapy to prevent folate deficiency is necessary as well. Transfusion is usually indicated only in aplastic crisis.

Course and prognosis. Most patients have no or mild anemia, fluctuating degrees of jaundice, and episodes of aplastic or hemolytic anemia. Splenectomy is usually curative; however, the subsequent risk of acquiring a serious infection is significant.

Glucose-6-Phosphate Dehydrogenase Deficiency

Etiology and pathogenesis. An example of an RBC intracellular defect caused by an enzyme deficiency is glucose-6-phosphate

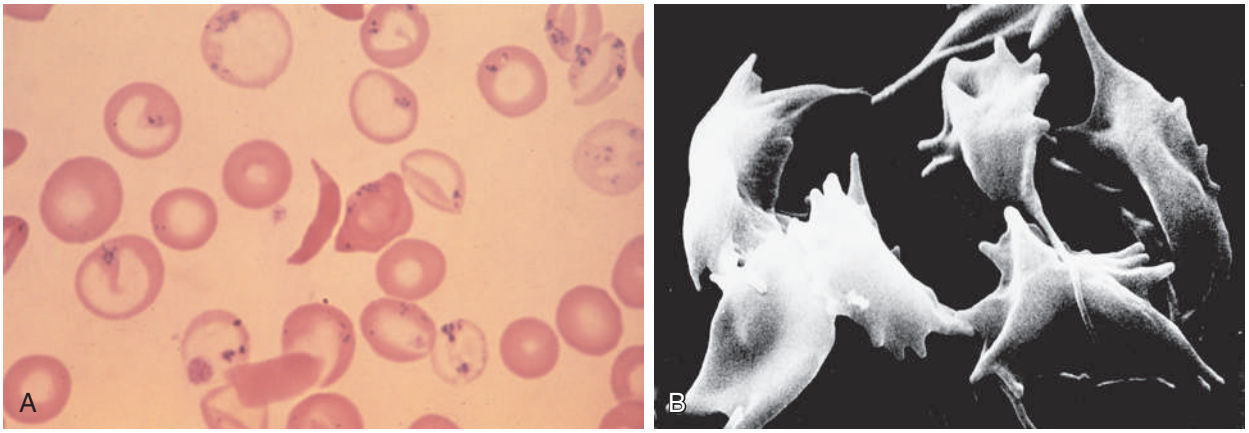


FIG 13.20 A, Blood smear showing sickle cells in sickle cell anemia. **B,** Scanning electron micrograph of deoxygenated sickled red cells. (A, Courtesy Beth Payne, Sacred Heart Medical Center, Spokane, WA. B, From Young NS et al, editors: *Clinical hematology,* Philadelphia, 2006, Mosby, p 39.)

TABLE 13.7 Vasoocclusive Consequences of Sickle Cell Disease

Event	Incidence	Features
Acute		
Painful episodes	>50% of patients with HbSS and HbS β-thalassemia	Mild to severe pain; one or several areas
Chest syndrome	10%–20% of adults	Difficult to distinguish from pneumonia; may involve entire lung
Priapism	10%–40% of males	Can have a more chronic form; causes impotence
Cerebrovascular accidents	1%–10% of children	Usually subarachnoid bleeding in adults
Hepatopathy	<2% of adults	Bilirubin may reach >80 mg/dL
Chronic		
Aseptic bone necrosis	10%–25% of adults	Hips and shoulders, common in HbSC
Proliferative retinopathy	50% of adults with HbSC; <5% HbSS 10%	Can lead to retinal detachment
Leg ulcers		Can be severe and disabling
Functional asplenia and autosplenectomy	Starts in infancy; >90% of adults with HbSS	Predispose to sepsis
Nephropathy	Renal failure in older patients	Nephritic syndrome, renal failure

Adapted from Sternberg MH: Hemoglobinopathies and thalassemias. In Stein JH et al, editors: *Internal medicine,* ed 5, St Louis, 1998, Mosby, p 658.
HbS, Hemoglobin S (sickle hemoglobin); *HbSC,* hemoglobin SC disease; *HbSS,* hemoglobin SS (sickle cell anemia).

dehydrogenase (G6PD) deficiency. The energy required for RBC membrane function and cellular integrity is derived from the anaerobic metabolism of glucose. Traditionally, hemolytic anemias caused by enzyme deficiencies have been called *nonspherocytic* to distinguish them from classic hereditary spherocytosis. When black soldiers receiving the antimalarial drug primaquine began suffering hemolytic episodes, a type of hemolytic anemia caused by a deficiency of G6PD (an enzyme in the red cell glycolytic pathway) was discovered. When G6PD-deficient RBCs are challenged by one of several drugs, glutathione is depleted and glucose utilization is inhibited. These events cause RBC membrane damage, which results in removal of the damaged cells by mononuclear phagocytes. Except in rare instances, G6PD-deficient persons do not have hemolytic anemia unless challenged by drugs. G6PD deficiency is the most common metabolic disease of the RBC, affecting hundreds of millions of people worldwide. This gene is found in 11% of African American males and in Sephardic Jews. The responsible gene is an X-linked recessive gene, so close relatives of affected individuals should be screened. Because G6PD deficiency is found in areas where malaria was once endemic, G6PD deficiency is thought to have conferred selective advantage against *Plasmodium falciparum* malaria infection.

Laboratory features. Usually this anemia is first recognized during or after an infectious illness or after exposure to a suspect drug or chemical. The hematologic tests reflect the severity of the hemolytic episode. The diagnosis of G6PD deficiency is based on the generation of NADPH from NADP as detected either by quantitative spectrophotometric analysis or by a rapid fluorescent screening test.

Clinical manifestations. Most individuals have no clinical manifestations of this disease. When such manifestations occur, hemolytic anemia is triggered by drug administration, infection, diabetic acidosis, the newborn period, and, in one subset, exposure to fava beans.

Treatment. Treatment is usually preventive and consists of avoidance of drugs that trigger hemolytic episodes and aggressive infection management. Some patients may require transfusion therapy or exchange transfusion in the case of life-threatening hemolysis.

Course and prognosis. The prognosis is generally good because the episodes of hemolytic crisis are usually self-limiting, except in fava bean–susceptible individuals, in whom shock may develop in a short time.

ANEMIA RELATED TO EXTRINSIC RED CELL DESTRUCTION OR LOSS

The final category of types of absolute anemia includes those caused by extrinsic abnormalities. The most important of these category types is immune hemolytic anemia caused by antibodies to red cells. Immune hemolytic anemias are further subdivided into those caused by isoantibodies, which may be the result of accidental immunization of individuals (e.g., hemolytic disease of the newborn [HDNB]), and those caused by autoantibodies (in individuals whose bodies create antibodies against their own red cells).

Hemolytic Disease of the Newborn

Etiology and pathogenesis. When fetal red cells cross the placenta, they may stimulate the production of maternal antibodies against antigens on the fetal red cell not inherited from the mother. These maternal antibodies cross into the fetal circulation and cause destruction of fetal cells. In severe cases, hydrops fetalis may result. Fetal-maternal ABO incompatibility is the most common cause of HDNB, but Rh incompatibility is clinically more important because of the severity of the hemolytic disease in the fetus. With the introduction of Rh treatment, the total incidence of HDNB in Rh-negative women has been greatly reduced.

Laboratory features. Anemia, reticulocytosis (an increased number of circulating reticulocytes), and nucleated red cells are seen in the peripheral blood of the infant. There is a rough correlation between the hemoglobin levels and the severity of the disease. Untreated infants may experience a rapid drop in hemoglobin levels after birth. Leukocytosis is present, but platelet counts are usually normal. Infants with severe disease may have thrombocytopenia. Serum bilirubin, a hemolytic breakdown product, is readily transferred across the placenta. At birth, the infant's total bilirubin level reflects both the severity of the hemolytic process and the ability of the infant's liver enzyme system to conjugate and excrete bilirubin. Cord blood red cells show a characteristic positive direct antiglobulin test (Coombs test), reflecting the maternal antibodies attached to the infant's red cells.

During pregnancy, laboratory tests of amniotic fluid for bilirubin and antibodies and tests of the mother's peripheral blood for maternal sensitization are useful in predicting whether infants will be affected by HDNB.

Clinical manifestations. The clinical manifestations of HDNB are hemolytic anemia, extramedullary erythropoiesis, and hyperbilirubinemia. Jaundice, petechial hemorrhages, hepatomegaly, splenomegaly, heart failure (with pulmonary edema, pleural effusions, ascites, and edema), kernicterus (a condition in the newborn marked by severe neural symptoms associated with high levels of bilirubin in the blood), and diffuse intravascular coagulation are seen in these infants. Many infants die in utero.

Treatment. A standard dose of anti-Rh immunoglobulin (RhoGAM) is given to the mother before or after delivery. This immunoglobulin destroys the infant's RBCs before they can sensitize the mother. This dose protects the mother against 30 mL of Rh-positive blood. Amniocentesis and fetal blood sampling are used to evaluate the severity of the disease. In severe cases, in utero transfusion and early delivery have been performed on fetuses with severe erythroblastosis. Exchange transfusion lowers the serum bilirubin level and the antibody content of the neonatal blood and removes cells susceptible to hemolysis. Phototherapy and phenobarbital are used to lower the bilirubin level.

Course and prognosis. The consequences of HDNB range from death, to possible retardation, to a barely perceptible hemolytic process. Severe anemia correlates with equally severe hyperbilirubinemia and high risk of central nervous system complications. Many infants appear normal at birth, only to develop jaundice within 2 to 3 hours. Petechial hemorrhages develop soon after birth, and kernicterus is usually seen late in the second day of significant jaundice. Successful RhoGAM administration prevention programs have reduced the perinatal death rate to about 1% to 2%.

Antibody-Mediated Drug Reactions

Etiology and pathogenesis. Drug-induced immune hemolytic anemia is an example of a disease in which exposure to a drug causes destruction and lysis of the sensitized person's own red cells. Drugs can lead to red cell hemolysis by four different immune mechanisms (Table 13.8).

Hapten mechanisms. In the hapten mechanism, which is seen with penicillin, cephalosporins, and tetracycline, the drug combines with a component of the RBC membrane. An antibody is developed against the drug. When the drug is given again, it coats the red cells, and the antibody attaches to the drug–red cell complex. The antigen–antibody complex then causes hemolysis.

Neoantigen formation. The old terminology for neoantigen formation is *immune complex formation.* In this situation, the drug combines with the RBC membrane and the antibody reacts with the new antigenic sites created by the combination of the drug and membrane. The RBC is hemolyzed. The immune complex can also bind to platelet and leukocyte

TABLE 13.8 Mechanisms of Drug-Induced Hemolysis or Positive Direct Antiglobulin Test

	Drug Absorption	Neoantigen	Autoimmune	Nonimmune Absorption
Prototype drug	Penicillin	Quinidine/stibophen	α-Methyldopa	First-generation cephalosporins
Role of drug	Cell-bound hapten	Antibody binds drug + RBC	Induces drug-independent RBC antibody	Modifies RBC membrane; absorbs proteins nonantibody-specifically
Typical DAT	IgG	C3	IgG	Nonimmunoglobulin
Antibody reactions	Reacts only with drug-coated cells	Reacts only with drug present	Drug independent; panagglutinin	No antibody present
Typical clinical presentation	Subacute onset; mild to severe hemolysis	Acute onset; severe hemolysis	Insidious onset; chronic mild hemolysis	No hemolysis

From Greer JP et al, editors: *Wintrobe's clinical hematology,* ed 11, Philadelphia, 2004, Lippincott Williams & Wilkins, p 1176.
C3, Complement third component; *DAT,* direct antiglobulin test; *IgG,* immunoglobulin G; *RBC,* red blood cell.

membranes, causing anemia, leukopenia, and thrombocytopenia. Quinidine, hydrochlorothiazide, sulfonamides, isoniazid, tetracycline, and cephalosporin are common drugs that cause this type of reaction.

Membrane modification. In membrane modification, seen in cephalosporin sensitivity, the drug alters the RBC membrane protein. Plasma proteins attach to the altered RBC protein and cause a positive serologic test but no cell hemolysis.

Autoantibody induction. This mechanism was first studied in cases of hemolytic anemia with patients who were taking the antihypertensive agent methyldopa (Aldomet). The drug appears to induce antibody formation to red cell membrane Rh antigens. About 29% of the patients receiving this drug develop a positive antiglobulin test response.

Laboratory features. The laboratory features for all mechanisms show increased red cell turnover and anemia if hemolysis exceeds the rate of RBC production. Serologic tests, such as the direct antiglobulin test, will be positive. In hapten antibody–mediated drug reactions and in immune complex formation, the antiglobulin reaction is positive for immunoglobulins. In autoantibody induction, the antiglobulin reaction is positive for complement. Fragmented RBCs may be seen on the peripheral blood smear. These fragments are called *schistocytes* (Fig. 13.21). Leukopenia and thrombocytopenia are sometimes seen with drug-induced platelet or leukocyte destruction.

Clinical manifestations. Types of immune drug-induced hemolytic anemia vary in symptoms and severity, depending on the mechanism involved. Hapten (e.g., penicillin) and autoimmune (e.g., methyldopa) drug-induced hemolytic anemias have an insidious onset of symptoms over a period of weeks. The neoantigen formation (e.g., quinine or quinidine) may present with sudden, severe hemolysis with hemoglobinuria and result in acute renal failure. Other clinical manifestations include acute respiratory distress syndrome and respiratory arrest.

Treatment. Recognition and discontinuation of the responsible drug are usually the only treatment necessary. Steroid therapy and transfusions may be required in cases of severe hemolysis.

Course and prognosis. Immune hemolytic anemia attributable to drugs is usually mild and the prognosis is good; however, with severe hemolysis, death can occur. Laboratory findings for erythrocyte disorders are summarized in Table 13.5.

Acute Blood Loss

Etiology and pathogenesis. Acute blood loss anemia may present after trauma or secondary to a disease process. Acute blood loss anemia rapidly decreases the overall blood volume and impairs oxygen delivery.

Laboratory features. A decrease occurs in both hematocrit level and hemoglobin concentration attributable to blood loss. The hematocrit level is less than 40% in men and less than 37% in women. The hemoglobin concentration is less than 14 g/dL in men and less than 12 g/dL in women. Anemia may not be apparent in the early stages because the cells and plasma are diminishing simultaneously. As replacement fluids move into the intravascular space, the anemia becomes apparent in later laboratory tests. The cells have normal MCV, MCH, and MCHC values.

Clinical manifestations. In a normal 70-kg person with a 5000-mL total blood volume, 10% loss of blood (500 mL) rarely causes any clinical signs except occasional vasovagal syncope. A 20% loss (1000 mL) usually causes no clinical symptoms at rest, but tachycardia is seen with exercise, and a slight postural drop in blood pressure occurs. A person with a 30% loss (1500 mL) usually presents with flat neck veins when supine, postural hypotension, and exercise tachycardia. A 40% loss (2000 mL) causes the central venous pressure, cardiac output, and arterial blood pressure to fall below normal while the patient is supine and at rest, with associated air hunger, tachycardia, and cold, clammy skin. A 50% loss of total blood volume (2500 mL) often causes shock and death.

Treatment. Blood volume replacement therapy with crystalloid solutions, colloid solutions (plasma protein, albumin, or dextran), and fresh whole blood is essential in the early management of acute hemorrhage to restore blood volume and to prevent shock. Complete reliance on fresh whole blood for managing acute blood loss is contraindicated and should be reserved for patients with a low red cell mass in whom tissue hypoxia is a threat. Replacement of red cell mass by increased red cell production is a gradual process, which occurs over 2 to 5 days as the marrow stem cells proliferate and mature. Maximal red cell production is seen by the tenth day after hemorrhage.

Course and prognosis. With adequate replacement therapy, the prognosis is excellent; however, the underlying cause must be identified and managed.

Other Extrinsic Abnormalities

Other mechanisms, such as mechanical heart valves or cardiopulmonary bypass machines, may cause physical damage to the red cells, resulting in hemolysis. Drugs and chemicals, physical agents (e.g., burns), or infectious diseases (e.g., malaria) may result in anemia. Venom from bee and wasp stings, spider and scorpion bites, and snake bites has been associated with hemolytic anemia. Finally, hypersplenism and splenomegaly can cause anemia, leukopenia, or thrombocytopenia severe enough to require splenectomy.

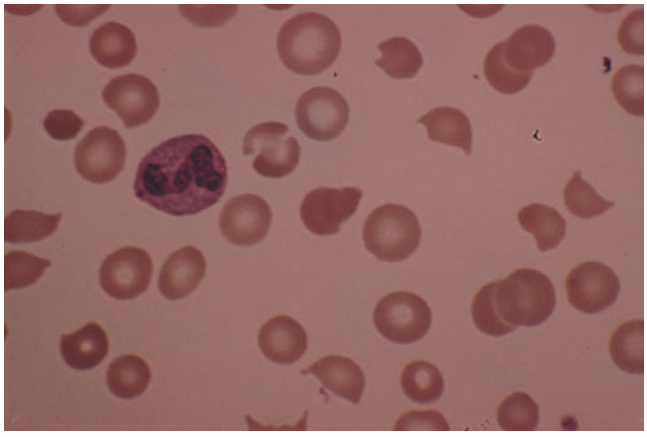


FIG 13.21 Schistocytes are fragments of red blood cells produced by hemolytic pathologies. (Courtesy Beth Payne, Sacred Heart Medical Center, Spokane, WA.)

KEY POINTS

- The general effects of anemia are due to tissue hypoxia and efforts to compensate for low oxygen-carrying capacity. Vasoconstriction, pallor, tachypnea, dyspnea, tachycardia, ischemic pain, lethargy, and lightheadedness may be present. In addition, signs and symptoms relating to the specific cause of the anemia may be present. These accompanying manifestations are helpful in determining the cause of the anemia.
- Anemia may be due to abnormally low production of red cells and/or excessive loss or destruction. Decreased production of red cells may be due to stem cell failure (aplastic anemia); lack of erythropoietin (renal disease); or nutritional deficiencies of iron, vitamin B_{12}, or folate. Excessive red cell loss may be due to hemolysis (e.g., ABO and Rh incompatibility, drugs) or bleeding (e.g., surgery, trauma). Inherited disorders of red cells often impair production and increase destruction of red cells.

- Determination of the cause of anemia is based on the history, differential signs and symptoms, and results of laboratory studies. The important differentiating features of the major types of anemia are as follows:
  - *Aplastic anemia:* History of toxic or radiation injury to bone marrow. Accompanying leukopenia and thrombocytopenia. Red cells are normocytic and normochromic.
  - *Chronic renal failure:* History of renal disease. Decreased erythropoietin level and erythropoietin responsiveness. Red cells are normocytic and normochromic.
  - *Vitamin B_{12} and folate deficiency:* History of poor nutrient intake or GI disease. Accompanying neurologic dysfunction. Red cells are megaloblastic (macrocytic).
  - *Iron deficiency:* History of poor nutrient intake or chronic blood loss. Decreased serum ferritin and iron levels. Red cells are microcytic and hypochromic.
  - *Hemolytic:* History of ABO or Rh incompatibility or drug exposure. Increased bilirubin level, jaundice, positive direct antiglobulin test. Red cells are normocytic and normochromic.
  - *Acute blood loss:* History of trauma, surgery, or known bleeding. Accompanying manifestations of volume depletion. Red cells are normal. Anemia may not be apparent until fluid loss is replaced.
- Inherited disorders of the red cell (thalassemia, sickle cell anemia, spherocytosis, G6PD deficiency) predispose red cells to early destruction because of abnormalities in hemoglobin structure, cell shape, membrane structure, or energy production. Manifestations of hemolysis (e.g., bilirubin, jaundice) are often present.
- The general management of anemia is aimed at removing the cause, if possible; restoring oxygen-carrying capacity with blood transfusion when necessary; and preventing the complications of ischemia (e.g., with rest, oxygen therapy) and hemolysis (e.g., increased fluid intake, management of high bilirubin levels).

TRANSFUSION THERAPY

Medical indications for transfusion therapy are restoration or maintenance of oxygen-carrying capacity, blood volume, hemostasis, and leukocyte function. Red cell transfusions are administered to improve tissue oxygenation in the context of anemia or acute blood loss. Adaptive responses to a declining hemoglobin concentration include increased cardiac output, augmented oxygen extraction, blood flow redistribution to the heart and brain, a right shift in the oxyhemoglobin dissociation curve, and increased red cell production by the marrow. These compensatory mechanisms help ensure continued oxygen delivery. A summary of blood components, indications, actions, contraindications, precautions, and hazards is presented in Table 13.9. Before transfusion therapy can occur, various donor tests are performed on the blood unit sample. These include ABO and Rh(D); syphilis; HIV antigen and antibodies; hepatitis B and C antigens; and human T-cell lymphotropic virus. Blood centers that produce plasma for fractionation also test for alanine aminotransferase. Specific pretransfusion testing using blood samples from the recipient and the donor unit must be done to ensure that the blood component will not harm the recipient and that the blood component will have an acceptable survival time when transfused. ABO and Rh typing and RBC antibody detection tests are performed, and then a cross-match between the donor unit and the recipient is performed. Types of transfusion reactions, signs and symptoms, usual causes, treatment, and precautions are summarized in Table 13.10. Two serious complications of transfusion therapy include transfusion-related acute lung injury (TRALI) and transfusion-related circulatory overload (TACO). TRALI is a syndrome of acute hypoxia as a result of noncardiogenic pulmonary edema that follows transfusion. All blood components have been implicated in TRALI, but plasma-containing products are more prevalent. Supportive care is the mainstay of therapy in TRALI with (1) oxygen supplementation, (2) aggressive respiratory support, and (3) intravenous administration of fluids, as well as vasopressors, which are essential for blood pressure support. Corticosteroids can be beneficial, and diuretics should be avoided in TRALI. TACO is defined as an expansion in the intravascular volume when the infused blood components and/or other fluids exceed the cardiovascular ability to handle the additional workload, and results in congestive heart failure. Diuretics are given in TACO along with the supportive care noted previously.

POLYCYTHEMIA

In polycythemia, red cells are present in excess, increasing blood viscosity, which in turn causes clinical manifestations such as hypertension. The three types of polycythemia are classified according to cause. *Polycythemia vera* is associated with neoplastic transformation of bone marrow stem cells. *Secondary polycythemia* is due to chronic hypoxemia, with a resultant increase in erythropoietin production. *Relative polycythemia* is due to dehydration, which causes a spurious increase in the RBC count.

Polycythemia Vera

Etiology and pathogenesis. Polycythemia vera, or primary polycythemia, is a type of chronic panmyelosis (see Chapter 11) and is part of the spectrum of myeloproliferative disorders. Polycythemia vera arises from the transformation of a single stem cell into a cell with a selective growth advantage that gradually becomes the predominant source of marrow precursors. There is an overproduction of normal red cells, white cells, and platelets. As with many malignancies, the cause is unknown. Possible mechanisms for the proliferation include (1) unregulated neoplastic proliferation of stem cells, (2) presence of abnormal myeloproliferative factor acting on normal stem cells, and (3) increase of stem cell sensitivity to erythropoietin and other hematopoietins. Some researchers have postulated that it is damage to the undifferentiated stem cell by a virus, radiation, drugs, or other agents that causes mutation and subsequent neoplastic transformation

Laboratory features. The diagnosis depends primarily on results of laboratory studies, which show an absolute increase in red cell mass and leukocytosis and thrombocytosis. The bone marrow shows hyperplasia of red cells, white cells, and platelets and extension of active hematopoietic marrow into bones of the extremities. Uric acid concentration is increased because of excessive cell proliferation, which results in the destruction of an increased number of cells. Arterial oxygen saturation is normal, which differentiates polycythemia vera from the more common secondary (hypoxemic) polycythemia. Additional findings include increased erythropoietin, elevated serum vitamin B_{12}, and elevated leukocyte alkaline phosphatase levels (Fig. 13.22).

Clinical manifestations. Symptoms include headache, backache, weakness, fatigue on exertion, pruritus, dizziness, sweating, visual disturbances, weight loss, paresthesias, dyspnea, joint complaints, and epigastric distress and pressure. Common clinical manifestations include hypertension, thrombosis, and mucosal hemorrhage attributable to engorgement of retinal and sublingual veins, but each phase of the disease presents somewhat differently. Most of the clinical symptoms of polycythemia vera are related to the increased red cell mass, which gives rise to an increased blood viscosity. The liver and spleen become congested, which increases the risk of clots, acidosis, and organ infarction. The onset is insidious, with variable manifestations in virtually any organ system. Clinical symptoms appear between 60 and 80 years of age, and they appear more often in men and Caucasians. The disorder

Text continued on p. 292

TABLE 13.9 Summary of Blood Components

Component*	Major Indications	Properties of Component	Not Indicated for These Conditions	Special Precautions	Hazards	Rate of Infusion
Whole blood	Symptomatic anemia with large volume deficit.	Increases oxygen-carrying capacity. Increases blood volume.	Condition responsive to specific component. Treatment of coagulopathy.	Must be ABO identical. Labile coagulation factors deteriorate within 24 hr after collection.	Infectious diseases. Hemolytic, septic/toxic, allergic, and febrile reactions. TACO. TRALI. TA-GVHD.	For massive loss, as fast as patient can tolerate.
Whole blood irradiated	See *Whole blood.* Increased risk for TA-GVHD. (e.g., stem cell transplant, IUT, and selected immunodeficiencies, HLA-matched platelets or transfusions from blood relatives.)	See *Whole blood.* Donor lymphocytes are inactivated, reducing risk of TA-GVHD.	See *Whole blood.*	See *Whole blood.*	See *Whole blood.*	See *Whole blood.*
RBCs, RBCs (adenine-saline added)†	Symptomatic anemia; Red cell exchange transfusion.	Increases oxygen-carrying capacity.	Pharmacologically treatable anemia. Volume expansion.	Must be ABO-compatible.	Infectious diseases. Hemolytic, septic/toxic, allergic, and febrile reactions. Iron overload. TACO. TRALI. TA-GVHD.	For massive loss, as fast as patient can tolerate, but less than 4 hr.
RBCs, deglycerolized	See *RBCs.* IgA deficiency with anaphylactoid reactions	See *RBCs.* Deglycerolization removes plasma proteins. Risk of allergic and febrile reactions reduced.	See *RBCs.*	See *RBCs.*	See RBCs. Hemolysis due to incomplete deglycerolization can occur.	See *RBCs.*
RBCs, irradiated	See *RBCs.* Increased risk for TA-GVHD.	See *RBCs.* Donor lymphocytes are inactivated, reducing risk of TA-GVHD.	See *RBCs.*	See *RBCs.*	See *RBCs.*	See *RBCs.*
RBCs, leukocytes reduced; apheresis red blood cells, leukocytes reduced	See *RBCs.* Symptomatic anemia. Reduction of febrile reactions, HLA alloimmunization and CMV infection.	See *RBCs.* Increases oxygen-carrying capacity. Reduces risk of febrile reactions from leukocyte antibodies, HLA alloimmunization, and CMV infection.	See *RBCs.* Pharmacologically treatable anemia. Leukocyte reduction should not be used to prevent TA-GVHD.	See *RBCs.* Must be ABO-compatible. Hypotensive reaction may occur if bedside leukocyte reduction filter is used.	Infectious diseases. Hemolytic, septic/toxic, allergic and febrile reactions (unless plasma also removed [e.g., by washing]) TACO. TRALI. TA-GVHD.	For massive loss, as fast as patient can tolerate, but less than 4 hr.

Continued

TABLE 13.9 Summary of Blood Components—cont'd

Component*	Major Indications	Properties of Component	Not Indicated for These Conditions	Special Precautions	Hazards	Rate of Infusion
RBCs washed	See *RBCs*. IgA deficiency with anaphylactoid reaction. Recurrent severe allergic reactions to unwashed red cell products.	See *RBCs*. Washing reduces plasma proteins. Risk of allergic reactions is reduced.	See *RBCs*.	See *RBCs*.	See *RBCs*.	See *RBCs*.
Fresh-frozen plasma (FFP)	Clinically significant plasma protein deficiencies when no specific coagulation concentrates are available. TTP.	Source of labile and nonlabile plasma proteins, including all coagulation factors.	Volume expansion. Coagulopathy can be more effectively treated with specific therapy.	Must be ABO-compatible	Infectious diseases. Hemolytic, septic/toxic, allergic, and febrile reactions. TACO. TRALI. TA-GVHD	<4 hr
Liquid plasma, plasma, and thawed plasma	Bleeding patients with deficit of stable coagulation factors. TTP	Source of plasma proteins and nonlabile factors. Levels and activation state of coagulation proteins in thawed plasma are variable and change over time.	Volume expansion. Coagulopathy that can be more effectively treated with specific therapy.	Must be ABO-compatible.	Infectious diseases. Allergic reactions. TACO. TRALI.	<4 hr
Plasma, cryoprecipitate reduced	TTP	See *FFP*. Plasma protein replacement for plasma in FFP. Deficient in fibrinogen, vWF, and factors VIII and XIII. Deficient in high molecular weight vWF multimers compared with FFP	Volume expansion. Deficiency of coagulation factors known to be depleted in this product, fibrinogen, vWF, and factors VIII, and XIII.	Must be ABO-compatible.	See *FFP*.	<4 hr
Cryoprecipitated AHF; pooled cryoprecipitated AHF	Hypofibrinogenemia. Factor XIII deficiency. Second line therapy of von Willebrand disease, hemophilia A and uremic bleeding.	Provides fibrinogen, vWF, factors VIII and XIII‡.	Not indicated if specific concentrates are available. Deficiencies of any plasma protein other than those enriched in cryoprecipitated AHF.	Frequent repeat doses may be necessary.	Infectious diseases. Allergic, febrile reactions.	<4 hr
Platelets; platelets pooled	Bleeding from thrombocytopenia or platelet function abnormality, including antiplatelet drugs. Prevention of bleeding from marrow hypoplasia.	Improves hemostasis.	Plasma coagulation deficits. Some conditions with rapid platelet destruction (e.g., ITP, TTP) unless life-threatening hemorrhage.	Must be ABO-compatible with plasma. Should not use some filters (check manufacturer's instructions).	Infectious diseases. Hemolytic, septic/toxic, allergic, febrile reactions. TACO. TRALI. TA-GVHD	<4 hr

Platelets, apheresis‡	See *Platelets.*	See *Platelets.* May be HLA or other antigen selected.	See *Platelets.*	See *Platelets.*	See *Platelets.*	<4 hr
Platelets, irradiated; platelets, pooled irradiated; platelets, apheresis irradiated	See *Platelets.* Increased risk of TA-GVHD	See *Platelets.* Donor lymphocytes are inactivated, reducing risk of TA-GVHD.	See *Platelets.*	See *Platelets.*	See *Platelets.*	<4 hr
Platelets, leukocytes reduced; pooled platelets, leukocytes reduced; apheresis platelets, leukocytes reduced	See *Platelets.* Reduction of febrile reactions; reduction of HLA alloimmunization and CMV infection.	See *Platelets.* Reduction of leukocytes reduces risk of febrile reactions, HLA alloimmunization, and CMV infection.	See *Platelets.* Leukocyte reduction should not be used to prevent TA-GVHD.	See *Platelets.*	See *Platelets.*	<4 hr
Granulocytes, apheresis	Neutropenia with infection, unresponsive to appropriate antibiotics.	Provides granulocytes with or without platelets.	Infection responsive to antibiotics, eventual marrow recovery not expected.	Must be ABO-compatible. Should not use some filters (check manufacturer's instructions); do not use depth-type microaggregate filters.	Infectious diseases. Hemolytic, allergic, febrile reactions. TACO. TRALI. TA-GVHD. Maintain caution. Pulmonary reactions may occur in patients receiving concomitant amphotericin B.	One unit over 2–4-hr period. Observe closely for reactions
Granulocytes, pheresis irradiated; granulocytes, platelets irradiated	See *Granulocytes;* see *Platelets*	Provides granulocytes with or without platelets.	See *Granulocytes;* see *Platelets*	See *Granulocytes;* see *Platelets*	See *Granulocytes;* see *Platelets*	See *Granulocytes;* see *Platelets*

*For all cellular components there is a risk that the recipient may become alloimmunized.
†RBCs and platelets may be processed in a manner that yields leukocyte-reduced components for which the main indications are prevention of febrile, nonhemolytic transfusion reactions and prevention of leukocyte alloimmunization. Risks are the same as those for standard components, except for reduced risk of febrile reactions.
‡When virus-inactivated concentrates are not available.
A HF, Antihemophilic factor; *CMV,* cytomegalovirus; *FFP,* fresh-frozen plasma; *HLA,* human leukocyte antigen; *ITP,* idiopathic thrombocytopenic purpura; *TACO,* transfusion-associated circulatory overload; *TA-GVHD,* transfusion-associated graft-versus-host disease; *TRALI,* transfusion-related acute lung injury; *IUT,* intra uterine transfusion; *TTP,* thrombotic thrombocytopenic purpura; *vWF,* von Willebrand factor.

TABLE 13.10 Transfusion Reactions

Type	Signs and Symptoms	Usual Cause	Treatment	Precautions
Acute intravascular hemolytic (immune)	Hemoglobinemia and hemoglobinuria, fever, chills, anxiety, shock, disseminated intravascular coagulation (DIC), dyspnea, chest pain, flank pain, nausea/vomiting, headache, pain at needle site and along venous tract	Incompatibility because of clerical errors; involves ABO (primarily) or other erythrocyte antigen-antibody incompatibility	Stop transfusion; hydrate; support blood pressure and respiration; induce diuresis; treat shock and DIC	Positively identify donor and recipient blood types and groups before transfusion is begun; verify with one other nurse or physician. Transfuse blood slowly for first 15–20 min and/or initial one-fifth volume of blood; remain with patient. In event of signs or symptoms, stop transfusion immediately, maintain patent IV line, and notify physician. Save donor blood to re–cross-match with patient's blood. Monitor blood pressure for shock. Insert urinary catheter and monitor hourly outputs. Send sample of patient's blood and urine to laboratory to determine presence of hemoglobin (indicates intravascular hemolysis). Observe for signs of hemorrhage resulting from DIC. Support medical therapies to reverse shock.
Delayed extravascular hemolytic (immune)	Fever, malaise, indirect hyperbilirubinemia, increased urine urobilinogen, falling hematocrit and hemoglobin	Occurs in previously RBC-alloimmunized patients in whom antigen on transfused red cells provokes anamnestic production of antibody; destruction of RBCs; usually involves non-ABO antigen-antibody incompatibility occurring 2–14 days posttransfusion	Monitor hematocrit, renal function, coagulation profile; no acute treatment generally required	Observe for posttransfusion anemia and decreasing benefit from successive transfusion.
Graft versus host disease (GVHD) or TA-GHVD		Viable T lymphocytes react against tissue antigens in recipient Immunocompromised recipients most at risk		Use γ-irradiated components to prevent TA-GVHD.
Febrile	Fever, chills, rarely hypotension	Antibodies to leukocytes or plasma proteins	Stop transfusion; give antipyretics, acetaminophen (or aspirin if patient not thrombocytopenic)	Use of leukocyte-poor RBCs is less likely to cause reaction.
Allergic	Urticaria (hives), flushing, wheezing, laryngeal edema, rarely hypotension or anaphylaxis	Antibodies to plasma proteins	Stop transfusion; give antihistamine; if severe, give epinephrine and/or steroids	Administer pretransfusion antihistamine; use washed RBC components.
Hypervolemic or TACO	Dyspnea, rales, hypertension, pulmonary edema, cardiac dysrhythmias, precordial pain, cyanosis, dry cough, distended neck veins	Transfusion-associated circulatory overload from too rapid or excessive blood transfusion	Induce diuresis; phlebotomy; support cardiorespiratory system as needed	Transfuse blood slowly. Prevent overload by using packed RBCs or administering divided amounts of blood. Use infusion pump to regulate and maintain flow rate. If signs of overload, stop transfusion immediately. Place patient in semi-Fowler position to increase venous resistance.

TRALI	Acute onset of hypoxemia within 6 hr of a blood or blood component transfusion; dyspnea, pulmonary edema, normal cardiac pressures	Anti-HLA or antileukocyte antibodies	Support blood pressure and aggressive respiratory support that may require intubation and mechanical ventilation	Use washed RBCs; avoid unnecessary transfusion.
Hypothermia	Chills, low temperature, irregular heart rate, possible cardiac arrest	Rapid infusion of cold blood products	Monitor temperature; if markedly subnormal, stop transfusion	Allow blood to warm at room temperature (<1 hr). Use an electric warming coil to rapidly warm blood.
Electrolyte disturbances, hyperkalemia	Nausea, diarrhea, muscular weakness, flaccid paralysis, paresthesia of extremities, bradycardia, apprehension, cardiac arrest	Massive transfusions or in patients with renal problems	Kayexalate enemas if potassium >5.0 mEq/L	Use washed RBCs or fresh blood if patient at risk.
Citrate intoxication (hypocalcemia)	Tingling in fingers, tetany, muscular cramps, carpopedal spasm, hyperactive reflexes, convulsions	Massive transfusion of blood	Stop transfusion; administer IV calcium if severe	Infuse blood slowly (citrate reaction less likely to occur). If signs of tetany occur, clamp tubing immediately, maintain patent intravenous line, and notify physician.
Air emboli	Sudden difficulty in breathing, sharp pain in chest, apprehension, respiratory or cardiac arrest	Air emboli from blood administered under pressure	Stop transfusion; turn patient on left side; aspirate right atrial/ventricular air emboli	When infusing blood under pressure before container is empty: if air is observed in tubing, clamp tubing immediately below air bubble, clear tubing of air by aspirating air with syringe or disconnecting tubing and allowing blood to flow until air has escaped.
Bacterial sepsis	Shock, chills, high fever	Bacterial contamination of blood component or endotoxin reaction seen more commonly with platelet components stored at room temperature	Stop transfusion; support blood pressure; give antibiotics	Use care in blood collection and storage.
Delayed reactions, transmission of infection	Signs of infection after transfusion (e.g., jaundice from hepatitis; bacterial or toxin contamination—high fever, severe headache or substernal pain, hypotension, intense flushing, vomiting/diarrhea)	Hepatitis, AIDS, malaria, syphilis, bacteria, viruses, other	Stop transfusion; do culture and sensitivity tests; treat specific infection	Blood is tested for HBsAg (hepatitis B), syphilis, and, in most centers, HIV (AIDS); positive units are destroyed. Individuals at risk for carrying certain viruses are deferred from donation. Observe for signs of infection.

Red Cross, America's Blood Centers and the Armed Services Blood Program: *Circular of information for the use of human blood and blood components,* Washington, DC, December 2009, American Red Cross.

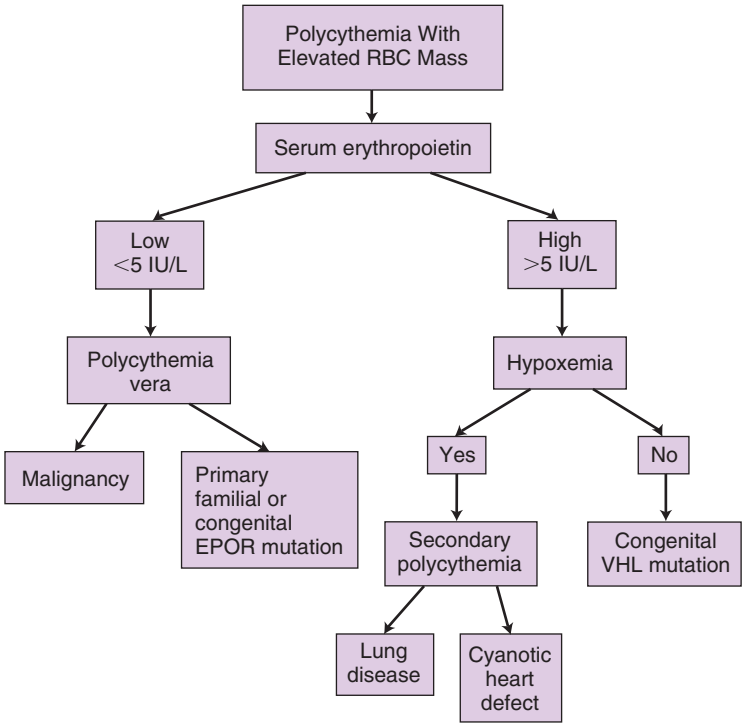


FIG 13.22 Algorithm for differentiation of polycythemia. *EPOR,* Erythropoietin receptor gene; *VHL,* von Hippel-Landau gene.

is rarely seen in children. In the preerythrocytic or developmental phase, hepatosplenomegaly, night sweats, and postbathing pruritus are common. Other patients experience mild thrombohemorrhagic symptoms or erythromelalgia (painful erythematous palms and soles from an increased number of circulating platelets).

The evolution of polycythemia vera is shown in Fig. 13.23. The phases include an asymptomatic phase, a plethoric or erythrocytic phase, an inactive phase, and a spent phase when anemia develops. The final evolutionary phase of polycythemia vera is that of acute myeloid leukemia.

In the erythrocytic phase, occlusive vascular lesions, such as transient ischemic attacks, cerebrovascular accidents (strokes), myocardial ischemia or infarctions, portal venous obstruction, or superficial venous thrombosis, occur and may be the first indication of the presence of the disease. The hyperviscosity produces symptoms of reduced cerebral blood flow, such as headaches, dizziness, and visual disturbances. Walking may induce leg pain and spasm, called *intermittent claudication.*

Mucosal hemorrhagic manifestations include epistaxis, ecchymosis, and GI and genitourinary bleeding. Progressive splenomegaly, intermittent claudication, peptic ulcer, hyperuricemia, and gout are often seen. The most striking feature is a ruddy or florid face, telangiectasias (chronic dilation of capillaries and small arterial branches, producing small, reddish tumors of the skin) of the cheeks and nose, and purplish cyanosis of the lips and ears. Hypertension is seen in about half of patients. Distention of the retinal veins with a dark purple coloration is another important clinical finding. As the disease develops into the spent or postpolycythemia myeloid metaplasia phase, many patients complain only of asthenia (weakness, fatigue); however, progressive hepatosplenomegaly, severe anemia, hemorrhage (particularly cutaneous), weight loss, and wasting often occur. The final phase is the development of acute myeloid leukemia.

Treatment. There is no cure. Treatment is directed at reducing the increased blood volume, blood viscosity, red cell mass, and platelet counts by use of phlebotomy and myelosuppressive therapy with radioactive phosphorus and chemotherapeutic agents. Phlebotomy of 450 to 500 mL every 2 to 4 days until a normal hematocrit level is reached alleviates many symptoms for most patients. Phlebotomy of only 200 to 300 mL should be considered for elderly patients or those with cardiovascular disease. In the past, a hematocrit of 50% was used as the upper limit of hematocrit tolerated before phlebotomy was used. Studies have found that increased vascular complications, decreased cerebral blood flow, and decreased mental alertness occurred when hematocrit levels exceeded 45%. Hematocrit levels should be maintained at 42% for females and 45% for males. Phlebotomy is effective in controlling red cell mass, but myelosuppressive therapy is needed when the platelet count increases to more than 800,000 to 1,000,000/μL to control hepatosplenomegaly and thrombocytosis. The agent of choice for myelosuppressive therapy is hydroxyurea, but radioactive phosphorus is also used. Hyperuricemia is treated with allopurinol (100 to 300 mg/day), pegylated interferon (90 to 180 μg/week), or interferon-α (3×106 units three times a week; alter dose depending on response and toxicity). Pegylated interferon, which can be administered once weekly, should be considered. As the disease progresses, thrombosis or hemorrhage, systemic symptoms, severe pruritus refractory to histamine antagonists, and painful splenomegaly occur, and a splenectomy is indicated.

Course and prognosis. Unmanaged polycythemia vera has a poor prognosis, with a survival of less than 2 years. The prognosis depends on the nature and severity of the complications, the duration of the erythrocytotic phase, and the duration of the acute myeloid leukemia phase. Treatment in the erythrocytotic phase is essential, or the patient is at extremely high risk for thromboses. The development of thrombosis, hemorrhage, and myeloproliferative syndromes is common. Treated patients have a median survival of 10 to 15 years, with the most common causes of death being thrombosis, hemorrhage, leukemia, and other myeloproliferative conditions.

Secondary Polycythemia

Etiology and pathogenesis. Secondary polycythemia is absolute erythrocytosis caused by increased stimulation of RBC production, usually in response to tissue hypoxia caused by, for example, high altitude

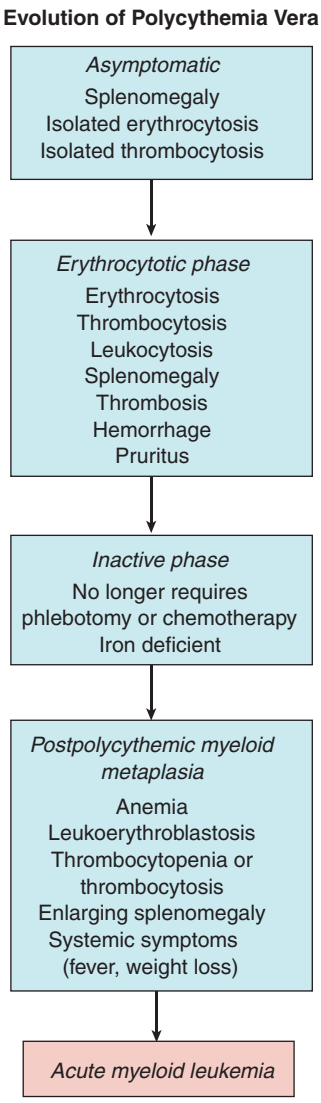


FIG 13.23 Clinical evolution of polycythemia vera. (From Hoffman R et al, editors: *Hematology: basic principles and practice,* ed 7, New York, 2018, Churchill Livingstone, p 1216.)

or lung disease. There are other less common types of secondary polycythemia that are caused by renal or other organ tumors, which cause an increase in erythropoietin production.

Because this type of polycythemia demonstrates an increase in red cell mass with no involvement of other marrow elements, it is most commonly seen in association with a known hypoxic stimulus, increased erythropoietin concentration, or excess levels of adrenocortical steroids or androgens.

Laboratory features. The laboratory findings confirm increased red cell production with no increase in white cells or platelets. Erythropoietin levels are increased.

Clinical manifestations. The symptoms are those of the underlying disease state, such as cardiovascular disease with right-to-left shunt, chronic lung disease or alveolar hypoventilation, low barometric pressure and/or high altitude, or abnormal hemoglobin concentration.

Treatment. Because this condition is a physiologic compensation, the clinical treatment is directed at identifying and managing the underlying cause. Phlebotomy has been used to reduce cardiovascular work and appears to be helpful in both cardiovascular and chronic obstructive pulmonary disease. Oxygen administration is helpful in chronic lung diseases.

Course and prognosis. The course and prognosis are influenced by the underlying disease process.

Relative Polycythemia

Etiology and pathogenesis. Relative (spurious) polycythemia is characterized by an increased hematocrit level in the presence of normal or decreased total RBC mass. Two types of patients manifest this characteristic. In the first group, the laboratory finding is secondary to an obvious disturbance in fluid balance, such as is seen in severe dehydration or endocrinologic disorders. Patients in the other group, often described as having stress polycythemia, present with hypertension, increased hematocrit levels, and no increase in total RBC mass or obvious fluid loss. Research is continuing on the etiologic process and pathogenesis.

Laboratory features. All hematologic tests are normal except for elevated hematocrit and hemoglobin levels and RBC count. The size and color of the red cell are normal. Increased levels of cholesterol and uric acid are common.

Clinical manifestations. The manifestations are contingent on the underlying cause. In dehydration, the patient will have flat neck veins, decreased skin turgor, thirst, tachycardia, and, in severe cases, low cardiac output and blood pressure. If the underlying condition is stress related, the symptoms are those of a catecholamine stress response. Patients are usually Caucasian middle-aged men. In patients with spurious polycythemia caused by smoking, the problem is usually chronic, and the symptoms attributable to the hyperviscosity described for polycythemia vera are often found.

Treatment. Because this is a spurious form of polycythemia, it is important to recognize and manage the underlying cause. Fluid administration and management will resolve dehydration; however, spurious polycythemia is likely to be associated with a long-term condition that will require concurrent medical management. When the condition is a result of stress, identification of the stressors and stress management are indicated, with long-term follow-up. In spurious polycythemia attributable to smoking, the patient must stop smoking in order for the condition to resolve.

Course and prognosis. The long-term prognosis is excellent if the underlying condition is identified and resolved, but patients with chronic anxiety or an inability to quit smoking may experience the same complications related to erythrocytosis as are seen in polycythemia vera.

KEY POINTS

- Three types of polycythemia have been identified, according to cause. Polycythemia vera is associated with neoplastic transformation of bone marrow stem cells. Secondary polycythemia is due to chronic hypoxemia, with a resultant increase in erythropoietin production. Relative polycythemia is due to dehydration, which causes a spurious increase in RBC count.
- Differential diagnosis of the type of polycythemia is based on the history and accompanying manifestations:
  - *Polycythemia vera:* Absence of hypoxemia and dehydration, accompanied by leukocytosis and thrombocytosis.
  - *Secondary polycythemia:* History of lung disease or living at high altitude. Hypoxemia evident on blood gas evaluation. Erythropoietin level is elevated.
  - *Relative polycythemia:* History of fluid loss or poor intake. Accompanying manifestations of dehydration.
- Treatment of polycythemia is aimed at removing the cause, if possible. Phlebotomy and bone marrow–suppressing agents may be used for polycythemia vera. Major complications of polycythemia are increased blood viscosity and the risk of thrombi.

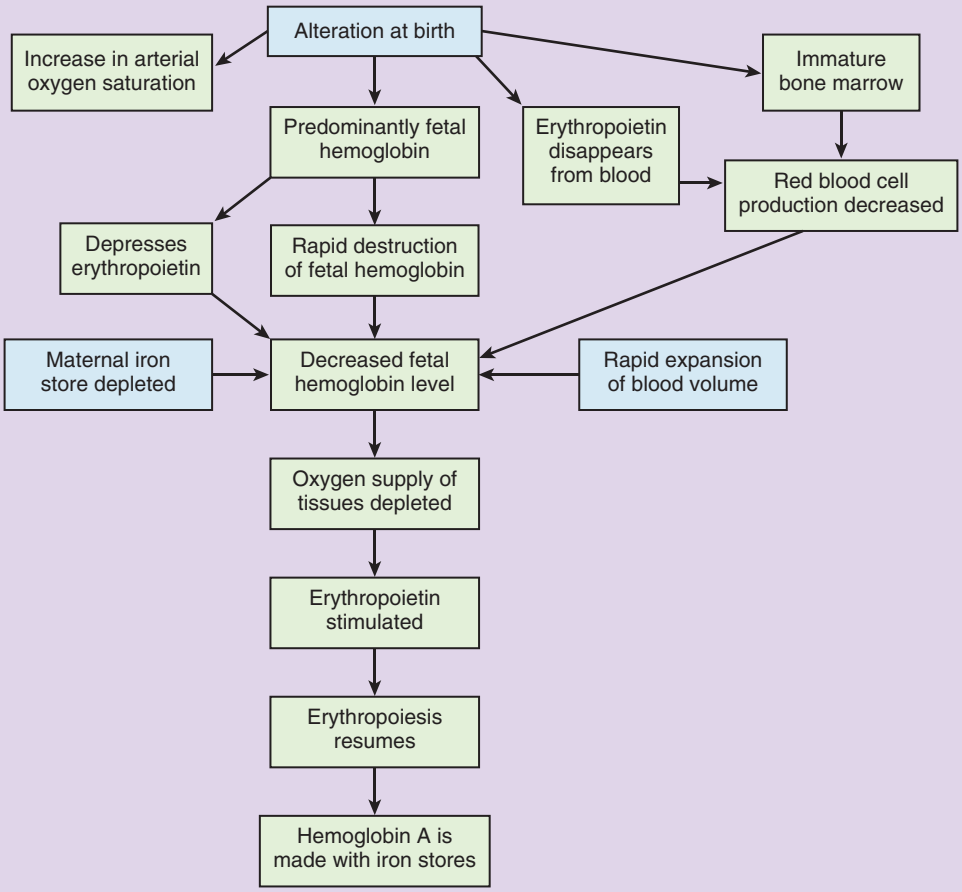
PEDIATRIC CONSIDERATIONS

Hemoglobin Synthesis in Infants

When an infant is born, many mechanisms occur to decrease the infant's hemoglobin level. At birth, erythropoietin, a hormone that stimulates RBC production, disappears from blood plasma; there is an increase in arterial oxygen saturation; and the infant is born with immature bone marrow. These mechanisms cause a slow rate of RBC production at birth. In addition, infants primarily have fetal hemoglobin, accounting for about 70% of their total hemoglobin. Although fetal hemoglobin is a more efficient oxygen carrier, it has a shorter life span than adult hemoglobin. This causes RBCs to be turned over every 70 to 90 days, instead of 120 days for adult RBCs. Also, fetal hemoglobin is thought to suppress production of erythropoietin.

Hemoglobin levels gradually decrease in the infant over the first 2 to 3 months because of the rapid destruction of fetal hemoglobin, decreased RBC production, and depressed erythropoietin production. Additionally, the infant experiences rapid growth during this time, which creates quick expansion of blood volume that further dilutes the supply of hemoglobin. Maternal iron stores are rapidly depleting at this time and will gradually diminish by 6 months.

As fetal hemoglobin is metabolized, the iron is released and stored. The body has enough iron to synthesize hemoglobin, but it is not stimulated to create hemoglobin at this time. Hemoglobin levels will continue to decrease until the oxygen needs of the tissues in the body are sufficiently depleted enough to stimulate erythropoietin production. Release of erythropoietin causes erythropoiesis to resume. Adult hemoglobin is made at this time with the iron stored in the body. Hemoglobin level will increase steadily in the infant starting around 6 months of age, and fetal hemoglobin is replaced by adult hemoglobin.



SUMMARY

The purpose of the erythron is to ensure adequate oxygen delivery with respect to oxygen demand. This is enhanced by the unique ability of hemoglobin in RBCs to carry and release oxygen at a suitable tension to support energy-generating systems in the body tissues. Anemia, a deficit in RBCs, poses a serious threat to oxygen transport and to the ability of the body to receive adequate oxygenation. Intense research in RBC physiology and pathophysiology continually yields new information for a better understanding of erythrocyte disorders, improved treatment modalities, and improved prognoses.

RESOURCES

Red Cell Production and Laboratory Analysis

Ahrens TS, Powers KC: Pulmonary clinical physiology. In Kinney MR, Packa DR, Dunbar SB, editors: *AACN's clinical reference for critical care nursing*, ed 4, St Louis, MO, 1998, Mosby.

American Edwards Laboratories: *Continuous SvO_2 monitoring: theory and applications*, Irvine, CA, n.d., The Laboratories.

Carmel R, Rosenblatt DS: Disorders of cobalamin and folate metabolism. In Handin RL, Lux SE, Stossel TP, editors: *Blood: principles and practice of hematology*, ed 2, Philadelphia, 2003, Lippincott.

Chow A, Frenettee PS: Hematopoiesis: origin and development of blood cells. In Greer JP, Arber DA, Glader B, et al, editors: *Wintrobe's clinical hematology*, ed 13, Philadelphia, 2014, Lippincott Williams & Wilkins. http://www.amazon.com/Wintrobe's-Clinical-Hematology. (Accessed 29 October 2015).

Damon LE, Andreadis C: Blood disorders. In Papadakis MA, McPhee SJ, Rabow MM, editors: *Current medical diagnosis & treatment*, New York, 2016, McGraw-Hill. http://accessmedicine.mhmedical.com/content.aspx?bookid=1585&sectionid=96306831.

Diggs L, Sturm D, Bell A: *The morphology of human blood cells*, Abbott Park, IL, 1985, Abbott Laboratories, 1 to 86.

Ellstrom K: The pulmonary system. In Alspach JG, editor: *Core curriculum for critical care nursing*, ed 6, St Louis, MO, 2006, Saunders.

Hall JE: Red blood cells, anemia, polycythemia. In Hall JE, editor: *Guyton and Hall textbook of medical physiology*, ed 13, Philadelphia, 2016, Elsevier, pp 413–422.

Hall JE: Transport of oxygen and carbon dioxide in blood and tissue fluids. In Hall JE, editor: *Guyton and Hall textbook of medical physiology*, ed 13, Philadelphia, 2016, Elsevier.

Hillman RS: Normal erythropoiesis. In Hillman RS, Ault KA, Leporrier M, Rinder HH, editors: *Hematology in clinical practice*, ed 5, Philadelphia, 2010, McGraw-Hill.

Jorde L, Roberts HR: Genetic principles and molecular biology. In Kaushansky K, Lichtman MA, Prchal JT, et al, editors: *Williams hematology*, ed 9, New York, 2016, McGraw-Hill. http://accessmedicine.mhmedical.com/content.aspx?bookid=1581&sectioned=94302226. (Accessed 20 October 2015).

Kaushansky K: Hematopoietic stem cells, progenitors, and cytokines. In Kaushansky K, Lichtman MA, Prchal JT, et al, editors: *Williams hematology*, ed 9, New York, 2016, McGraw-Hill. http://accessmedicine.mhmedical.com/content.aspx?bookid=1581&sectionid=94302625. (Accessed 20 October 2015).

McKenzie SB: Megaloblastic and nonmegaloblastic macrocytic anemias. In *Textbook of hematology*, ed 2, Baltimore, MD, 1996, Williams & Wilkins. http://www.barnesandnoble.com/s/textbook+of+hematology++McKenzie+2nd+ed. (Accessed 29 October 2015).

Nagel RL: Disorders of hemoglobin function and stability. In Handin RL, Lux SE, Stossel TP, editors: *Blood: principles and practice of hematology*, ed 2, Philadelphia, 2003, Lippincott.

Narla M: Structure and composition of the erythrocyte. In Kaushansky K, Lichtman MA, Prchal JT, et al, editors: *Williams hematology*, ed 9, New York, NY, 2016, McGraw-Hill. http://accessmedicine.mhmedical.com/content.aspx?bookid=1581&sectionid=94303279. (Accessed 20 October 2015).

Papayannopoulou T, Migliaccio AR: Biology of erythropoiesis, elytroid differentiation, and maturation. In Hoffman RL, Benz E, Silberstein LE, et al, editors: *Hematology: basic principles and practice*, ed 6, New York, NY, 2013, Elsevier. http://www.clinicalkey. (Accessed 20 October 2015).

Platt W: Introduction to hematology. In Platt W, editor: *Color atlas and textbook of hematology*, Philadelphia, 1979, Lippincott, pp 1–6.

Prchal JT: Clinical manifestations and classification of erythrocyte disorders. In Kaushansky K, Lichtman MA, Prchal JT, et al, editors: *Williams hematology*, ed 9, New York, NY, 2016, McGraw-Hill. http://accessmedicine.mhmedical.com/content.aspx?bookid=1581&sectionid=94303608. (Accessed 20 October 2015).

Prchal JT: Production of erythrocytes. In Kaushansky K, Lichtman MA, Prchal JT, et al, editors: *Williams hematology*, ed 9, New York, NY, 2016, McGraw-Hill. http://accessmedicine.mhmedical.com/content.aspx?bookid=1581&sectionid=94303394. (Accessed 20 October 2015).

Quigley JG, Means RT, Glader B: The birth, life, and death of red blood cells: erythropoiesis, the mature red blood cells, and cell destruction. In Greer JP, Arber DA, Glader B, et al, editors: *Wintrobe's clinical hematology*, ed 13, Philadelphia, 2014, Lippincott Williams & Wilkins. http://www.amazon.com/Wintrobe's-Clinical-Hematology. (Accessed 29 October 2015).

Ryan DH: Examination of the marrow. In Kaushansky K, Lichtman MA, Prchal JT, et al, editors: *Williams hematology*, ed 9, New York, NY, 2016, McGraw-Hill. http://accessmedicine.mhmedical.com/content.aspx?bookid=1581&sectionid=94301405. (Accessed 20 October 2015).

Ryan DH, Dave UP, Koury M, et al, editors: Structure of the marrow and the hematopoietic micro environment. In *Williams hematology*, ed 9, New York, NY, 2016, McGraw-Hill. http://accessmedicine.mhmedical.com/content.aspx?bookid=1581&sectionid=94301664. (Accessed 20 October 2015).

Anemias

Adamson JW: Iron deficiency and other hypoproliferative anemias. In Kasper DL, Fauci AS, Hauser SL, et al, editors: *Harrison's principles of internal medicine*, ed 19, New York, 2015, McGraw-Hill. http://accessmedicine.mhmedical.com/content.aspx?bookid=1130&sectioned=79731112. (Accessed 20 October 2015).

Antony AC: Megaloblastic anemias. In Hoffman RL, Benz E, Silberstein LE, et al, editors: *Hematology: basic principles and practice*, ed 6, New York, NY, 2013, Elsevier. http://www.clinicalkey. (Accessed 23 October 2015).

Bagby GC: Aplastic anemia and related bone marrow failure states. In Goldman L, Schaefer AI, editors: *Goldman's Cecil medicine*, ed 25, Philadelphia, 2016, Saunders, an imprint of Elsevier. http://www.clinicalkey. (Accessed 23 October 2015).

Bargman JM, Skorecki K: Chronic kidney disease. In Kasper DL, Fauci AS, Hauser SL, et al, editors: *Harrison's principles of internal medicine*, ed 19, New York, NY, 2015, McGraw-Hill. http://accessmedicine.mhmedical.com/content.aspx?bookid=1130&sectioned=79746512. (Accessed 20 October 2015).

Benz EJ: Disorders of hemoglobin. In Kasper DL, Fauci AS, Hauser SL, et al, editors: *Harrison's principles of internal medicine*, ed 19e, New York, NY, 2015, McGraw-Hill. http://accessmedicine.mhmedical.com/content.aspx?bookid=1130&sectioned=79731188. (Accessed 20 October 2015).

Borgna-Pignatti C, Galanello R: The thalassemias and related disorders: quantitative disorders of hemoglobin synthesis. In Greer JP, Arber DA, Glader B, et al, editors: *Wintrobe's clinical hematology*, ed 13, Philadelphia, 2014, Lippincott Williams & Wilkins. http://www.amazon.com/Wintrobe's-Clinical-Hematology. (Accessed 29 October 2015).

Brittenham GM: Disorders of iron metabolism: iron deficiency and overload. In Hoffman RL, Benz E, Silberstein LE, et al, editors: *Hematology: basic principles and practice*, ed 6, New York, NY, 2013, Elsevier. http://www.clinicalkey.com. (Accessed 23 October 2015).

Cappellini MD: Thalassemias. In Goldman L, Schaefer AI, editors: *Goldman's Cecil medicine*, ed 25, Philadelphia, 2016, Saunders. http://www.clinicalkey.com. (Accessed 23 October 2015).

Coetzer TL: Erythrocyte disorders. In Kaushansky K, Lichtman MA, Prchal JT, et al, editors: *Williams hematology*, ed 9, New York, NY, 2016, McGraw-Hill. http://accessmedicine.mhmedical.com/content.aspx?bookid=1581&sectionid=94304557. (Accessed 20 October 2015).

Cohen D, Valer AM: Treatment of irreversible renal failure. In Goldman L, Schaefer AI, editors: *Goldman's Cecil medicine*, ed 25, Philadelphia, 2013, Saunders, an imprint of Elsevier. http://www.clinicalkey. (Accessed 20 October 2015).

Fasano RM, Hendrickson JE, Luban NLC: Alloimmune hemolytic disease of the fetus and newborn. In Kaushansky K, Lichtman MA, Prchal JT, et al, editors: *Williams hematology*, New York, NY, 2016, McGraw-Hill. http://accessmedicine.mhmedical.com/content.aspx?bookid=1581&sectionid=94305868. (Accessed 27 October 2015).

Friedberg RC, Johari VP: Autoimmune hemolytic anemias. In Greer JP, Arber DA, Glader B, et al, editors: *Wintrobe's clinical hematology*, ed 13, Philadelphia, 2014, Lippincott Williams & Wilkins. http://

www.amazon.com/Wintrobe's-Clinical-Hematology. (Accessed 29 October 2015).

Gallagher PG: Hemolytic anemias: red cell membrane and metabolic defects. In Goldman L, Schaefer AI, editors: *Goldman's Cecil medicine*, ed 25, Philadelphia, 2016, Saunders. https://www.clinicalkey.com. October 23, 2015 A.

Gallagher PG: Red blood cell membrane disorders. In Hoffman RL, Benz E, Silberstein LE, et al, editors: *Hematology: basic principles and practice*, ed 6, New York, NY, 2013, Elsevier. http://www.clinicalkey.com. (Accessed 23 October 2015).

Gallagher PG, Glader BE: Hereditary spherocytosis, elliptocytosis, disorders associated with abnormalities of erythrocyte membrane. In Greer JP, Arber DA, Glader B, et al, editors: *Wintrobe's clinical hematology*, ed 13, Philadelphia, 2014, Lippincott Williams & Wilkins. http://www.amazon.com/Wintrobe's-Clinical-Hematology. (Accessed 29 October2015).

Ganz T: Anemia of chronic disease. In Kaushansky K, Lichtman MA, Prchal JT, et al, editors: *Williams hematology*, ed 9, New York, NY, 2016, McGraw-Hill. http://accessmedicine.mhmedical.com/content.aspx?bookid=1581&sectionid=94303965. (Accessed 20 October 2015).

Ganz T: Iron metabolism. In Kaushansky K, Lichtman MA, Prchal JT, et al, editors: *Williams hematology*, ed 9, New York, NY, 2016, McGraw-Hill. http://accessmedicine.mhmedical.com/content.aspx?bookid=1581&sectionid=101238028. (Accessed 20 October 2015).

Ginder GD: Microcytic and hypochromic anemias. In Goldman L, Schaefer AI, editors: *Goldman's Cecil medicine*, ed 25, Philadelphia, 2016, Saunders. http://www.clinical.key.com.

Glader BE: Hereditary hemolytic anemia due to red blood cell enzyme disorders. In Greer JP, Arber DA, Glader B, et al, editors: *Wintrobe's clinical hematology*, ed 13, Philadelphia, 2014, Lippincott Williams & Wilkins. http://www.amazon.com/Wintrobe's-Clinical-Hematology. (Accessed 29 October 2015).

Goodnaugh LT, Rerreth E: Iron deficiency and related disorders. In Greer JP, Arber DA, Glader B, et al, editors: *Wintrobe's clinical hematology*, ed 13, Philadelphia, 2014, Lippincott Williams & Wilkins. http://www.amazon.com/Wintrobe's-Clinical-Hematology. (Accessed 29 October 2015).

Green R: Folate, cobalamin, and megaloblastic anemias. In Kaushansky K, Lichtman MA, Prchal JT, et al, editors: *Williams hematology*, ed 9, New York, NY, 2016, McGraw-Hill. http://accessmedicine.mhmedical.com/content.aspx?bookid=1581&sectionid=101237678. (Accessed 20 October 2015).

Guinan EC, Shimamura A: Acquired aplastic anemia. In Greer JP, Arber DA, Glader B, et al, editors: *Wintrobe's clinical hematology*, ed 13, Philadelphia, 2014, Lippincott Williams & Wilkins. http://www.amazon.com/Wintrobe's-Clinical-Hematology. (Accessed 29 October 2015).

Jager U, Lechner K: Autoimmune hemolytic anemia. In Hoffman RL, Benz E, Silberstein LE, et al, editors: *Hematology: basic principles and practice*, ed 6, New York, NY, 2013, Elsevier. http://www.clinicalkey.com. (Accessed 23 October 2015).

Lee A, Lim W: Hematologic manifestations of liver disease. In Hoffman RL, Benz E, Silberstein LE, et al, editors: *Hematology: basic principles and practice*, ed 6, New York, NY, 2013, Elsevier. http://www.clinicalkey. (Accessed 20 October 2015).

Luzzatto L: Hemolytic anemias and anemia due to acute blood loss. In Kasper DL, Fauci AS, Hauser SL, et al, editors: *Harrison's principles of internal medicine*, ed 19, New York, NY, 2015, McGraw-Hill. http://accessmedicine.mhmedical.com/content.aspx?bookid=1130&sectioned=79720773. (Accessed 24 October 2015).

Means RT: Anemias secondary to chronic disease and systemic disorders. In Greer JP, Arber DA, Glader B, et al, editors: *Wintrobe's clinical hematology*, ed 13, Philadelphia, 2014, Lippincott Williams & Wilkins. http://www.amazon.com/Wintrobe's-Clinical-Hematology. (Accessed 29 October 2015).

Mitch WE: Chronic kidney disease. In Goldman L, Schaefer AI, editors: *Goldman's Cecil medicine*, ed 25, Philadelphia, 2013, Elsevier. http://www.clinicalkey. (Accessed 23 October 2015).

Natarajan K, Kutlar A: Disorders of hemoglobin structure: Sickle cell anemia and related abnormalities. In Kaushansky K, Lichtman MA, Prchal JT, et al, editors: *Williams hematology*, ed 9, New York, NY, 2016, McGraw-Hill. http://accessmedicine.mhmedical.com/content.aspx?bookid=1581&sectionid=101238113. (Accessed 20 October 2015).

Packman CH: Hemolytic anemia resulting from immune injury. In Kaushansky K, Lichtman MA, Prchal JT, et al, editors: *Williams hematology*, ed 9, New York, NY, 2016, McGraw-Hill. http://accessmedicine.mhmedical.com/content.aspx?bookid=1581&sectionid=94305662. (Accessed 27 October 2015).

Prchal JT: Clinical manifestations and classification erythrocyte disorders. In Kaushansky K, Lichtman MA, Prchal JT, et al, editors: *Williams hematology*, ed 9, New York, NY, 2016, McGraw-Hill. http://accessmedicine.mhmedical.com/content.aspx?bookid=1581&sectionid=94305662. (Accessed 27 October 2015).

Price EA, Otis S, Schrier SL: Red blood cell enzymopathies. In Hoffman RL, Benz E, Silberstein LE, et al, editors: *Hematology: basic principles and practice*, ed 6, New York, NY, 2013, Elsevier. http://www.clinicalkey.com. (Accessed 23 October 2015).

Quinn CT, Eder AF, Manno CS: Alloimmune hemolytic disease of the fetus and newborn. In Greer JP, Arber DA, Glader B, et al, editors: *Wintrobe's clinical hematology*, ed 13, Philadelphia, 2014, Lippincott Williams & Wilkins. http://www.amazon.com/Wintrobe's-Clinical-Hematology. (Accessed 29 October 2015).

Rao KV: Drug-induced hematologic disorders. In DiPiro JT, Talbert RL, Yee GC, et al, editors: *Pharmacotherapy: A Pathophysiologic Approach*, ed 9, New York, NY, 2014, McGraw-Hill. http://accesspharmacy.mhmedical.com/content.aspx?bookid=689&Sectionid=48811451. (Accessed 9 November 2015).

Rezvani AR, Lowsky R, Negrin RS: Hematopoietic cell transplantation. In Kaushansky K, Lichtman MA, Prchal JT, et al, editors: *Williams hematology*, ed 9, New York, NY, 2015, McGraw-Hill. http://accessmedicine.mhmedical.com/content.aspx?bookid=1581&sectionid=101237285. (Accessed 20 October 2015).

Sauntharajah Y, Vichinsky EP: Sickle cell disease—clinical features and management. In Hoffman R, et al, editors: *Hematology: basic principles and practice*, ed 6, New York, NY, 2013, Elsevier. https://www.clinicalkey.com. (Accessed 23 October 2015).

Segel GB, Lichtman MA: Aplastic anemia: acquired and inherited. In Kaushansky K, Lichtman MA, Prchal JT, et al, editors: *Williams hematology*, ed 9, New York, NY, 2016, McGraw-Hill. http://accessmedicine.mhmedical.com/content.aspx?bookid=1581&sectionid=94303673. (Accessed 20 October 2015).

Steinberg MH: Hemoglobins with altered oxygen affinity, unstable hemoglobins, M-hemoglobins, and dyshemoglobinemias. In Greer JP, Arber DA, Glader B, et al, editors: *Wintrobe's clinical hematology*, ed 13, Philadelphia, 2014, Lippincott Williams & Wilkins. http://www.amazon.com/Wintrobe's-Clinical-Hematology. (Accessed 29 October 2015).

Steinberg MH: Sickle cell disease and other hemoglobinopathies. In Goldman L, Schaefer AI, editors: *Goldman's Cecil medicine*, ed 25, Philadelphia, 2016, Saunders. At: https://www.clinicalkey.com. (Accessed 23 October 2015).

van Solinge WW, van Wijk R: Disorders of red cells resulting from enzyme abnormalities. In Kaushansky K, Lichtman MA, Prchal JT, et al, editors: *Williams hematology*, ed 9, New York, NY, 2016, McGraw-Hill. http://accessmedicine.mhmedical.com/content.aspx?bookid=1581&sectionid=94304881. (Accessed 20 October 2015).

Vulliamy TJ, Luzzatto L: Glucose-6-phosphate dehydrogenase deficiency and related disorders. In Handin RL, Lux SE, Stossel TP, editors: *Blood: principles and practice of hematology*, ed 2, Philadelphia, 2003, Lippincott.

Weatherall DJ: The thalassemias: disorders of globin synthesis. In Kaushansky K, Lichtman MA, Prchal JT, et al, editors: *Williams hematology*, ed 9, New York, NY, 2016, McGraw-Hill. http://accessmedicine.mhmedical.com/content.aspx?bookid=1581&sectionid=94305138. (Accessed 20 October 2015).

Young NS, Maciejewski JP: Aplastic anemia. In Hoffman RL, Benz E, Silberstein LE, et al, editors: *Hematology: basic principles and practice*, ed 6, New York, NY, 2013, Elsevier. http://www.clinicalkey. (Accessed 20 October 2015).

Blood Transfusion Therapy

American Association of Blood Banks: *American Red Cross, America's Blood Centers and the Armed Services Blood Program: Circular of information for the use of human blood and blood components*, 2013. Washington, DC, November, American Red Cross.

Galel S, Nguyen DD, Magali JF, et al: Transfusion medicine. In Greer JP, Arber DA, Glader B, et al, editors: *Wintrobe's clinical hematology*, ed 13, Philadelphia, 2014, Lippincott Williams & Wilkins. http://www.amazon.com/Wintrobe's-Clinical-Hematology. (Accessed 29 October 2015).

Herrmann PC: Erythrocyte disorders as a result of chemical and physical agents. In Kaushansky K, Lichtman MA, Prchal JT, et al, editors: *Williams hematology*, ed 9, New York, NY, 2016, McGraw-Hill. http://accessmedicine.mhmedical.com/content.aspx?bookid=1581&sectionid=94305547. (Accessed 27 October 2015).

McCullough J, Refaai MA, Cohn CS: Blood procurement and red cell transfusion. In Kaushansky K, Lichtman MA, Prchal JT, et al, editors: *Williams hematology*, ed 9, New York, NY, 2016, McGraw-Hill. http://accessmedicine.mhmedical.com/content.aspx?bookid=1581&sectionid=94306075. (Accessed 27 October 2015).

Polycythemia

Fauci AS, Braunwald E, Kasper DL, et al: Polycythemia vera and other myeloproliferative diseases. In Kasper DL, Fauci AS, Hauser SL, et al, editors: *Harrison's principles of internal medicine*, ed 19, New York, NY, 2015, McGraw-Hill. http://accessmedicine.mhmedical.com/content.aspx?bookid=1130&sectioned=79720773. (Accessed 24 October 2015).

Gilliland DG, Dunbar CL: Myelodysplastic syndromes. In Handin RL, Lux SE, Stossel TP, editors: *Blood: principles and practice of hematology*, ed 2, Philadelphia, 2003, Lippincott.

Kremyanskay M, Nasjfield V, Mascarenhas J, Hoffman R: Hoffman RL, Benz E, Silberstein LE, et al, editors: *Hematology: basic principles and practice*, ed 6, New York, NY, 2013, Elsevier. http://www.clinicalkey.com. (Accessed 27 October 2015).

Means RT: Erythrocytosis. In Greer JP, Arber DA, Glader B, et al, editors: *Wintrobe's clinical hematology*, ed 13, Philadelphia, 2014, Lippincott Williams & Wilkins. http://www.amazon.com/Wintrobe's-Clinical-Hematology. (Accessed 29 October 2015).

Means RT: Polycythemia vera. In Greer JP, Arber DA, Glader B, et al, editors: *Wintrobe's clinical hematology*, ed 13, Philadelphia, 2014, Lippincott Williams & Wilkins. http://www.amazon.com/Wintrobe's-Clinical-Hematology. (Accessed 29 October 201).

Prchal JT: Primary and secondary polycythemias (erythrocytosis). In Kaushansky K, Lichtman MA, Prchal JT, et al, editors: *Williams hematology*, ed 9, New York, NY, 2016, McGraw-Hill. http://accessmedicine.mhmedical.com/content.aspx?bookid=1581&sectionid=94306075. (Accessed 27 October 2015).

14

Alterations in Hemostasis and Blood Coagulation

Cheryl Rockwell

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- How do platelets and factors of the clotting cascade contribute to hemostasis?
- What findings from the patient history, physical examination, or laboratory studies would indicate a potential bleeding disorder?
- How are laboratory tests used to differentiate the various coagulation disorders?
- What vascular alterations result in abnormalities of hemostasis?
- What are the common causes of platelet deficiencies, excesses, and dysfunction?
- What are the common causes of inherited and acquired disorders of coagulation?

CHAPTER OUTLINE

The Process of Hemostasis, 298
Stages of Hemostasis, 298
Platelets, 299
Blood Coagulation Factors, 299
Fibrin Clot, 299
Fibrinolysis, 301
Evaluation of Hemostasis and Coagulation, 301
Clinical Assessment, 301
Laboratory Tests, 304
Vascular and Platelet Disorders, 304
Vascular Disorders, 304
Vascular Purpura, 304
Hereditary Hemorrhagic Telangiectasia, 306
Platelet Disorders, 306
Thrombocytopenia, 306
Thrombocytosis, 307
Qualitative Platelet Disorders, 307
Coagulation Disorders, 308
Hemophilia, 308
von Willebrand Disease, 309
Vitamin K Deficiency Bleeding in Infancy, 309
Acquired Vitamin K Deficiency, 309
Disseminated Intravascular Coagulation (DIC), 310
Hepatic Disease, 310

The term *hemostasis* means arrest of bleeding or prevention of blood loss after a blood vessel is injured. Hemostasis is accomplished via a complex interaction involving the vessel wall, circulating platelets, and plasma coagulation proteins. If hemostasis is inadequate, bleeding results; if hemostasis is excessive, inappropriate clotting or thrombosis results.

This chapter reviews the process of hemostasis and describes how that process is evaluated by means of clinical assessment and laboratory tests. The focus of this chapter is disorders of hemostasis and coagulation that result in bleeding. Disorders that result in thrombosis are discussed in Chapter 15.

THE PROCESS OF HEMOSTASIS

Stages of Hemostasis

Primary hemostasis, the initial response to vascular injury, involves the interaction between platelets and the endothelium of the injured blood vessel. The immediate response of the vessel to trauma is vasoconstriction to reduce blood loss. Although nervous reflex may play a part, this vasoconstriction results primarily from local myogenic spasm that may last from minutes to hours. The more trauma to the vessel, the greater the degree of vascular spasm.

The second component of primary hemostasis is formation of a platelet plug. Platelets not only adhere to endothelial collagen exposed by injury but also aggregate (clump together) at the site of vessel injury. The formation of this platelet plug is usually completed within 3 to 7 minutes.

Secondary hemostasis involves the formation of a fibrin clot, or coagulation, at the site of injury to maintain the hemostasis already initiated. Clotting factors are activated via the intrinsic pathway or extrinsic pathway and participate in a series of events that catalyze or facilitate the conversion of fibrinogen to fibrin. This process takes an average of 3 to 10 minutes.

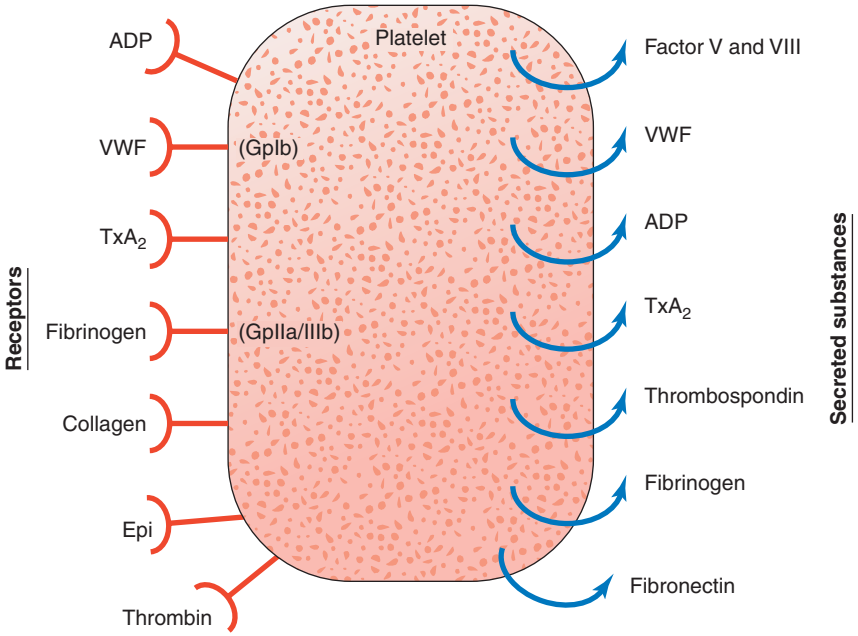


FIG 14.1 Platelets are complex cell fragments containing numerous chemical mediators that are released when platelets are activated. Platelets display a variety of cell surface receptors that mediate both adhesion to exposed subendothelium and aggregation with other platelets. *ADP,* Adenosine diphosphate; *Epi,* epinephrine; *TxA2,* thromboxane A2; *vWF,* von Willebrand factor.

Clot retraction, the final stage of clot formation, occurs when the components of the fibrin clot—the platelet plug, fibrin strands, and trapped red blood cells—are compressed or contracted to form a firm clot. This stage takes approximately 1 hour.

Platelets

Platelets have an integral role in hemostasis; thus it is important to review their nature and function (Fig. 14.1). A normal platelet count is between 150,000 and 400,000 platelets/mm^3 of blood. Platelets, also known as *thrombocytes,* are the smallest of the formed elements in the blood. They are produced in the bone marrow from megakaryocytes, which are derived from the pluripotent stem cell. Most of the platelets are found in the circulation, and about 25% are sequestered in the liver and spleen. Factors such as the stress response, epinephrine, and exercise may stimulate platelet production. The average life span of a platelet is 7 to 12 days. On completion of its life span, a platelet is eliminated from the circulation by the tissue macrophage system.

Platelets play a complex role in the process of hemostasis. Initially, platelets adhere to subendothelial collagen exposed by trauma (Fig. 14.2). After adhesion, the platelets become activated and initiate degranulation, the release of α granules and dense bodies. α Granules release platelet thrombospondin, fibrinogen, fibronectin, von Willebrand factor (vWF), and coagulation factors V and VIII. The dense granules release adenosine diphosphate (ADP), adenosine triphosphate (ATP), and serotonin. The presence of ADP and collagen encourages arachidonic acid formation, which leads to formation of thromboxane A_2 (TxA_2, a potent platelet aggregation agonist). Aspirin and other cyclooxygenase enzyme inhibitors can be used to block this cascade. Thromboxane A_2 stimulates the glycoprotein IIb/IIIa (GpIIb/IIIa) receptors on platelets to be expressed and further promotes platelet adhesion. The glycoprotein IIb/IIIa blockers (e.g., eptifibatide) are useful antiplatelet agents.

In addition to the major role platelets play in primary hemostasis, they are involved in secondary hemostasis and clot retraction. Platelets catalyze interactions between activated coagulation factors, accelerating the conversion of prothrombin to thrombin. Platelets also have a role in clot retraction.

Blood Coagulation Factors

With the exception of tissue factor (factor III; tissue thromboplastin) and calcium, blood coagulation factors are plasma proteins that circulate in the bloodstream in an inactive state. These factors are listed in Table 14.1 according to the internationally standardized nomenclature. The factors are numbered in the order of their discovery, not the order in which they participate in the clotting cascade. Factors with both active and inactive forms are differentiated with the letter "a" after the Roman numeral to designate the active form.

The liver is responsible for the synthesis of coagulation factors, with the exception of part of factor VIII. Factors II, VII, IX, and X; protein C; and protein S are dependent on vitamin K for synthesis and normal activity. Some of the coagulation proteins also can be synthesized by other cells, such as megakaryocytes and endothelial cells. Antithrombin III (ATIII) and protein C are protein complexes that promote anticoagulation. Antithrombin is a potent anticoagulant that binds to and inactivates free thrombin, preventing its binding and cleaving of fibrinogen. Protein C, a plasma protein that inactivates factors V and VIII, prevents clot formation. Protein S assists protein C in binding to phospholipase and stimulates release of tissue plasminogen activator, initiating fibrinolysis. Low-molecular-weight heparins and heparin work by enhancing the activity of antithrombin III (Fig. 14.3).

Fibrin Clot

In normal hemostasis, the fibrin clot is produced through activation of the intrinsic or extrinsic pathway and, in turn, the common final pathway. Effective hemostasis is the result of interactions between all of these pathways and is commonly referred to as the *coagulation cascade.*

Fig. 14.4 illustrates the coagulation cascade. The intrinsic pathway of coagulation begins when blood comes into contact with altered vascular

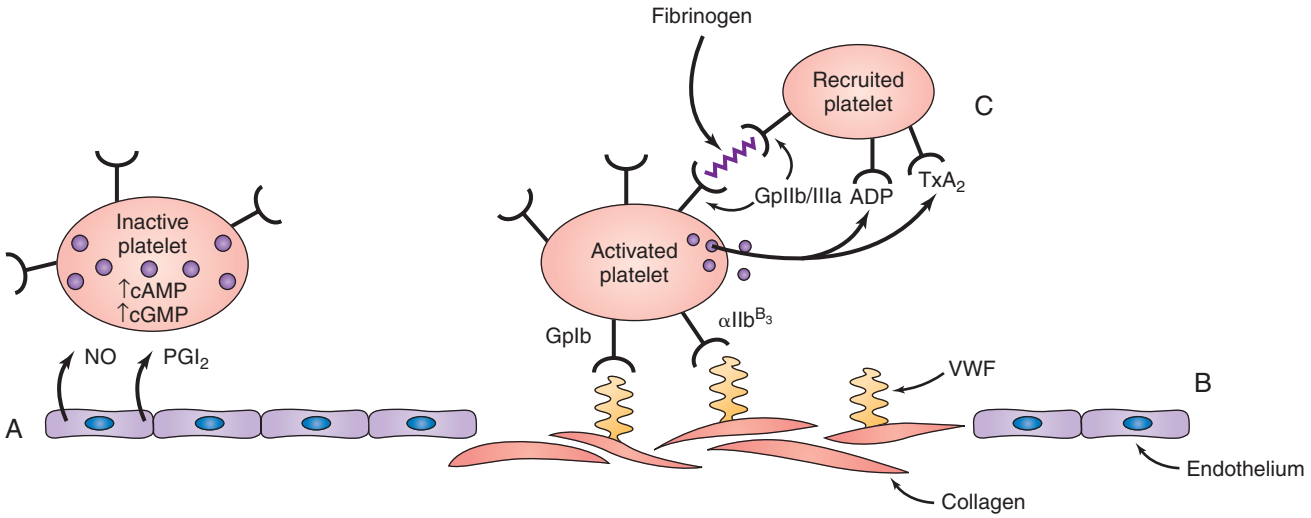


FIG 14.2 **A,** Endothelial cells normally prevent platelet adhesion by releasing nitric oxide (NO) and prostaglandin I2 (PGI2), which increases platelet cGMP and cAMP levels, reducing the likelihood of their being activated. **B,** Injury to the vessel wall exposes collagen and von Willebrand factor (vWF), which are bound by specific receptors on platelets, causing them to adhere and become activated. **C,** Activated platelets release numerous chemical mediators that bind to and stimulate other nearby platelets. Groups of platelets aggregate together by binding to fibrinogen molecules through their GpIIb/IIIa receptors. *ADP,* Adenosine diphosphate; *TxA2,* thromboxane A2.

TABLE 14.1 The Clotting Factors

Factor	Action
I: Fibrinogen	Factor I is converted to fibrin by the enzyme thrombin. Individual fibrin molecules form fibrin threads, which are the scaffold for clot formation and wound healing.
II: Prothrombin	Factor II is the inactive precursor of thrombin. Prothrombin is activated to thrombin by coagulation factor X (Stuart–Prower factor). After it is activated, thrombin converts fibrinogen (coagulation factor I) into fibrin and activates factors V and VIII. Synthesis is vitamin K–dependent.
III: Tissue thromboplastin	Factor III interacts with factor VII to initiate the extrinsic clotting cascade.
IV: Calcium	Calcium (Ca^{2+}), a divalent cation, is a cofactor for most of the enzyme-activated processes required in blood coagulation. Calcium also enhances platelet aggregation and makes red blood cells clump together.
V: Proaccelerin	Factor V is a cofactor for activated factor X, which is essential for converting prothrombin to thrombin.
VI: Discovered to be an artifact	No factor VI is involved in blood coagulation.
VII: Proconvertin	Factor VII activates factors IX and X, which are essential in converting prothrombin to thrombin. Synthesis is vitamin K dependent.
VIII: Antihemophilic factor	Factor VIII together with activated factor IX enzymatically activates factor X. In addition, factor VIII combines with another protein (von Willebrand factor) to help platelets adhere to capillary walls in areas of tissue injury.
IX: Plasma thromboplastin component (Christmas factor)	Factor IX, when activated, activates factor X to convert prothrombin to thrombin. This factor is essential in the common pathway between the intrinsic and extrinsic clotting cascades. A lack of factor IX is the basis for hemophilia B. Synthesis is vitamin K dependent.
X: Stuart–Prower factor	Factor X, when activated, converts prothrombin into thrombin. Synthesis is vitamin K dependent.
XI: Plasma thromboplastin antecedent	Factor XI, when activated, assists in the activation of factor IX. However, a similar factor must exist in tissues. People who are deficient in factor XI have mild bleeding problems after surgery but do not bleed excessively as a result of trauma.
XII: Hageman factor	Factor XII is critically important in the intrinsic pathway for the activation of factor XI.
XIII: Fibrin-stabilizing factor	Factor XIII assists in forming cross-links among the fibrin threads to form a strong fibrin clot.

endothelium or another negatively charged surface, such as glass. This contact phase of coagulation involves four factors: (1) factor XII, (2) high-molecular-weight kininogen (HMWK), (3) prekallikrein, and (4) factor XI. Factor XII is activated to factor XIIa, which in turn activates XI to XIa and prekallikrein to its active form, kallikrein (KAL). KAL liberates bradykinin from HMWK. The release of bradykinin produces an initial vasodilation followed by release of angiotensin II and vasoconstriction. The major role of factor XIa is activation of factor IX to factor IXa in the presence of calcium. Factor IXa then activates factor X to factor Xa in the presence of factor VIII, calcium, and phospholipid. This activation usually takes place on the membrane of stimulated platelets. The common final pathway is initiated by factor Xa.

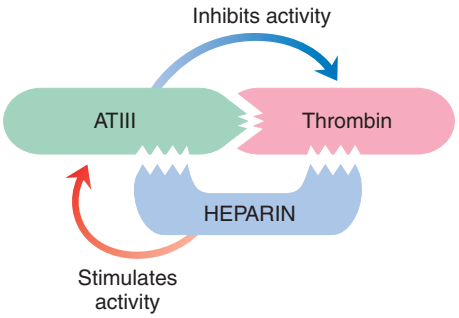


FIG 14.3 Antithrombin III (ATIII) can bind and neutralize the activity of thrombin. Heparin is a catalyst that increases the activity of ATIII, making it more effective. Thrombin is a potent inducer of clot formation; thus ATIII and heparin have significant anticoagulant properties.

The extrinsic pathway of coagulation begins when the vascular wall is traumatized, as in a crush injury. Tissue factor (factor III) from injured tissue activates factor VII. Factor VIIa activates factor X to Xa, which in turn initiates the common final pathway. Factor VIIa also activates factor IX in the intrinsic system.

The common final pathway of coagulation is initiated by factor X, which is activated by both the intrinsic and extrinsic pathways. Factor Xa, in the presence of factor V, calcium, and phospholipid, converts prothrombin (factor II) to thrombin. This conversion is facilitated by the presence of activated platelets. Thrombin then cleaves fibrinogen to form an insoluble fibrin clot. Thrombin also activates factor XIII, which promotes fibrin stabilization. The clot is further stabilized by clot retraction. Thrombin also helps to perpetuate the clotting cascade by continuing to activate factors V and VIII.

Fibrinolysis

At the same time the fibrin clot is forming, the process of fibrinolysis, or clot dissolution, is initiated (Fig. 14.5). Factor XII, HMWK, KAL, and thrombin are involved in the release of plasminogen activators. The plasminogen activators cleave plasminogen, a plasma protein that has been incorporated into the fibrin clot, to its active form, plasmin. Plasmin digests fibrinogen and fibrin and inactivates blood coagulation factors V and VIII. Fibrin split products, or fibrin degradation products, result from the dissolution of the fibrin clot.

The control of fibrinolysis is complex. The Kupffer cells of the liver and macrophages located in the spleen and bone marrow clear the circulation of activated clotting factors and fibrin degradation products. Antiplasmins that inhibit plasmin exist to prevent inappropriate fibrinolysis. All these factors and mechanisms are present to create a balance between clot production and clot dissolution.

KEY POINTS

- The several critical steps of hemostasis include vasospasm, formation of a platelet plug, and activation of the clotting cascade to form a fibrin clot.
- Factors released from platelets contribute to hemostasis by enhancing vasoconstriction, platelet aggregation, and vessel repair.
- Fibrin clot formation can be initiated by the intrinsic or extrinsic pathway. Each pathway requires the sequential activation of specific clotting factors, ultimately resulting in enzymatic cleavage of fibrinogen to form an insoluble fibrin clot.
- Initiation of fibrinolysis occurs simultaneously with clot formation to prevent excessive clotting and vessel occlusion.

TABLE 14.2 Clues From Patient History Regarding Bleeding Disorders

Clue From Patient History	Possible Cause
Family history of bleeding in both males and females	von Willebrand disease
Family history of bleeding in males	Hemophilia A or B
Newly acquired bruising	Drugs (especially aspirin and NSAIDs, anticoagulant therapy), thrombocytopenia
Excessive bleeding/bruising during/after surgery	Mild to severe deficiency of coagulation factors, von Willebrand disease; thrombocytopenia, drug ingestion
Bleeding after initial hemostasis	Factor XIII deficiency

NSAIDs, Nonsteroidal antiinflammatory drugs.

EVALUATION OF HEMOSTASIS AND COAGULATION

Data obtained from clinical assessment and laboratory tests facilitate the identification and evaluation of a hemostatic abnormality. Evaluation of a patient for a bleeding tendency is indicated in the following circumstances: when there is a personal or family history of bleeding, during active bleeding that is unresponsive to standard interventions, as part of screening before surgery, and for ongoing evaluation of anticoagulation therapy. A bleeding tendency may be inherited or acquired and may result from defects in blood vessels, platelets, or coagulation factors. The purpose of the evaluation process is to determine whether a problem exists and to ascertain the underlying cause so that appropriate management can be initiated.

Clinical Assessment

Both the family history and the personal history are important in the evaluation of a bleeding problem (Table 14.2). A family history of bleeding in males is often linked to one of the types of hemophilia, which accounts for the majority of serious inherited coagulation problems. The location, severity, duration, and setting in which bleeding occurs are also important clues to the type of defect that is present. Bleeding associated with vascular or platelet defects usually occurs immediately after trauma (e.g., dental extraction), involves skin or mucous membranes, and is brief. Delayed bleeding or bleeding into muscles or joints is more typical of a coagulation defect.

Systemic diseases, such as renal failure, liver disease, systemic lupus erythematosus, and malignancies, may be associated with a bleeding problem. Medication history, including use of over-the-counter medications, is another important aspect in the evaluation of a hemostatic defect. A common cause of acquired bleeding problems is drug ingestion. Specific drugs that alter hemostasis include aspirin and aspirin-containing preparations, nonsteroidal antiinflammatory agents, some antibiotics, anticoagulants, alcohol, and chemotherapeutic and thrombolytic agents.

Many of the physical findings of bleeding are manifested in the skin and mucous membranes. The individual may appear pale or jaundiced. Pallor is associated with a marked decrease in hemoglobin level; jaundice is associated with liver or gallbladder disease, possible coagulation disorders, and excessive red blood cell destruction.

Petechiae are flat, pinpoint, nonblanching red or purple spots caused by capillary hemorrhages in the skin and mucous membranes (Fig. 14.6). Petechiae are commonly seen with vascular and platelet disorders. They are usually present on dependent areas of the body, such as the

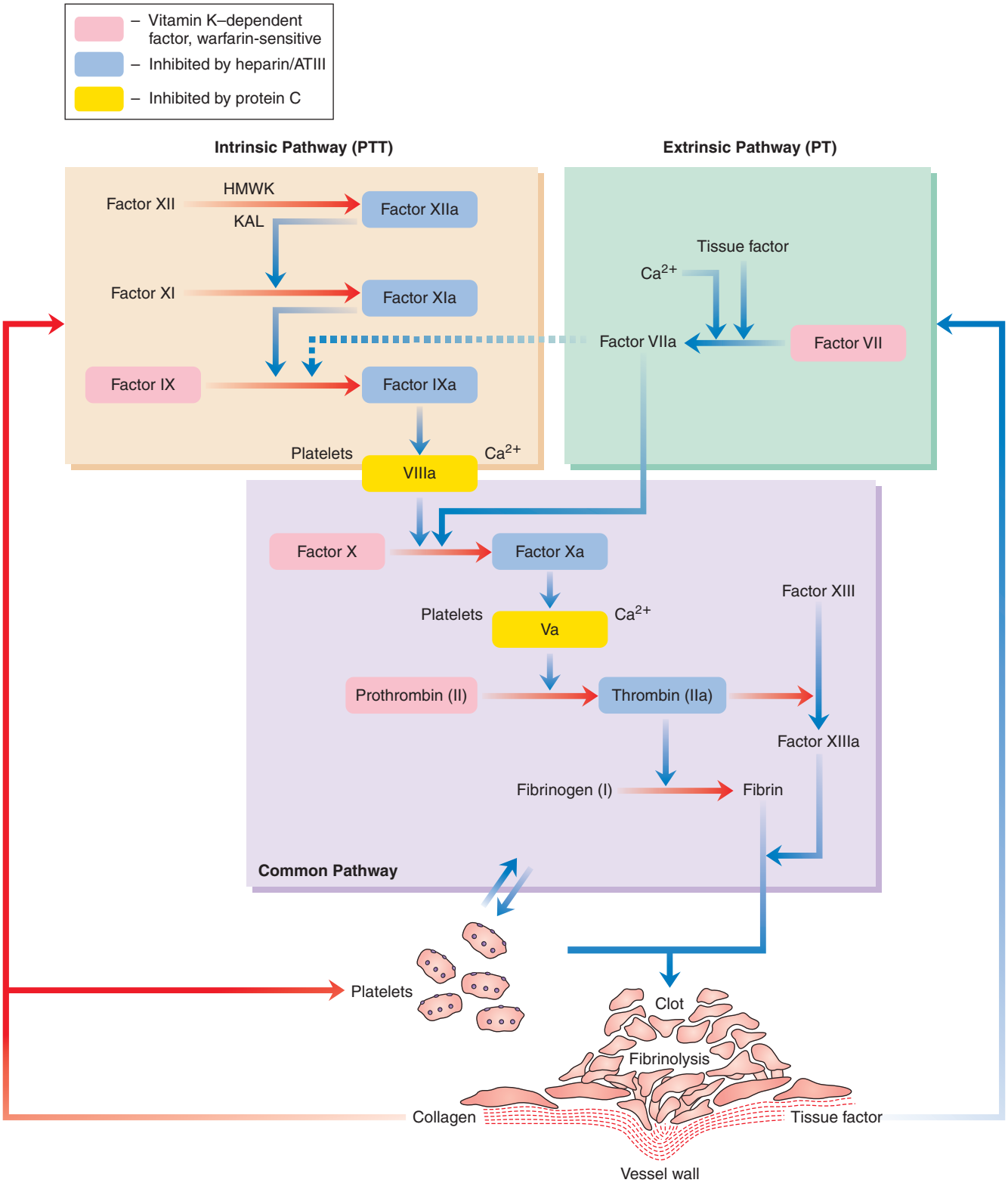


FIG 14.4 Coagulation cascade. *PT,* Prothrombin time; *PTT,* partial thromboplastin time.

legs, or on areas constricted by tight clothing. Not all petechiae indicate a bleeding problem. Petechiae found on other body areas not constricted by tight clothing, such as the abdomen or thorax, may be associated with infectious disease or other pathophysiologic sources. Petechiae may be seen in the newborn as a result of the trauma of delivery, not as a result of a bleeding problem.

When petechiae occur in groups or patches, the term *purpura* is used (Fig. 14.7). Purpuric lesions are often pruritic (itchy). Fever and malaise may be present, as may effusions into joints or viscera, manifested by joint or abdominal pain.

Ecchymosis occurs when blood escapes into the tissues, producing a bruise (Fig. 14.8). If the area is raised, it is called a *hematoma.*

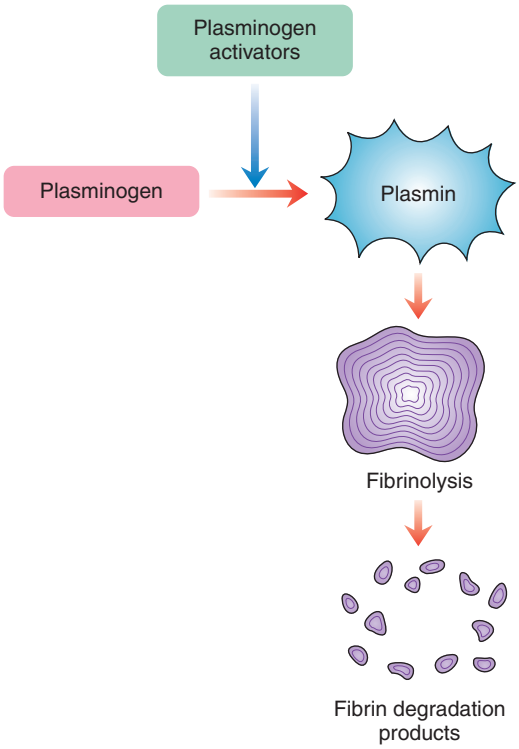


FIG 14.5 Fibrinolysis. Plasmin, activated from plasminogen, enzymatically cleaves fibrin proteins in the clot. This results in fibrin split products, which can be measured.

FIG 14.6 Petechiae. (From Dockery GL: *Cutaneous disorders of the lower extremity,* Philadelphia, 1997, Saunders.)

FIG 14.7 Purpura. (From Hurwitz S: *Clinical pediatric dermatology: a textbook of skin disorders of childhood and adolescence,* ed 2, Philadelphia, 1993, Saunders, p 269.)

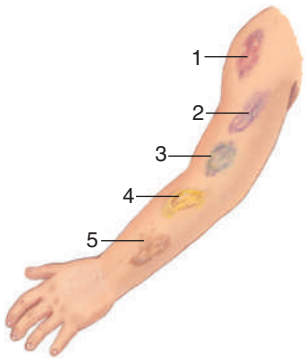


FIG 14.8 Ecchymosis. A large patch of capillary bleeding into tissues. Color in a light-skinned person is first red-blue or purple (1) immediately after or within 24 hours of trauma and generally progresses to blue to purple (2), blue-green (3), yellow (4), and brown to disappearing (5). (From Jarvis C: *Physical examination and health assessment,* ed 6, Philadelphia, 2012, Saunders.)

Hemarthrosis, manifested by swelling and pain, is bleeding into a joint. Large ecchymoses, hematomas, and hemarthroses are seen in coagulation disorders.

Telangiectasia is a lesion created by dilation of capillaries and small arteries, typically on the lips, tongue, tips of the fingers and toes, and sometimes in visceral vessels (Fig. 14.9). These thin, dilated, tortuous vessels are red to violet in color, blanch with pressure, and tend to bleed with minimal trauma. Spider telangiectasia branch into the subcutaneous and dermal layers of the skin and are often associated with liver disease.

Other significant findings indicative of a bleeding disorder include blood (bright red, rusty, or black) in drainage or excreta, such as feces (hematochezia or melena), urine (hematuria), vomitus (hematemesis), nasal drainage (epistaxis), gastric drainage, or sputum (hemoptysis). Excessive menstrual bleeding may occur (menorrhagia). Acute abdominal or flank pain may indicate internal bleeding. Hypovolemia from bleeding may produce a shock state and present as hypotension, tachycardia,

FIG 14.9 Telangiectasia (spider or star angioma). A fiery red, star-shaped marking with a solid circular center. Capillary radiations extend from the central arterial body. With pressure, note a central pulsating body and blanching of extended legs. Develops on face, neck, or chest; may be associated with pregnancy, chronic liver disease, or estrogen therapy; or may be normal. (From Hurwitz S: *Clinical pediatric dermatology: a textbook of skin disorders in childhood and adolescence,* ed 2, Philadelphia, 1993, Saunders, p 266.)

pallor, altered mentation, and decreased urine output. The two sites at which bleeding is most life threatening are the oropharynx (resulting in airway compromise) and within the brain tissue. One of the leading causes of death in patients experiencing severe disorders of coagulation is intracerebral hemorrhage.

Laboratory Tests

Many laboratory tests are available to aid in the diagnosis of hemostasis problems (Table 14.3). Basic screening includes a complete blood cell count (CBC), including a platelet count and peripheral blood smear, bleeding time, prothrombin time (PT) or international normalized ratio (INR), activated partial thromboplastin time (aPTT), and thrombin time. These screening tests evaluate both primary and secondary hemostasis. The CBC determines whether anemia is present, the platelet count determines the number of platelets, and the peripheral smear indicates the number and gross morphologic characteristics of platelets. The bleeding time evaluates vascular status and platelet function. The PT and INR assess the extrinsic pathway of coagulation, and the aPTT assesses the intrinsic pathway. Reporting prothrombin activity as a percentage of PT in seconds can pose difficulty in the adjustment of anticoagulation therapy because the PT varies with each laboratory and the reagent used at that laboratory. Laboratories have tried to compensate for this variation by using the ratio of the patient's value to the laboratory's control value, which again varied with the reagent. The INR is a standardized PT value used worldwide that controls for this reagent variability. Thrombin time measures the time needed to convert fibrinogen to fibrin; this reflects the quantity and quality of fibrinogen as well as the influence of any inhibitors. The D-dimer assay reflects fibrinolysis.

Further laboratory investigation is necessary if abnormalities are identified on the screening tests or if, despite normal screening test results, a bleeding problem obviously exists. Specific tests are available to assess abnormal platelet function, the presence of circulating anticoagulants or fibrin split products, and the levels of individual coagulation factors. Table 14.4 reflects the alterations in laboratory values seen with the major disorders of hemostasis.

KEY POINTS

- Bleeding tendencies may be inherited or acquired. Abnormal bleeding, liver disease, and anticoagulant drug use may be important risk factors. Physical findings of petechiae, purpura, ecchymoses, telangiectasia, and occult or frank bleeding are indicative.
- Usual laboratory tests include platelet count, bleeding time, PT/INR (extrinsic pathway), and aPTT (intrinsic pathway).

VASCULAR AND PLATELET DISORDERS

Vascular Disorders

Vascular disorders of hemostasis and coagulation are those in which the primary cause of bleeding is a problem with the vascular component of primary hemostasis. The vascular defect may be acquired (e.g., related to ingestion of a specific drug) or inherited.

Vascular Purpura

Etiology. Vascular purpura is a disorder in which purpura—patches of petechiae, or pinpoint hemorrhages, on the skin—are present. The primary cause of the purpura, or more extensive bleeding in some cases, is an abnormality of the vessels or the tissues that support them (see Fig. 14.7).

Allergic purpura (anaphylactoid purpura, Henoch–Schönlein purpura) is most often seen in children under age 8, with an average age of 6 years. Drug-induced purpura may result from many drugs, including atropine, chloral hydrate, and other sedatives; sulfa drugs; procaine penicillin; and warfarin (Coumadin). Purpuric lesions and perhaps severe hemorrhage are components of Ehlers–Danlos syndrome and osteogenesis imperfecta, both of which are inherited disorders of connective tissue. Acquired disorders of connective tissue such as scurvy (vitamin C deficiency), senile purpura (seen in the elderly), and corticosteroid purpura (associated with chronic steroid drug therapy) may also result in purpuric lesions.

Pathogenesis. The allergic purpuras are thought to result from an autoimmune process that produces inflammation or vasculitis of small vessels. As a result, perivascular infiltration and serosanguineous effusion occur in surrounding tissues to produce the characteristic purpuric lesion. The pathophysiologic process of drug-induced purpura is not well understood.

Structural abnormalities of vessels and perivascular supportive tissue provide the mechanism for bleeding in many of the vascular purpuras. These abnormalities may be inherited or acquired. In Ehlers–Danlos syndrome and osteogenesis imperfecta, the vascular abnormality is thought to result from decreased amounts or poor quality of collagen and elastin; both are necessary for perivascular support. Vitamin C deficiency, which causes scurvy, results in defective collagen synthesis. The lack of proper collagen support for the vessels leads to bleeding. In the elderly (senile purpura), loss of subcutaneous fat and changes in connective tissue allow for more mobility of the skin. Shearing force then causes rupture of small vessels. Steroids induce catabolism of proteins in supportive tissues, decreasing the mechanical strength of the microvasculature.

Clinical manifestations. The purpuric lesions characteristically appear and fade or disappear in groups. The lesions are not elevated and do not blanch with pressure.

With allergic purpura, the lesions tend to be palpable and are found on the proximal extremities, especially on the legs and buttocks; they may be accompanied by fever, pruritus, arthralgia, and paresthesia. Bleeding from the lesions themselves and generalized bleeding are

TABLE 14.3 Select Laboratory Tests Used to Assess Bleeding

Test	Normal Value*	Purpose or Significance
Platelet count	150,000–400,000/mm^3	Determines number of platelets; decreased in ITP, anemias, DIC, infection, chemotherapy; increased in leukemia, cancer, splenectomy
Bleeding time	3–10 min	Assesses platelet and vascular response; increased in thrombocytopenia, vascular defects, severe liver disease, DIC, von Willebrand disease, aspirin ingestion
Prothrombin time	10–14 sec; 100%	Evaluates extrinsic pathway of coagulation; increased in vitamin K deficiency, hemorrhagic disease of the newborn, liver disease, DIC, anticoagulant therapy. Evaluates all coagulation factors except VIII and XII.
International normalized ratio	1.5 (low-level anticoagulation for atrial fibrillation) 2.0–3.0 (medium-level anticoagulation for DVT, pulmonary embolism, MI, stroke prophylaxis) 2.5–3.5 (high-level anticoagulation for mechanical heart valve)	Evaluates extrinsic pathway of coagulation (as prothrombin time); provides uniformity worldwide, independent of reagents
Activated partial thromboplastin time	33–45 sec	Evaluates intrinsic pathway of coagulation; increased in hemophilia, vitamin K deficiency, liver disease, DIC, circulating anticoagulants, heparin therapy
Thrombin time	15 sec, or control + 5 sec	Measures conversion of fibrinogen to fibrin; increased in DIC, liver disease, low fibrinogen <100 mg/dl, multiple myeloma
Fibrinogen	200–400 mg/dl	Measures fibrinogen level; decreased in liver disease, DIC
Fibrin split products or fibrin degradation products†	<3 μg/ml	Measures byproducts from breakdown of fibrin clot; increased in DIC, hypoxia, leukemia, thromboembolic disorders
Clot retraction†	1 hr: evidence of shrinking and increased firmness 24 hr: 50% of volume is clot, 50% is serum	Approximate measure of platelet function; decreased in thrombocytopenia, von Willebrand disease
Platelet aggregation†	Visible aggregates form in <5 min	Measures rate and percentage of aggregation; decreased in mononucleosis, ITP, von Willebrand disease, leukemia, aspirin ingestion, thrombasthenia, Bernard–Soulier syndrome
Tourniquet test (Rumpel–Leede test, capillary fragility test)†	No petechiae or occasional petechiae	Evaluates vascular fragility and platelet function; positive test in thrombocytopenia, vascular purpuras, thrombasthenia
Euglobulin lysis time†	No lysis of fibrin clot at 37° C for 3 hr; clot is observed for 24 hr	Assesses fibrinolysis; increased lysis in DIC, incompatible blood transfusion, cirrhosis, cancer, obstetric complications
Plasma D-dimer assay	<200 ng/ml	Assesses fibrinolysis; increased in DVT, pulmonary embolism (highly nonspecific), DIC (high negative predictive value)

*Value may vary, depending on source.
†Tests not included in a routine coagulation screen.
DIC, Disseminated intravascular coagulation; *DVT,* deep vein thrombosis; *ITP,* immune thrombocytopenic purpura; *MI,* myocardial infarction.

TABLE 14.4 Alterations in Laboratory Values Seen With Major Disorders of Hemostasis

	LABORATORY ALTERATIONS							
Disorder	**Platelet Count**	**Bleeding Time**	**PT/INR**	**aPTT**	**TT**	**FSP**	**FVIII**	**FIX**
Idiopathic thrombocytopenic purpura	↓	Prolonged	N	N	N	N	N	N
Hemophilia A	N	N/Prolonged	N	↑	↑	↓	↓	N
Hemophilia B	N	N/Prolonged	N	↑	↑	N	N	↓
von Willebrand disease	N	Prolonged	N	↑	↑	N	↓	N
Vitamin K deficiency	N	Prolonged	↑	N/↑	↑	N	N	N/↓
Disseminated intravascular coagulation	↓	Prolonged	↑	↑	↑	↑	↓	↓
ASA/NSAIDs	N	Prolonged	N	N	N	N	N	N
Heparin	N/↓	Prolonged	N	↑	↑	N	N	N
Coumadin	N	Prolonged	↑	N/↑	↑	N	N	N
Vascular purpura	N	N/Prolonged	N	N	N	N	N	N
Liver disease	N/↓	Prolonged	↑	N/↑	N	N	N	N/↓

aPTT, Activated partial thromboplastin time; *ASA,* acetylsalicylic acid; *FIX,* factor IX; *FSP,* fibrin split products; *FVIII,* factor VIII; *INR,* International normal ratio; *N,* normal; *NSAIDs,* nonsteroidal antiinflammatory drugs; *PT,* prothrombin time; *TT,* thrombin time.

uncommon. Usually, allergic purpura is self-limited, and the prognosis is good.

Generalized purpura is characteristic of drug-induced vascular purpura. The lesions quickly subside when the drug is discontinued. Other bleeding manifestations are uncommon.

The purpuric lesions associated with inherited connective tissue disorders, such as Ehlers–Danlos syndrome, often are accompanied by large ecchymoses and hematomas. Although not common, bleeding into the brain tissue may result in cerebrovascular accident (stroke).

The purpuric lesions seen with scurvy typically occur around hair follicles and on the medial surfaces of the thighs and buttocks. Ecchymoses and large hematomas may also occur.

Senile purpura and corticosteroid purpura generally occur on the dorsum of the hands and forearms and are aggravated by trauma. Other bleeding is uncommon.

Diagnosis and treatment. The diagnosis of vascular purpura is one of exclusion after platelet disorders and coagulation disorders have been ruled out. An abnormal tourniquet test (positive Rumpel–Leede test) in the setting of a normal or increased bleeding time, normal platelet count, and normal coagulation study results suggest a problem with the vascular component of hemostasis. The tourniquet test is an assessment for abnormal capillary fragility. To perform the test, a blood pressure cuff is applied and inflated to a point between the systolic and diastolic blood pressures for 5 minutes. The test is positive if there are more than 20 petechiae per square inch.

Treatment for vascular purpura includes removal or avoidance of the causative agent if one is identified (e.g., penicillin) and interventions to relieve symptoms such as pruritus. If more extensive bleeding accompanies the purpura, identification of the cause and interventions to control the bleeding are necessary.

Hereditary Hemorrhagic Telangiectasia

Etiology. A telangiectasia is a dilated or tortuous small blood vessel found in the skin or mucous membranes that has a tendency to bleed spontaneously or after minor trauma (see Fig. 14.9). Hereditary hemorrhagic telangiectasia (Osler–Weber–Rendu disease) is transmitted as an autosomal-dominant trait; the vascular abnormalities can be seen in children but become more prominent after puberty, peaking between the fourth and fifth decades. As the telangiectases—the skin spots resulting from the vascular lesion—become more prominent, the frequency and severity of the bleeding increase.

Pathogenesis. The telangiectases result from an abnormality in vascular development. The vessel wall is composed of a single layer of endothelium; thus support and contractile properties are deficient, leading to spontaneous bleeding or bleeding as a result of minor trauma. Any mucosal surface (e.g., respiratory, gastrointestinal, and genitourinary tracts) may be involved. Arteriovenous malformations in the lung, liver, and brain are the most serious complications.

Clinical manifestations. Bright red or purple lesions, ranging from pinpoint to 3 mm in diameter, can be found on the nasal mucous membranes, lips, palate, tongue, face, trunk, palms of the hands, and the soles of the feet. A hallmark symptom is recurrent nosebleeds (epistaxis), with increasing frequency as the patient ages. The severity of the disorder is linked to age of onset. Typically the lesions are flat and blanch with pressure. The most common clinical problem is mucous membrane bleeding, especially epistaxis. However, bleeding may occur from telangiectases in any area. Frequent bleeding episodes may result in anemia.

Diagnosis and treatment. The diagnosis is confirmed by the presence of multiple telangiectases, repeated episodes of bleeding, or a family history of bleeding in both genders. If telangiectases are not easily visible, the diagnosis is more difficult to make.

Treatment is primarily supportive and includes humidification of the nasal cavity, use of topical hemostatic agents or cauterization if the bleeding site is accessible, tamponade of the nasal cavity, use of iron replacement therapy, laser treatment for cutaneous lesions, embolization, and administration of estrogen or estrogen with progesterone for epistaxis. Tranexamic acid is used to control severe hemorrhage. Blood transfusions or surgical intervention for uncontrolled bleeding may be considered in selected cases.

Platelet Disorders

Platelet disorders of hemostasis and coagulation are those in which the primary cause of bleeding is an abnormality in the quantity or the quality of platelets.

Thrombocytopenia

Etiology. Thrombocytopenia is a common cause of generalized bleeding. Some of the many causes of thrombocytopenia are listed in Box 14.1.

Immune thrombocytopenia (ITP), previously called *idiopathic thrombocytopenic purpura,* is an acquired immune-mediated disorder. ITP is classified according to the duration of the disease: newly diagnosed, persistent (lasting 3 to 12 months), or chronic (lasting 12 months or longer). ITP occurs at any age and in both genders until midlife, when it is more common in women 30 to 60 years of age. ITP in children may follow a viral illness or have no evidence of previous illness and typically spontaneously resolves in 6 months. ITP in adults has a more insidious onset and tends to be more chronic. A pregnant woman with ITP can deliver a thrombocytopenic infant because the antiplatelet antibody crosses the placenta. Adult ITP may precede or occur in association with diseases of altered immunity, such as systemic lupus erythematosus (see Chapter 10), lymphoproliferative disease (see Chapter 11), or AIDS (see Chapter 12).

Pathogenesis. Four general mechanisms are responsible for thrombocytopenia: decreased platelet production, decreased platelet survival,

BOX 14.1 Some Causes of Thrombocytopenia

Decreased Platelet Production
Folate/B_{12} deficiency
Radiation therapy
Chemotherapy
Drugs (e.g., alcohol, thiazides, phenytoin)
Aplastic anemia
Cancer in bone marrow

Decreased Platelet Survival
Drugs (e.g., thiazides, digoxin, heparin, furosemide, certain antibiotics)
Mechanical prosthetic heart valves
Viral and bacterial infections
Circulating immune complexes
Increased destruction in the spleen
Disseminated intravascular coagulation

Splenic Sequestration (Pooling)
Splenomegaly
Hypothermia

Platelet Dilution
Massive transfusions with blood stored for more than 24 hours

splenic sequestration (pooling), and intravascular dilution of circulating platelets (see Box 14.1). Regardless of the mechanism responsible, fewer platelets are available, and inadequate hemostasis is the potential result.

Platelets are produced by bone marrow megakaryocytes. Production declines when the number of megakaryocytes is reduced or when the process of platelet production (thrombocytopoiesis) is ineffective. Although numerous causes of decreased platelet production are listed in Box 14.1, drugs are often responsible. Bone marrow suppression from chemotherapy, recent immunizations, and alcohol ingestion are common causes of platelet level reduction.

The average life span of a platelet is 7 to 9 days. Decreased platelet survival may be the result of an antibody-mediated immune mechanism that destroys platelets (e.g., ITP, a possible adverse effect of heparin) or the result of increased consumption of platelets, as seen in disseminated intravascular coagulation (DIC). Direct trauma to platelets from vascular or valvular prostheses also may be responsible for decreased platelet survival.

Normally, about 25% of the total number of platelets can be found in the spleen and the remaining 75% are circulating. When the spleen is enlarged (splenomegaly), as much as 90% of the platelets may be pooled or sequestered in the spleen; thus the circulating number of platelets is markedly decreased. If the spleen cannot be palpated on physical examination, platelet sequestration can be ruled out as the primary mechanism of the thrombocytopenia.

The final mechanism responsible for thrombocytopenia is dilution of circulating platelets by administration of massive transfusions. Platelets degenerate in stored blood after 24 hours; thus when a large amount of blood deficient in platelets is transfused, thrombocytopenia results.

Clinical manifestations. Clinical manifestations of thrombocytopenia are generally absent until the platelet count drops below 100,000/mm^3. Petechiae and purpura are prominent with platelet counts below 50,000/mm^3. Spontaneous mucosal, deep tissue, and intracranial bleeding may be seen with platelet counts less than 20,000/mm^3, though recent evidence suggests serious bleeding is more likely to occur at platelet counts less than 10,000/mm^3.

Diagnosis. Thrombocytopenia is diagnosed by the presence of a low platelet count on peripheral blood laboratory examination. The bleeding time is prolonged, and clot retraction is poor or absent. PT/INR, partial thromboplastin time, and other coagulation studies are normal. The CBC will indicate if the thrombocytopenia is isolated or if an associated problem, such as anemia or leukopenia, is present. Gross morphologic analysis of platelets, evaluated from the peripheral blood smear, and bone marrow examination provide additional information regarding the mechanism for the thrombocytopenia. Because many drugs are associated with thrombocytopenia, careful review of all medications the patient is taking is also necessary in the search for the cause of the thrombocytopenia.

Treatment. The treatment for thrombocytopenia is based on the identified cause or mechanism and may include any of the following: discontinuation of any suspected drug; avoidance of aspirin and pharmacodynamically similar drugs that alter normal platelet function; and administration of corticosteroids, immunosuppressants, intravenous immunoglobulin, rituximab, and thrombin receptor agonists such as romiplostim and eltrombopag to increase platelet production. Splenectomy may be helpful in some cases because it results in removal of a major site of platelet destruction and eliminates a source for production of antiplatelet antibodies.

Thrombocytosis

Etiology. Thrombocytosis is generally defined as a platelet count greater than 400,000/mm^3. Transitory thrombocytosis is seen after stress or physical exercise. Secondary or reactive thrombocytosis occurs as a response to hemorrhage, inflammatory diseases, malignancy, infection, hemolysis, or splenectomy. Primary thrombocytosis is seen with polycythemia vera and chronic granulocytic leukemia.

Pathogenesis. In all types of thrombocytosis, the number of platelets is increased, but the mechanism of the increase varies. Transitory thrombocytosis results from release of preformed platelets, not increased production. As the name implies, the elevation in platelet count is transient. Secondary thrombocytosis results from an actual increase in platelet production via an unknown mechanism. With primary thrombocytosis, there is abnormal proliferation of megakaryocytes in the bone marrow, resulting in as much as a 15-fold increase in platelet production.

Clinical manifestations. In general, transitory thrombocytosis and secondary thrombocytosis do not result in hemorrhage or thrombotic complications. Hemorrhage into the skin and mucous membranes and gastrointestinal bleeding may be seen with primary thrombocytosis. The pathogenesis of excessive bleeding in the presence of excessive levels of platelets is not well understood. Thrombosis resulting in peripheral vascular ischemia or pulmonary embolism may further complicate the clinical picture. Thromboembolic events are the most common cause of death. However, the course of thrombocytosis is benign in most patients.

Diagnosis and treatment. The diagnosis is made on the basis of a high platelet count. Bleeding time may be normal or prolonged, and platelet aggregation is normal or impaired. The history and clinical presentation, as well as additional laboratory tests such as bone marrow examination, aid in determining the type of thrombocytosis.

No treatment is necessary with transitory and secondary thrombocytosis. To manage primary thrombocytosis, the use of cytotoxic agents or interferon may be used. Antiplatelet therapy (e.g., aspirin or dipyridamole) also may be used. In the presence of acute bleeding or thrombosis, plasma exchange may be used to temporarily control the platelet count.

Qualitative Platelet Disorders

Etiology. Although the number of platelets may be normal, the ability of the platelets to function in the hemostatic process may be abnormal; thus a qualitative platelet disorder is present. Inherited defects in platelet function, such as Bernard–Soulier syndrome (giant platelet syndrome), von Willebrand disease, and thrombasthenia (Glanzmann disease), are rare. In contrast, acquired disorders of platelet function are common; they are often associated with drugs, especially aspirin; with renal failure; or with a coexisting hematologic disease, such as leukemia.

Pathogenesis. Whether the qualitative platelet disorder is inherited or acquired, at least one aspect of platelet function (adhesion, aggregation, or degranulation or release reaction) is abnormal and a bleeding tendency results. In both Bernard–Soulier syndrome and von Willebrand disease, platelet adhesion is abnormal. Platelet aggregation is the problem in thrombasthenia, due to the absence of the fibrinogen receptor necessary for normal platelet aggregation. Aspirin and other nonsteroidal anti-inflammatory agents inhibit production of thromboxane A_2 and thus impair both platelet aggregation and the platelet release reaction.

Clinical manifestations. The clinical presentation of a qualitative platelet disorder is some form of bleeding tendency, such as petechiae or purpura on skin and mucous membranes, epistaxis, gastrointestinal bleeding, or menorrhagia. Acquired platelet function defects also may result in excessive bleeding during and after surgical procedures.

Diagnosis and treatment. With qualitative platelet defects, the bleeding time is prolonged but the platelet count and other routine coagulation screening test results are normal. Although a bleeding time greater than 10 minutes is associated with a slight increase in bleeding

tendency, the risk is not significantly increased until the bleeding time exceeds 15 or 20 minutes.

Special laboratory tests that more specifically evaluate platelet function, such as platelet aggregation studies, are necessary to determine the exact cause of bleeding. Coexisting hematologic defects may make diagnosis of a platelet defect difficult.

If the platelet disorder is drug induced, the offending drug is discontinued. Transfusion with normal platelets is the usual intervention if treatment is necessary because of bleeding. Administration of desmopressin or cryoprecipitate is the treatment of choice when von Willebrand disease is the underlying cause of bleeding, as well as for patients with aspirin overdose and cirrhosis. von Willebrand disease is described in greater detail in the following section.

KEY POINTS

- Disorders of the vasculature that result in altered hemostasis include inflammation (allergic purpura), structural abnormalities (collagen diseases), and weakened vessel walls (telangiectasia).
- An insufficient quantity of platelets (fewer than 50,000/mm^3) results from decreased production, sequestration, increased destruction, or dilution. Important causes of thrombocytopenia include autoimmune destruction (ITP), DIC, and mechanical destruction (artificial valves).
- Excessive quantity of platelets (more than 400,000/mm^3) results from excessive production (proliferation of bone marrow cells). Thrombocythemia may result in excessive coagulation with thrombosis or excessive bleeding.
- A normal platelet count does not ensure adequate platelet function. Platelet adhesion, aggregation, and degranulation may be abnormal, resulting in a prolonged bleeding time. The usual cause is drug related (e.g., aspirin); rarely, the platelet defect is inherited (e.g., von Willebrand disease).

COAGULATION DISORDERS

Coagulation disorders, or coagulopathies, are defects of the normal clotting mechanism. They may cause bleeding as a result of problems with the formation, stabilization, or lysis of the fibrin clot. Alternatively, the coagulation disorder may be attributable to inappropriate activation of the coagulation cascade, producing excessive clot formation.

Hemophilia

Etiology. Hemophilia is rare in the general population, but it is the most common severe inherited coagulation disorder. Excessive bleeding after circumcision or the formation of a hematoma after vitamin K injection may lead to the diagnosis in the neonate. Some children will not develop bleeding problems until they begin crawling or walking.

Hemophilia A, the classic form of the disease, accounts for approximately 85% of cases of clinical hemophilia. Hemophilia A is caused by factor VIII deficiency. The majority of patients inherit this X-linked recessive disorder; hemophilia is transmitted by an asymptomatic carrier female to an affected son. Approximately 20% of patients with hemophilia A have a negative family history because of a spontaneous mutation of the hemophilic gene.

Less common than hemophilia A is hemophilia B, also known as *Christmas disease.* Factor IX is deficient in this form of hemophilia.

Hemophilia is often classified according to the extent to which the specific coagulation factor (factor VIII or IX) is deficient. Patients with severe hemophilia have less than 1% normal coagulation factor activity; patients with moderate hemophilia, 1% to 5% normal coagulation factor activity; and patients with mild hemophilia, 5% to 40% normal coagulation factor activity.

Of critical concern in the hemophilic patient are intracranial hemorrhage and other serious bleeding episodes. Because of advances in treatment, however, a normal life span is possible for many.

Pathogenesis. Hemophilia A results from factor deficiency or the abnormal function of factor VIII. Hemophilia B results from factor deficiency or the abnormal function of factor IX. A deficiency or malfunction in either factor interferes with the normal sequence of events in the intrinsic pathway of coagulation and, in turn, the eventual production of a fibrin clot. Inability to form a fibrin clot results in bleeding.

Clinical manifestations. Once clinical evidence of bleeding is present, hemophilia A and hemophilia B are indistinguishable. Patients with mild hemophilia may not experience symptoms until stressed by surgery or trauma. Prolonged bleeding from relatively minor trauma and occasional spontaneous bleeding episodes are characteristic of moderate hemophilia. With severe hemophilia, frequent episodes of spontaneous bleeding are likely.

Any of the following clinical manifestations may occur: easy bruising, prolonged bleeding from the nasal or oral mucosa, deep tissue hematomas, hemarthrosis, bleeding into muscles in the extremities, spontaneous hematuria, gastrointestinal bleeding, and intracranial bleeding. The hallmark of hemophilia is hemarthrosis. Knees, ankles, and elbows are the most often affected. Repeated episodes of hemarthrosis may result in joint deformity.

Major long-term complications of hemophilia include progressive joint deformity as a result of repeated hemarthroses, hepatitis B or C, cirrhosis, and HIV infection related to repeated transfusions or administration of virus-contaminated factor concentrates.

Diagnosis and treatment. Hemophilia is the suspected cause of bleeding when the family history is positive for bleeding in males, there is a history of joint bleeding and hematomas, and joint deformity is present on physical examination. Laboratory tests consistent with hemophilia include a normal or slightly prolonged bleeding time, a normal PT/INR, and a prolonged aPTT. Factor assay verifies a deficiency in factor VIII or IX. Early in pregnancy, chorionic villus biopsy or amniocentesis may be done to identify factor deficiency, making prenatal diagnosis of hemophilia possible.

The patient and family must learn about hemophilia, including recognition and appropriate response to bleeding episodes, necessary lifestyle changes, and the genetic nature of disease transmission. Prevention of injury and avoidance of aspirin and aspirin-like drugs, which alter platelet function, are important parts of treatment. Joint bleeding is managed by immobilization of the limb and application of ice.

With dental procedures requiring administration of a local anesthetic, prophylactic administration of factor VIII should be considered in the patient with hemophilia A. Bleeding episodes from hemophilia A are managed primarily by the administration of cryoprecipitate or other preparations of factor VIII concentrate. Recombinant DNA–derived factor concentrates contain no viruses and are now available. In addition, desmopressin and antifibrinolytics (e.g., tranexamic acid and aminocaproic acid) are possibilities to treat mild bleeding episodes. The goal of therapy is to obtain a factor VIII level of at least 40% to 60% for joint and most muscle bleeding, and 80% to 100% for iliopsoas muscle, throat or neck, central nervous system, or gastrointestinal bleeding. Up to 20% of patients with severe hemophilia develop factor VIII inhibitor, an antibody that rapidly inactivates transfused factor VIII. Plasmapheresis and immunosuppressive therapy are sometimes necessary to maintain adequate factor VIII levels in these patients.

Mild to moderate bleeding resulting from hemophilia B is managed with the administration of fresh or fresh frozen plasma or cryoprecipitate. Concentrates containing factors II, VII, IX, and X are another therapeutic option. These concentrates are now treated in a variety of fashions (heat, pasteurization, solvent detergents, immunoaffinity purification)

to prevent transmission of viruses. A highly purified factor IX concentrate that appears to be safe in terms of both adverse effects and viral transmission is also available.

von Willebrand Disease

Etiology. von Willebrand disease is inherited as an autosomal-dominant disorder of factor VIII carrier protein and platelet dysfunction. In rare cases, von Willebrand disease is an autosomal-recessive disorder. Several less common subtypes of the disease have been identified, but all have some defect in vWF, a plasma protein. von Willebrand disease occurs in both females and males. Bleeding manifestations of the disease tend to become more severe with age.

Pathogenesis. VWF and factor VIII normally circulate in plasma as a complex. VWF is necessary for stabilization of factor VIII in the circulation and for normal adherence of platelets to damaged vascular endothelium. In von Willebrand disease, the level of vWF is decreased or absent. Serum levels of factor VIII range from mildly to severely reduced. Absence of platelet adhesion at the site of vascular injury and deficient factor VIII activity in the intrinsic coagulation pathway contribute to the bleeding seen in von Willebrand disease.

Clinical manifestations. Epistaxis, mucosal bleeding, ecchymoses, gastrointestinal bleeding, and menorrhagia are common clinical manifestations of von Willebrand disease. Once hemostasis is achieved, it can usually be maintained. Hemarthrosis is rare. Although not common, von Willebrand disease should be considered a possible cause of excessive surgical bleeding. Bleeding manifestations may decrease during pregnancy because levels of VWF and factor VIII rise during this time.

Diagnosis and treatment. The history and clinical presentation initially suggest the possibility of von Willebrand disease as the cause of bleeding. Laboratory tests consistent with the disease include a prolonged bleeding time, prolonged aPTT, normal platelet count, and normal PT/INR. More specialized testing will verify that the level of plasma vWF is decreased and that factor VIII activity is reduced.

Mild forms of classic von Willebrand disease can be managed with desmopressin, which causes release of vWF and factor VIII from vascular endothelial cells. Excessive menstrual bleeding can be treated with hormonal suppression therapy. Severe bleeding is addressed by using cryoprecipitate that contains both factor VIII and vWF. Recombinant replacement therapy is available. Aspirin and aspirin-containing drugs, which inhibit normal platelet function in hemostasis, should be avoided in patients with von Willebrand disease.

Complications of therapy for severe von Willebrand disease include hepatitis and AIDS, related to transfusions with blood products. Antibodies that inhibit the activity of vWF may develop, but this is rare.

Vitamin K Deficiency Bleeding in Infancy

Etiology. As the name implies, this coagulation disorder is seen in the newborn, typically 48 to 72 hours after birth, through 6 months of age. Hemorrhagic disease of the newborn is more common in breast-fed babies (who do not receive vitamin K supplement) than in formula-fed babies. It is rare in Western countries because of routine administration of vitamin K to newborns.

Pathogenesis. This bleeding disorder results from a deficiency of the vitamin K–dependent coagulation factors II, VII, IX, and X. The levels of these factors are approximately 50% of normal in umbilical cord blood, and they decline rapidly after birth, reaching their lowest levels at 48 to 72 hours. In a small number of infants, the decline is so significant that severe bleeding occurs. After 72 hours, the levels of these coagulation factors gradually increase over the course of several weeks. This increase is primarily caused by absorption of vitamin K from the diet. The vitamin K content of human milk is very low compared with standard infant formulas; therefore breast-fed babies need vitamin K supplementation.

Hepatic immaturity may also contribute to hemorrhagic disease of the newborn. The liver may be unable to initially produce adequate levels of the vitamin K–dependent coagulation factors.

Clinical manifestations. Evidence of bleeding, such as melena (tarry, black feces composed of partially digested blood), bleeding from the umbilicus, and hematuria, appears on the second or third day of life. Life-threatening complications include intracranial hemorrhage and hypovolemic shock.

Diagnosis and treatment. The diagnosis is based on the clinical presentation, particularly the timing of the onset of bleeding. The PT/INR is prolonged; levels of vitamin K–dependent clotting factors are decreased.

Prophylactic administration of vitamin K to the newborn prevents the severe decline of the vitamin K–dependent coagulation factors and largely eliminates this coagulation disorder.

If evidence of hemorrhage is present, vitamin K should be administered. For severe hemorrhage, fresh plasma will replenish the deficient coagulation factors and stop the bleeding. Fresh whole blood will correct severe anemia and shock.

Premature infants may experience bleeding attributable to platelet abnormalities and a deficiency in several coagulation factors. Because of hepatic immaturity, vitamin K is ineffective therapy in these infants. Fresh plasma is the treatment of choice for the premature infant with bleeding complications.

Acquired Vitamin K Deficiency

Etiology. Acquired vitamin K deficiency may result in bleeding as a result of a coagulation defect. Vitamin K, a fat-soluble vitamin, is obtained by the ingestion of specific foods (e.g., liver, cheese, butter, egg yolks, and green, leafy vegetables) and by a synthetic process occurring in the intestinal flora. Vitamin K is then absorbed by the intestine and stored in the liver. Normal absorption is dependent on bile acids and adequate mucosal function in the intestine. Vitamin K is necessary for normal synthesis and function of coagulation proteins (factors II, VII, IX, and X) as well as coagulation inhibitors (proteins C and S).

Vitamin K deficiency, with its associated risk for bleeding, may occur with the following: malnutrition, malabsorption (including biliary disease), chronic hepatic disease, antibiotic therapy, and oral anticoagulation therapy with warfarin (Coumadin).

Pathogenesis. One of the many functions of the liver is the synthesis and transport of bile, which is necessary for fat digestion and normal absorption in the small intestine. Vitamin K is a fat-soluble vitamin; if fat malabsorption occurs because of a lack of bile, vitamin K is not absorbed, resulting in a vitamin K deficiency. In the newborn, especially the premature infant, vitamin K deficiency may be related to liver immaturity and the lack of vitamin K synthesis by the intestine until the gut is colonized with the flora that produce vitamin K. Coumadin-type drugs are vitamin K antagonists that inhibit the normal activity of vitamin K in the synthesis of clotting factors. The net effect is decreased clotting factor activity.

Although vitamin K is deficient, the liver continues to synthesize some vitamin K–dependent coagulation factors. However, the coagulation activity of these factors is impaired, resulting in bleeding.

Clinical manifestations. Evidence of bleeding may present in a variety of ways, including mucosal and gastrointestinal bleeding, ecchymoses, menorrhagia, and hematuria. Surgical bleeding may be a significant problem in the patient with a vitamin K deficiency.

Diagnosis and treatment. Vitamin K deficiency should be considered the cause for bleeding when the PT/INR is increased but other coagulation studies are normal. Of the vitamin K–dependent clotting factors, factor

VII (extrinsic pathway) has the shortest half-life; thus the PT/INR is prolonged first. Ultimately, the aPTT also will be prolonged as clotting factors in the intrinsic pathway become deficient.

Parenteral administration of vitamin K rapidly restores levels in the liver. Fresh frozen plasma, with an immediate supply of clotting factors, is the treatment of choice for severe hemorrhage. Correction or removal of the cause of vitamin K deficiency also is an important part of therapy.

Disseminated Intravascular Coagulation (DIC)

Etiology. DIC is an acquired hemorrhagic syndrome in which both clotting and bleeding occur simultaneously (Fig. 14.10). This syndrome is also known as *disseminated intravascular coagulopathy* or *disseminated intravascular consumption* in some references. Widespread clotting in small vessels leads to consumption of the clotting factors and platelets, which in turn leads to bleeding. DIC is either chronic or acute. The chronic form is seen mainly in the cancer patient with malignancy and presents in a less severe form with bleeding tendencies that are mild to moderate and thrombotic episodes. The liver and bone marrow have sufficient time to replenish consumed factors and platelets, which leads to a more pronounced thrombotic problem. Acute DIC occurs secondary to a variety of factors, including malignancy, sepsis, snake bite, abruptio placentae, trauma and crushing injuries, transfusions of incompatible blood, burns, shock, and severe liver disease. DIC is estimated to occur in 1 of every 900 to 2400 adult admissions in large, urban hospitals. Death rates are reported to range from 50% to 80%.

Pathogenesis. DIC represents a paradox of both thrombosis and hemorrhage. Accelerated intravascular clotting in small vessels is initiated by contact of the blood with damaged vascular endothelium (sepsis, burns), release of procoagulant substances into the blood (snake venom, malignancy), generation of procoagulants in the blood (incompatible blood transfusion), or stagnant blood flow (shock). Coagulation factors, especially prothrombin, platelets, factor V, and factor VIII, are rapidly consumed. At the same time, the fibrinolytic system is activated to break down the clots. The fibrin degradation products or fibrin split products that result act as circulating anticoagulants. The combination of coagulation, anticoagulation, and fibrinolysis ultimately leads to hemorrhage.

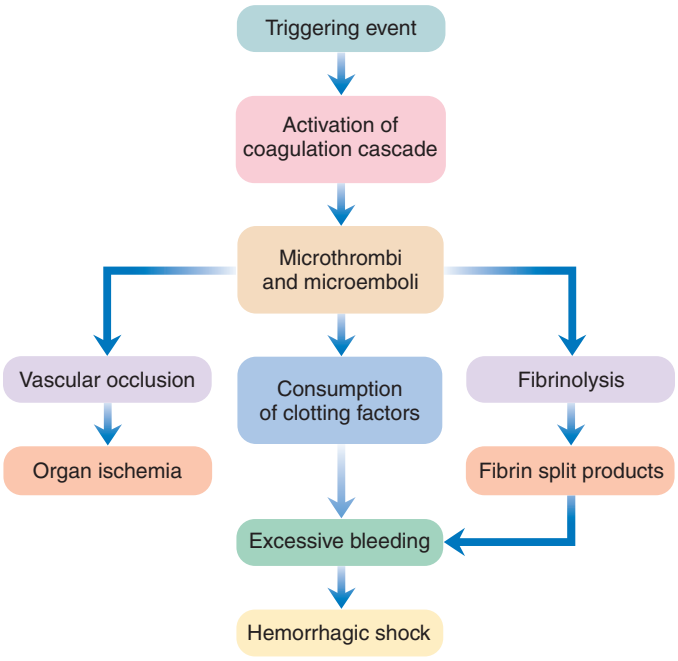


FIG 14.10 Pathophysiology of disseminated intravascular coagulation. Clotting and bleeding occur simultaneously, resulting in organ ischemia and hemorrhagic shock.

Clinical manifestations. Although both bleeding and clotting are part of the syndrome, initially bleeding is more apparent clinically. Petechiae and ecchymoses on skin and mucous membranes, as well as bleeding from orifices and any site of injury, such as venipuncture and injection sites, may be present. Acrocyanosis (cold, mottled fingers and toes) may be apparent, attributable to thrombi formation in the microvasculature of the extremities. Thrombi in the pulmonary microcirculation (small vessels) may result in dyspnea, hemoptysis, and crackles or rales, as blood fills alveoli. Patients with DIC are also predisposed to acute renal failure because of the presence of microthrombi in the renal microvasculature.

Diagnosis and treatment. The diagnosis of DIC is based on a high index of suspicion drawn from the history and presenting signs and symptoms. The typical clinical picture described previously, plus the presence of a predisposing cause, should make DIC a consideration in the differential diagnosis. Abnormal coagulation studies that help confirm the diagnosis include increased values for bleeding time, PT/INR, aPTT, fibrin split products, and thrombin time as well as decreased measurements for the fibrinogen level and platelet count. D-Dimer (a fibrin degradation product) is one of the most useful tests to measure fibrinolysis; this in conjunction with an elevated antithrombin complex is indicative of DIC.

The cornerstone of treatment for DIC is removal or correction of the underlying cause and support of major organ systems. Replacement of depleted clotting factors with fresh frozen plasma, packed red blood cells, platelets, or cryoprecipitate may be necessary. Antifibrinolytics (aminocaproic acid) may be used if there is life-threatening hemorrhage. Some studies have had promising results in decreasing mortality with administration of antithrombin concentrate or activated protein C. Improved mortality was seen in the groups with higher severity ratings. Although controversial, heparin may be used to minimize further consumption of clotting factors. The purpose of heparin therapy is to stop thrombin formation, thus preventing microemboli. Low-dose subcutaneous heparin appears to be as effective as high-dose heparin with fewer complications. Heparin has been found useful in chronic DIC.

Hepatic Disease

Etiology. A common complication of many hepatic disorders is abnormal hemostasis. With the exception of part of the antihemophilic factor, all plasma protein clotting factors and fibrinolytic factors and their inhibitors are synthesized totally or predominantly by the liver. If liver function is altered by disease, bleeding is one manifestation.

Pathogenesis. Several factors may contribute to the abnormal hemostasis seen in liver disease. Liver disease alters the synthesis and transport of bile, which is necessary for normal fat digestion and absorption. Impaired absorption and metabolism of vitamin K, which is fat soluble, results in decreased hepatic synthesis of coagulation factors II, VII, IX, and X. Altered liver function also results in decreased synthesis of fibrinogen and factors V and XI. A deficiency in any of the coagulation factors can interrupt the normal process of fibrin clot formation. In addition to synthesis of coagulation factors, the liver has a role in removing activated coagulation and fibrinolytic proteins from the circulation. Failure to filter these proteins adequately may result in an imbalance between clot formation and clot dissolution (fibrinolysis), manifesting clinically as DIC. Liver disease also may alter normal production of inhibitors of coagulation (antithrombin III, proteins C and S), which contributes to the hypercoagulable component of DIC.

Another factor contributing to the bleeding associated with liver disease is thrombocytopenia. A low platelet count is common in liver

disease. The exact mechanism is unknown but may relate to the splenomegaly associated with portal hypertension. Sequestration of platelets in the enlarged spleen depletes the number of platelets circulating and available for normal hemostasis. The portal hypertension that develops as blood flow through the liver is obstructed adds to the bleeding problem. While pressure in collateral circulatory beds increases, bleeding is manifested as esophageal varices and hemorrhoids (see Chapter 38).

Clinical manifestations. Patients with chronic rather than acute liver disease are more likely to have clinical evidence of a bleeding problem. Typical clinical features may include any of the following: petechiae, ecchymoses, spider telangiectasia, bleeding from venipuncture sites or esophageal varices, and gastrointestinal bleeding. DIC may complicate the clinical presentation. Bleeding may not be a problem until the patient has surgery or a biopsy.

Diagnosis and treatment. The patient with liver disease and associated bleeding will commonly have a decreased platelet count, normal or decreased fibrinogen levels, and prolonged PT/INR and aPTT values. More specific coagulation studies may be indicated in some situations.

Treatment may be instituted prophylactically before surgery or biopsy, or it may be mandated by a bleeding episode. The degree of abnormality on coagulation tests or the severity of the bleeding will influence the aggressiveness of therapy. Because of the high likelihood of vitamin K deficiency, administration of vitamin K may be the initial intervention. Platelet infusions are appropriate if significant thrombocytopenia is present. Fresh frozen plasma is the primary replacement product used to supply coagulation factors. Administration of large quantities of plasma carries the risk of precipitating hepatic encephalopathy and fluid overload. Transfusions of whole blood or, more commonly, packed red blood cells may be necessary to manage anemia from bleeding of significant proportions.

KEY POINTS

- Coagulation disorders result from defects in the clotting cascade or fibrinolytic process. These disorders may be inherited or acquired.
- Hemophilia is an inherited bleeding disorder that results from deficient clotting factor production. The most common types are hemophilia A (factor VIII) and hemophilia B (factor IX).
- von Willebrand disease is an inherited bleeding disorder caused by abnormal factor VIII carrier protein production. The disease results in a deficiency of factor VIII in the circulation and decreased platelet function.
- Vitamin K deficiency is associated with several coagulation disorders, including hemorrhagic disease of the newborn and bleeding related to malnutrition and liver disease. Vitamin K is a necessary cofactor for liver production of factors II, VII, IX, and X.
- Disseminated intravascular coagulation (DIC) is an acquired bleeding syndrome associated with a number of etiologic factors, including trauma, malignancy, burns, shock, and abruptio placentae. DIC is characterized by widespread clot formation in small vessels. Clotting factors and platelets are consumed, leaving the patient with deficient resources for appropriate clot formation. The platelet count and fibrinogen levels are typically decreased, and values for PT, aPTT, thrombin time, bleeding time, and fibrin split products are elevated.

SUMMARY

The presence of unexpected overt or occult bleeding may signal an acquired or inherited problem with hemostasis. A review of normal hemostasis, as well as information on selected disorders of hemostasis and coagulation, has been presented in this chapter. With a sound knowledge base, the health care professional is in a position to play a key role in the recognition, diagnosis, and management of a bleeding problem.

RESOURCES

Platelet Function and Hemostasis

Casper D: *Harrison's principles of internal medicine*, ed 19, New York, 2015, McGraw-Hill.

Hall JE: *Guyton and Hall textbook of medical physiology*, ed 13, Philadelphia, 2016, Saunders.

McCance K, Huether S, Brashers V, Rote N: *Pathophysiology the biologic basis for disease in adults & children*, ed 7, St Louis, MO, 2015, Mosby Elsevier.

Pagana KD, Pagana TJ: *Mosby's manual of diagnostic and laboratory tests*, ed 12, St Louis, MO, 2015, Mosby.

Quinn M: Platelet physiology. In Quinn M, et al, editors: *Platelet function*, New York, 2010, Humana Press.

Vascular and Platelet Disorders

Bolton-Maggs P, et al: A review of inherited platelet disorders with guidelines for their management, on behalf of the UKHCDO. *Br J Haemotol* 135:603–633, 2006.

Chen O, et al: Henoch-Schonlein purpura in children: clinical analysis of 120 cases. *Afr Health Sci* 13(1):1–6, 2013.

Faughnan M, et al: International guidelines for the diagnosis and management of hereditary haemorrhagic telangiectasia. *J Med Genet* 48:73–87, 2011.

Gallo RL: Vascular purpuras. In Kaushansky K, et al, editors: *Williams hematology*, ed 9, New York, 2010, McGraw-Hill.

George J: Platelets. *Lancet* 355:1531–1539, 2000.

Germain D: Ehlers-Danlos syndrome type IV. *Orphanet J Rare Dis* 2:32–40, 2007.

Israels SJ, Kahr WH, Blanchette VS, et al: Platelet disorders in children: a diagnostic approach. *Pediatr Blood Cancer* 56(6):975–983, 2011.

McDonald J, Bayrack-Toydemir P, Pyeritz R: Hereditary hemorrhagic telangiectasia: an overview of diagnosis, management and pathogenesis. *Genet Med* 13(7):607–616, 2011.

Newton J, et al: Fatigue in adult patients with primary immune thrombocytopenia. *Eur J Haemotol* 86:420–429, 2011.

Provan D, et al: International consensus report on the investigation and management of primary immune thrombocytopenia. *Blood* 115:168–186, 2010. doi:10.1182.

Roberts P, et al: Henoch-Schönlein purpura: a review article. *South Med J* 100(8):821–824, 2007.

Segal G, Feig S: Controversies in the diagnosis and management of childhood acute immune thrombocytopenic purpura. *Pediatr Blood Cancer* 53:318–324, 2009.

Shimaoka Y, Kosho T, et al: Clinical and genetic features of 20 Japanese patients with vascular-type Ehlers-Danlos syndrome. *Br J Dermatol* 163(4):704–710, 2010.

Stasi R, Newland A: ITP: a historical perspective. *Br J Haemotol* 153(4): 437–450, 2011.

Thakrar SV, Mallett SV: Thrombocytopenia in cirrhosis: Impact of fibrinogen on bleeding risk. *World J Hepatol* 9(6):318–325, 2017.

Coagulation Disorders

Baronciani L, Goodeve A, Peyvandi F: Molecular diagnosis of von Willebrand disease. *Haemophilia* 23(2):188–197, 2017.

Boral BM, Williams DJ, Boral LI: Disseminated intravascular coagulation. *Am J Clin Pathol* 146(6):670–680, 2016.

Hoffman R, et al: *Hematology basic principles and practices*, ed 6, Philadelphia, 2013, Saunders Elsevier.

Kasper C, Lin C: Prevalence of sporadic and familial haemophilia. *Haemophilia* 13:90–92, 2007.

Leebeek FW, Eikenboom JC: Von Willebrand's Disease. *N Engl J Med* 375(21):2067–2080, 2016.

Levi M: Disseminated intravascular coagulation. *Crit Care Med* 35(9): 2191–2195, 2007.

Lima H: The diagnosis and management of von Willebrand disease. *Infusion* 16(6):1–11, 2010.

Lippi G, Franchini M: Vitamin K in neonates: facts and myths. *Blood Transfus* 9:4–9, 2011.

Pacheco L, et al: von Willebrand disease and pregnancy: a practical approach for the diagnosis and treatment. *Am J Obstet Gynecol* 203(3):194–200, 2010.

Stachnik J: Hemophilia: etiology, complications, and current management. *Formulary* 45:218–227, 2010.

Van Winckel M, et al: Vitamin K an update for the paediatrician. *Eur J Pediatr* 168(2):127–134, 2009.

World Federation of Hemophilia: *Guidelines for the management of hemophilia*, 2005. www.wfh.org/2/docs/Publications/Diagnosis_and_Treatment/Guidelines_Mng_Hemophilia.pdf.

15

Alterations in Blood Flow

Teresa Grigsby Loftsgaarden

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- How do the structures of arteries, veins, capillaries, and lymphatics differ, and how do these differences reflect the functions of each?
- What is the relationship among vessel resistance, blood pressure, and blood flow?
- How is vascular resistance regulated centrally by the autonomic nervous system and locally by tissues?
- What are the determinants of transcapillary exchange of fluids, electrolytes, and nutrients?
- How do arterial and venous obstructions develop?
- What are the clinical consequences of acute and chronic arterial obstruction?
- What are the clinical consequences of superficial and deep venous obstructions?

CHAPTER OUTLINE

Organization of the Circulatory and Lymphatic Systems, 315
Vessel Structure, 315
Anatomy of Arteries and Veins, 316
Anatomy of Capillaries, 316
Lymphatic Structure, 317
Principles of Flow, 317
Hemodynamics of the Circulatory System, 317
Blood Flow, Pressure, and Resistance, 318
Velocity and Laminar and Turbulent Flow, 319
Wall Tension and Compliance, 320
Dynamics in the Microcirculation: Capillaries and Lymphatics, 321
Control of Flow, 322
Control of Blood Flow, 322
Extrinsic Mechanisms, 322
Intrinsic Mechanisms, 322
Control of Lymphatic Flow, 323
General Mechanisms That Cause Altered Flow, 323
Blood Vessels: Obstructions, 323
Thrombus, 323
Embolus, 324
Vasospasm, 325
Inflammation, 325
Mechanical Compression, 325
Blood Vessels: Structural Alterations, 325
Types of Structural Alterations, 325
Lymphatic Vessels, 326
Alterations in Arterial Flow, 326
Arteriosclerosis/Atherosclerosis, 326
Etiology and Pathogenesis, 326
Risk Factors, 327
Clinical Manifestations and Diagnosis, 330
Treatment, 330
Thromboangiitis Obliterans (Buerger Disease), 330
Raynaud Syndrome, 330
Aneurysms, 330
Classifications, 330
Clinical Manifestations and Diagnosis, 331
Treatment, 331
Acute Arterial Occlusion, 331
Alterations in Venous Flow, 332
Valvular Incompetence, 332
Etiology and Pathogenesis, 332
Clinical Manifestations and Treatment, 332
Varicose Veins, 332
Etiology and Pathogenesis, 332
Clinical Manifestations and Treatment, 332
Chronic Venous Insufficiency, 333
Etiology and Pathogenesis, 333
Clinical Manifestations and Treatment, 333
Deep Vein Thrombosis, 333
Etiology and Pathogenesis, 333
Clinical Manifestations and Treatment, 333
Alterations in Lymphatic Flow, 333
Lymphedema, 333
Etiology and Pathogenesis, 333
Clinical Manifestations, Diagnosis, and Treatment, 334

The primary functions of the circulatory system are the transportation of oxygen and nutrients and the removal of metabolic waste products within the body. To perform these functions, a complex circuitry of vessels traverses the body (Fig. 15.1), powered by the pumping action of the heart. Propulsion of blood through the lungs is provided by the right ventricle, whereas systemic blood flow is driven by the left ventricle.

Nutrients are absorbed into the blood as it moves through the gastrointestinal tract via the splanchnic circulation. Oxygen uptake and the release of carbon dioxide occur in the specialized vascular bed of the pulmonary circulation. The liver, with its extensive blood supply, has a major role in metabolism and generation of metabolic waste products. These, and other metabolic by-products, are carried by the blood to the

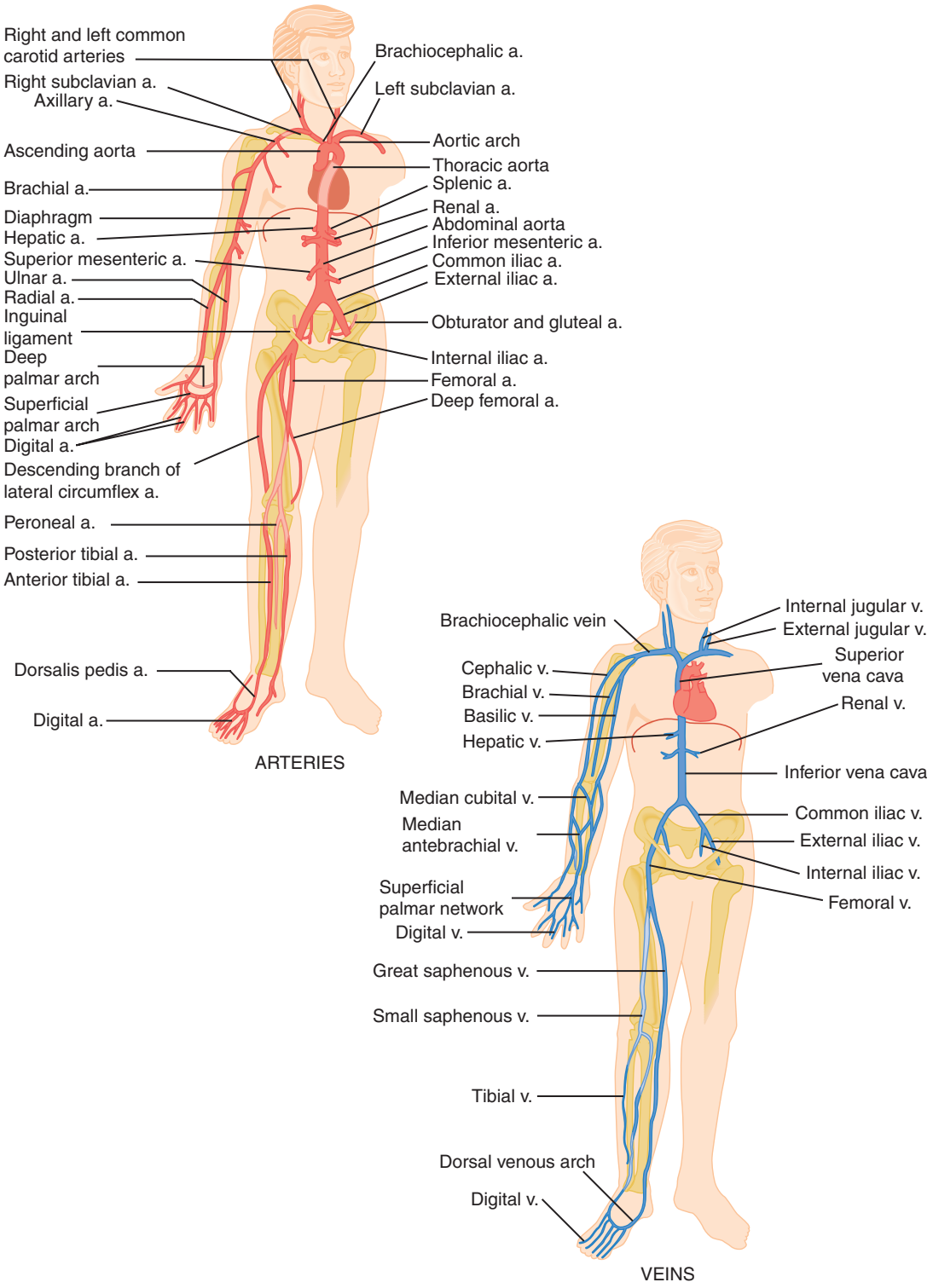


FIG 15.1 The primary systemic arterial and venous circulatory networks. a., Artery; v., vein. (From Black JM, Hawks J: *Medical-surgical nursing: clinical management for positive outcomes,* ed 8, Philadelphia, 2008, Saunders, p 1273.)

kidneys for elimination. Inadequate circulation in the lungs, liver, or kidneys may interfere with the removal of metabolic wastes from the body. Effective transportation of oxygen and nutrients and removal of waste materials depend on proper functioning of the circulatory system.

Aging produces significant changes in the circulatory system, altering the ability of the system to carry out its functions and increasing susceptibility to certain disease processes. The effects of the aging process on the circulatory system are summarized in the box "Geriatric Considerations: Changes in the Circulatory System."

ORGANIZATION OF THE CIRCULATORY AND LYMPHATIC SYSTEMS

After passing through the pulmonary circulatory system and leaving the left ventricle, blood flows through a graduated series of tubes to tissues of the body before returning to the right side of the heart. The powerful left ventricle propels the blood to the aorta, arteries, arterioles, and, finally, to the capillary beds. Here the proximity of capillary endothelium to the other cells of the body facilitates movement of nutrients and oxygen into the cells and removal of cellular metabolic wastes. Capillary blood is then collected by venules, which flow into veins, returning blood to the venae cavae and the right side \of the heart (Fig. 15.2). The complete process, moving approximately 5 L of blood through the entire circuit, takes only about 1 minute.

The lymphatic circulation is a specialized system of channels and tissues (nodes). It is not arranged in a circuit, as is the vascular system. Instead, the lymphatic vessels begin blindly, deep in the connective tissue. One of the functions of the lymphatic system is to reabsorb fluid that leaks out of the vascular network into the interstitium and return it to the general circulation. During the process of cellular exchange within the capillary bed, some fluid moves into the interstitium and fails to return to the vascular bed. This lost fluid can amount to as much as 2 to 4 L/day. At this circulatory level, lymphatic vessels lie in close proximity to the capillary vasculature. The fluid, now called *lymph,* is absorbed by the lymphatic vessels and returned to the venous circulation by way of the thoracic duct and the right lymphatic duct (Fig. 15.3). Lymphatic drainage of the breast is illustrated in Fig. 15.4.

Vessel Structure

To perform their specialized functions, the blood and lymphatic vessels are different in their structure. Knowledge of the morphology of these vessels is essential for an understanding of the alterations in function produced by disease.

The primary differences between the smaller arterial and venous vessels are in terms of the quantities of muscle and connective tissue

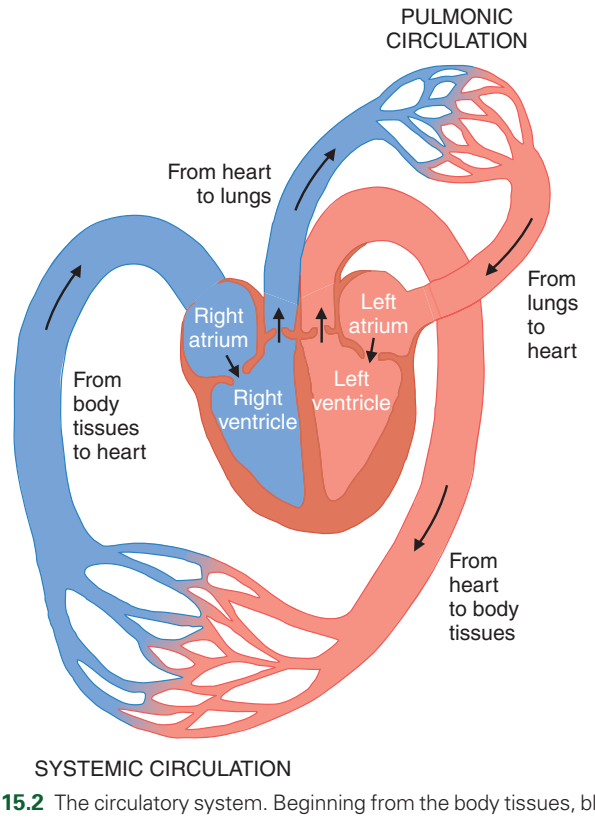


FIG 15.2 The circulatory system. Beginning from the body tissues, blood returns to the right side of the heart, through the right atria to the right ventricle, which propels it into the lungs. In the lungs, the metabolic waste carbon dioxide is removed and oxygen is replenished. Oxygenated blood leaves the pulmonic circulation and returns to the heart via the left atrium and then to the left ventricle. From the left side of the heart, the oxygenated blood enters the systemic circulation, where oxygen is delivered to the tissues in exchange for metabolic wastes. (From Black JM, Hawks J: *Medical-surgical nursing: clinical management for positive outcomes,* ed 8, Philadelphia, 2008, Saunders, p 1344.)

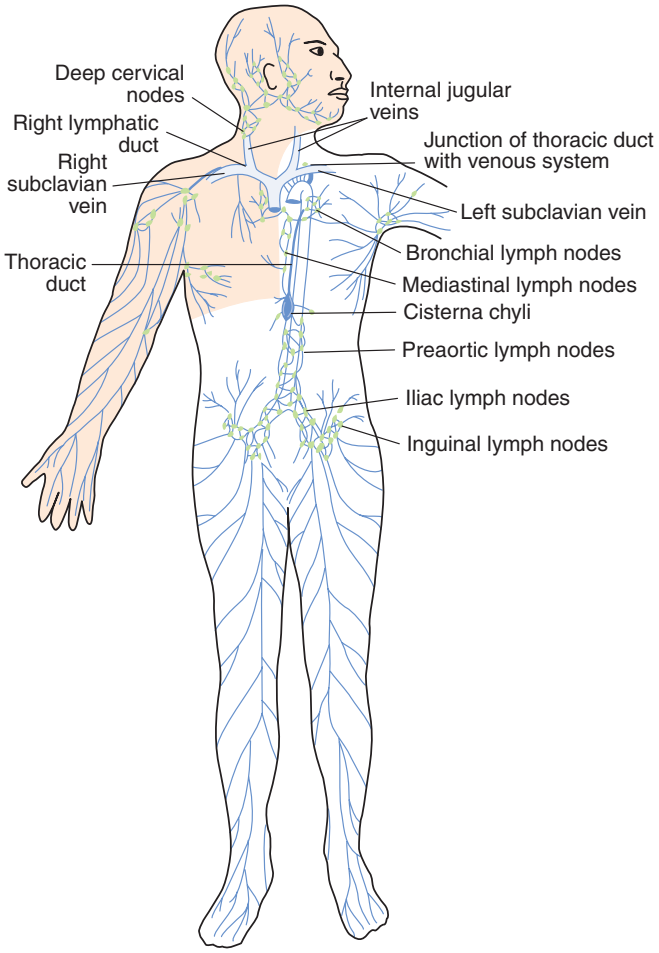


FIG 15.3 Anatomy of the lymphatic system. Lymphatic capillaries collect the excess fluid from the vascular capillaries, returning it to the venous circulation at the junction of the internal jugular and subclavian veins. (From Monahan FD et al: *Phipps' medical-surgical nursing: health and illness perspectives,* ed 8, Philadelphia, 2007, Mosby, p 936.)

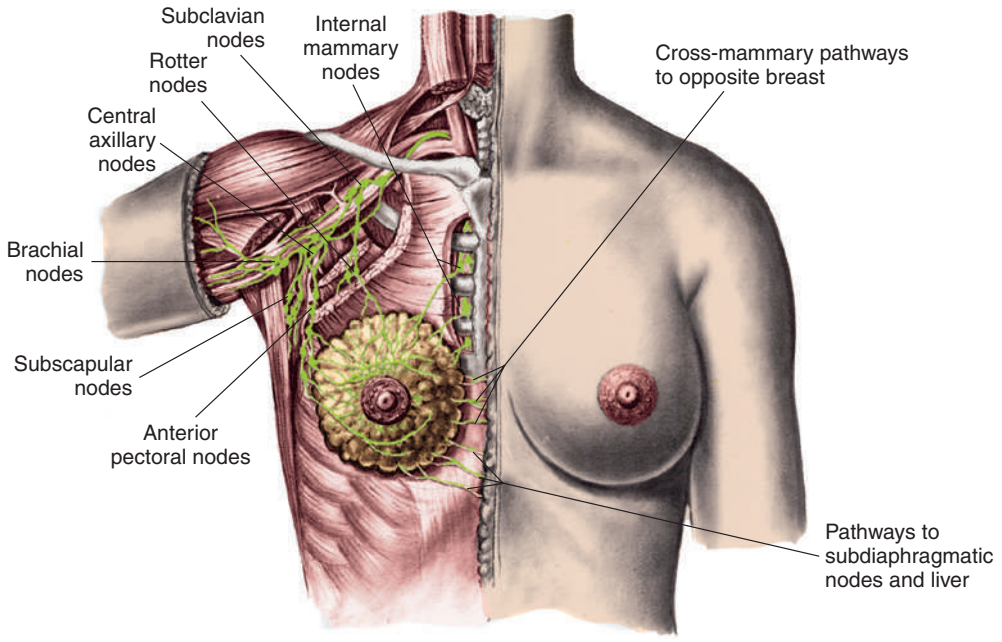


FIG 15.4 Lymphatic drainage of the breast to the axillary and subclavian nodes, then to the jugular and subclavian veins. (From Monahan FD et al: *Phipps' medical-surgical nursing: health and illness perspectives,* ed 8, Philadelphia, 2007, Mosby, p 1661.)

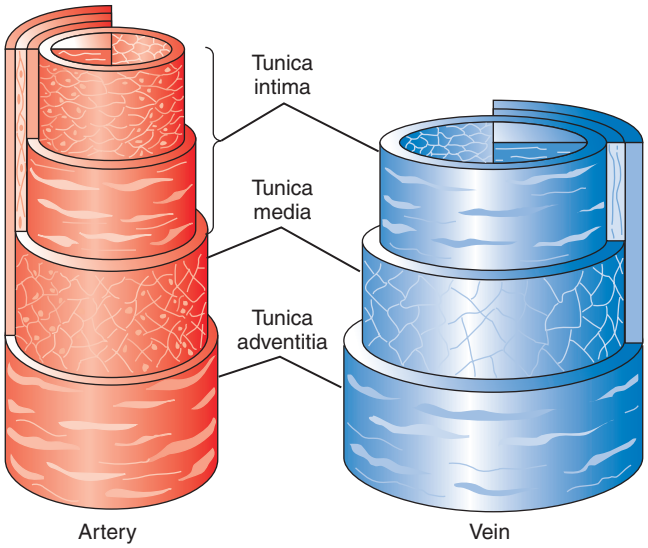


FIG 15.5 Tunicae of arteries and veins showing the thicker walls of the arteries.

present. In arterioles, the principal tissue is smooth muscle, whereas in venules, smooth muscle is scarce, and connective tissue dominates. The composition of the walls and the size and shape of the vessels also vary in larger arteries and veins. Capillary walls are composed of a single layer of endothelial cells. These simple structures carry out extraordinarily complex functions.

Anatomy of Arteries and Veins

The walls of both arteries and veins are composed of three microscopically distinct layers, or tunicae: the intima, the media, and the adventitia. The histologic constituents of these tunicae are similar in arteries and veins (Fig. 15.5). Generally, the walls of veins are not as thick as the walls of arteries, but the lumina are larger.

The intima consists of a layer of endothelial cells that is in direct contact with the blood as it flows through the vessel. Periodically, the intimal layer of veins protrudes into the lumen, creating valves that prevent the backflow of blood. Arterial intima is characterized by an inner elastic membrane next to the endothelial cells. This elastic membrane is thickest in the aorta and decreases in density until only scattered elastic fibers can be identified in the smallest arterioles. With increasing age, the intimal arterial wall becomes thicker and less elastic. This interferes with diffusion of nutrients into the wall, causing the internal elastic membrane to degenerate and calcify.

The media, or middle layer, exhibits the greatest difference between arteries and veins. In arteries the media is the thickest of all the tunicae. Large arteries have smooth muscle fibers arranged in a circular pattern and interspersed with elastic fibers. Progressing from arteries to ever-smaller arterioles, the smooth muscle remains, but the elastic tissue disappears. This thick, smooth muscle layer is responsible for the firmness and limited distensibility of arterial vessels. With advancing age, changes in the intima result in decreased nutrition reaching the media, causing degeneration of the smooth muscle tissue. In veins, the media also has smooth muscle, usually arranged in a circular pattern with some longitudinal strands. The quantity of smooth muscle decreases as the veins become larger. Venous media also contains collagenous connective tissue, but elastic tissue is rare, except in the large veins.

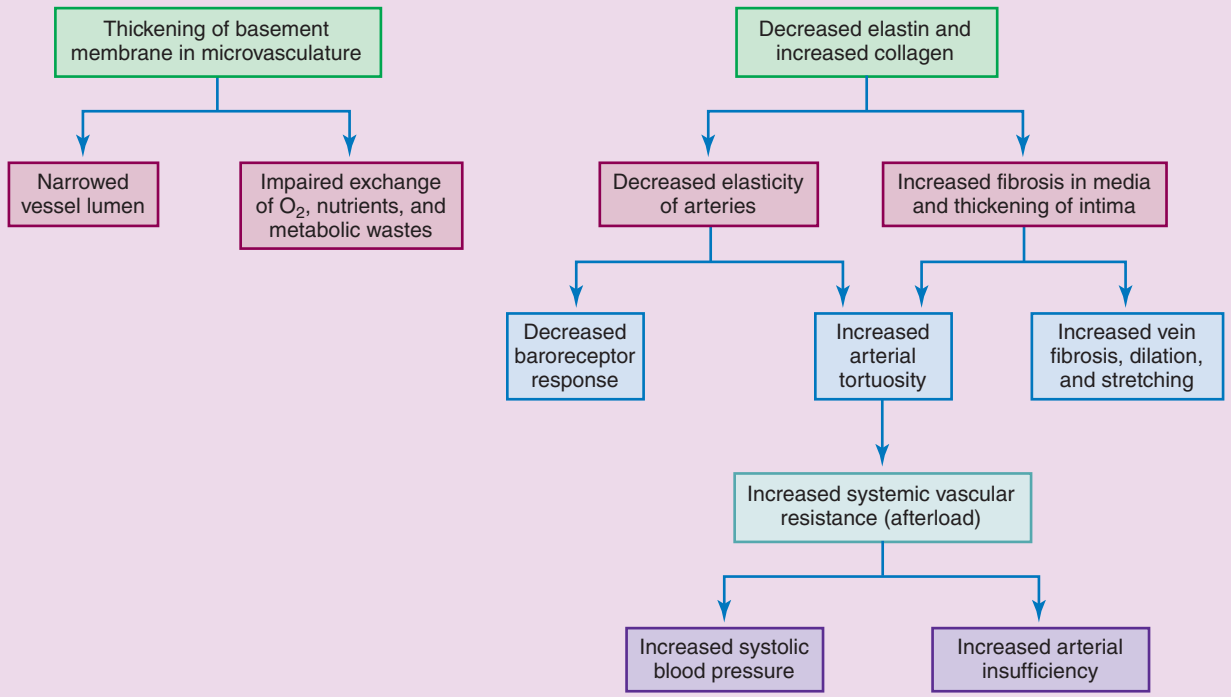
In veins, the adventitia is the thickest of the tunicae. It is composed of collagenous connective tissue and longitudinal smooth muscle. In larger arteries there is a discernible external elastic membrane in the adventitia. This membrane disappears as the arteries decrease in diameter. Arterial adventitia consists predominantly of collagenous connective tissue. Some larger vessels also contain isolated, longitudinally arranged fibers of smooth muscle.

Anatomy of Capillaries

Capillaries are composed of a single thickness of endothelial cells attached to a protein network called the *basement membrane.* Moving from the end of an arteriole to the beginning of a venule, capillaries narrow to

GERIATRIC CONSIDERATIONS

Changes in the Circulatory System



In the aging individual, changes occur throughout the vascular bed. The microvascular bed demonstrates thickening of the basement membrane. This change narrows the vessel lumen and impairs the free exchange of oxygen, nutrients, and metabolic wastes at the cellular level.

In both arteries and veins, the vascular changes occur first in the proximal portions. The intima becomes fibrotic and the endothelial cell variation increases. In the media, the amount of elastin and smooth muscle is reduced, whereas the amount of fibrotic and collagen tissue increases. With collagen cross-linking, the vessel walls lose elastic flexibility and recoil, becoming stiffer and less compliant. They become inflexible tubes with an increase in systemic vascular resistance (SVR). The increased SVR causes a reduction in tissue and organ blood flow and decreased perfusion.

Baroreceptor function is reduced because of decreased sensitivity of the receptors and diminished responsiveness of the vessels attributable to their rigidity. These factors decrease the body's ability to respond to hypotensive and hypertensive stimuli. The decreased compliance of the systemic vascular system increases afterload, forcing the left ventricle of the heart to work harder to meet the metabolic demands of the body.

a diameter barely sufficient for a single red blood cell (RBC) to pass through the aperture. In some tissues, one or two smooth muscle cells form a precapillary sphincter that controls flow through the vessel (Fig. 15.6).

There are spaces between the endothelial cells that vary in size among organ systems. These spaces, or pores, permit certain constituents to pass in and out of the capillaries. For example, capillary beds in the brain have little or no spaces and permit the passage of only certain molecules. The space between endothelial cells of the brain is so small that it is referred to as the *blood–brain barrier.* In parts of the kidneys, however, capillaries are more porous, allowing much larger molecules to move between the circulation and the filtrate (urine). The size of these spaces determines the *capillary permeability* of a specific capillary bed.

Lymphatic Structure

Lymphatic vessels are thin walled, and most resemble veins in their appearance. Like their counterparts in the circulatory system, they range in size from lymphatic capillaries to vessels of increasing diameter. Like veins, lymphatics have intermittent valves composed of folds of their inner layer that extend into the lumen (Fig. 15.7). The walls of lymphatic capillaries contain contractile fibers that are stimulated when stretched, causing the vessels to contract and propel lymph along the vessel.

KEY POINTS

- Arteries and veins have three distinct layers. The intima, the innermost layer, is composed of a single layer of endothelial cells. The media, or middle layer, is composed of smooth muscle and elastin. Media is thicker in arteries than in veins. The adventitia, the outermost layer, is composed of supporting connective tissue.
- Capillaries have only a single layer of endothelial cells attached to a basement membrane. The permeability of capillaries is determined by the tightness of the endothelial cell connection.
- Lymphatic vessels resemble veins, having thin walls and valves.

PRINCIPLES OF FLOW

Hemodynamics of the Circulatory System

The principles of blood flow are known as circulatory hemodynamics. These principles govern the quantity of blood passing by a given point in a specific period. Therefore blood flow is measured as a given number of liters, milliliters, or cubic centimeters per second, minute, or hour. A discussion of the hemodynamics of the circulatory system includes the concepts of pressure, resistance, velocity, laminar and turbulent flows, wall tension, and compliance.

Blood Flow, Pressure, and Resistance

Blood flow is accomplished by movement along a pressure gradient within the vascular bed. This means that blood moves from an area of higher pressure to an area of lower pressure. The arterial and arteriolar walls, with their muscular media coats, provide the high-pressure end of the gradient. Seeking a lower pressure, blood moves toward the venous system. The thinner, more pliable walls of the venous bed furnish the low-pressure portion of the pressure gradient. The greater the pressure difference, the greater the rate of blood flow.

The movement of blood through the vascular system is opposed by the force of *resistance*. The relationship between blood flow and resistance is an inverse one: as resistance increases, blood flow decreases. This force has several determinants, each of which can change resistance considerably; these determinants are represented in physiology by *Poiseuille's law:*

$$\text{Resistance} = \frac{8nl}{\pi r^4}$$

The number 8 is a mathematical constant, as is the value of π; n represents blood viscosity, l represents the length of a given vessel, and r is the radius of the vessel. Using this formula, the effects of changes on the components of resistance are very predictable.

Two of the determinants of resistance are *vessel length* and *vessel radius*. As predicted by Poiseuille's law, resistance changes directly with the length of the vessel, and these changes in resistance significantly affect flow. As illustrated in Fig. 15.8, given three vessels of the same radius, doubling the length increases the resistance and reduces the flow *(Q)* by 50%. Reducing the vessel length by half decreases resistance and increases the flow by 100%. These changes in flow occur when the pressure gradient remains constant and are caused solely by variations in vessel length. Resistance decreases as the radius of a vessel increases, and resistance is inversely related to the fourth power of the radius of a vessel, or r^4. Therefore increasing a vessel's radius markedly reduces resistance and produces an exponential increase in blood flow. Fig. 15.9 demonstrates the effect of doubling the radius of a vessel on the flow of blood if all other factors related to flow are held constant. The resulting flow of blood is 16 times greater in the larger-diameter vessel.

Although there is variability in the length of vessels throughout the circulatory system, vessels are incapable of altering their own length. They do, however, possess a considerable ability to change their diameters, and many disease processes (e.g., arteriosclerosis) and drug therapies (e.g., vasopressors) are associated with changes in the size of the vessel lumen. Even minor changes will produce major alterations in resistance and, hence, blood flow. This makes changes in diameter the most important determinant of resistance.

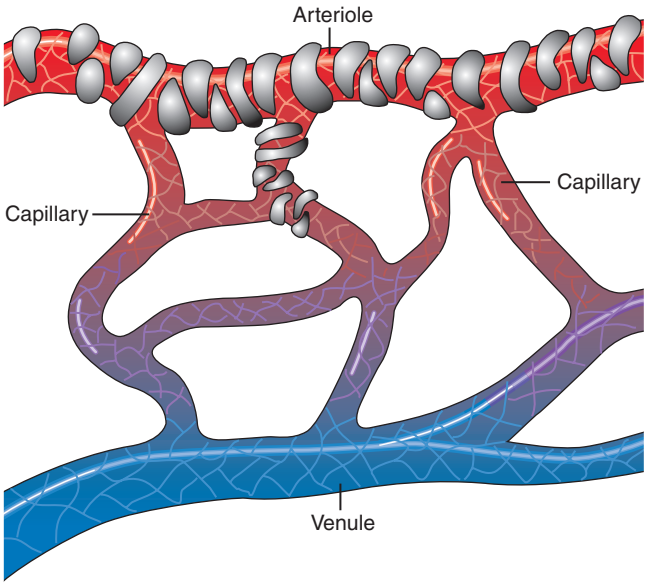


FIG 15.6 Capillary network.

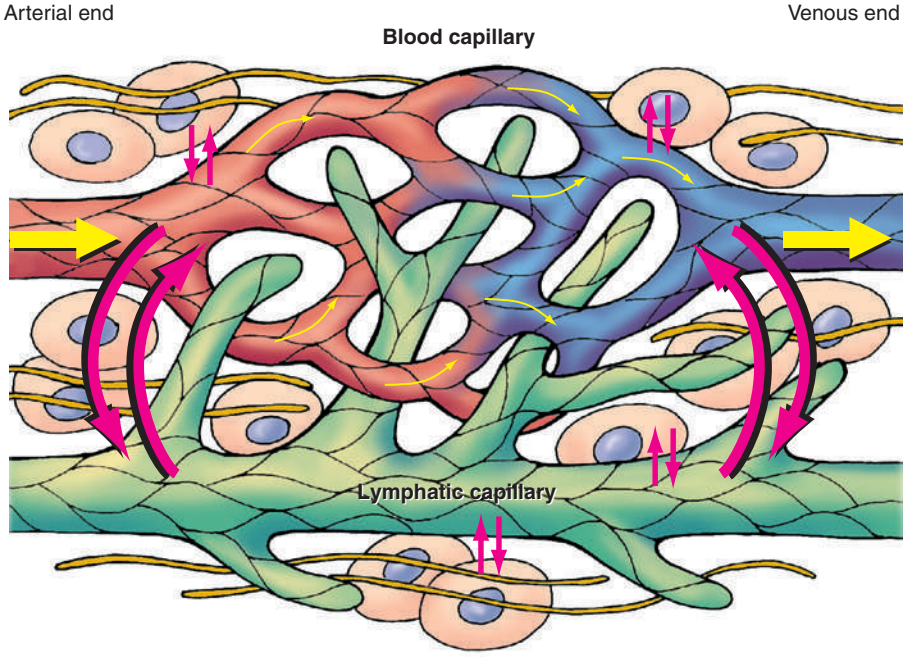


FIG 15.7 Lymphatic network. The lymphatic system is integrally related to the systemic vascular system. Excess fluid and plasma diffuse between the capillaries, interstitial spaces, and lymphatic vessels. Because lymphatic capillaries have larger spaces between endothelial cells, they can remove excess interstitial fluid or plasma that venous capillaries cannot reabsorb.

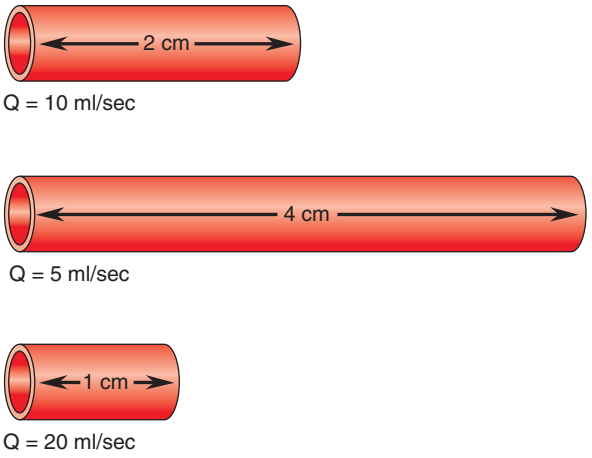


FIG 15.8 Relationship of vessel length to blood flow *(Q)* with a constant pressure gradient.

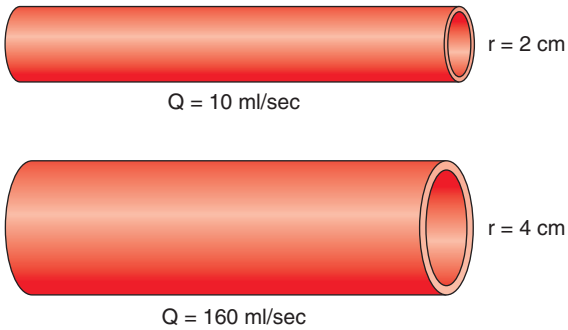


FIG 15.9 Relationship of vessel radius *(r)* to blood flow *(Q)* with a constant pressure gradient.

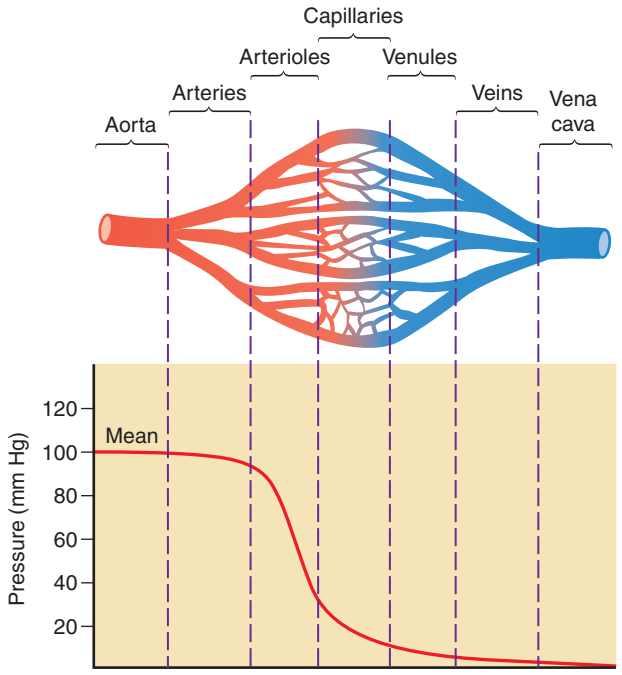


FIG 15.10 Mean pressure changes within the systemic vasculature. A significant decrease in pressure occurs as blood flows through the arterioles into the capillaries. The figure illustrates the role of the arterioles in the determination of vascular resistance. Because of the large number of capillaries, total resistance is not increased with the decreased radius of the capillaries.

The third determinant of resistance is the *viscosity* of the blood itself, represented in Poiseuille's law as *n*. Viscosity is defined as the thickness of a fluid. When the blood is more viscous, the friction between the cells and the liquid increases, and an increase in resistance to flow is produced. Blood is composed of a suspension of cellular material and plasma. Approximately 99% of the cellular constituents of the blood are RBCs. The ratio of RBCs to plasma is presented in the laboratory value *hematocrit*. Increasing the number of RBCs or decreasing the plasma component results in more viscous blood (increased hematocrit value), more resistance, and a slowing of blood flow. This is what occurs in dehydration, when the plasma component is relatively decreased, or in polycythemia, when the number of RBCs increases.

The relationship between the variables of driving pressure and resistance and their effect on blood flow is expressed by *Ohm's law,* as follows:

$$Q = P/R$$

Here, Q is the blood flow, P is the pressure difference between two points, and R is resistance. Altering any one of the determinants of resistance (vessel length, vessel radius, or blood viscosity) produces a change in flow. According to Ohm's law, a change in the pressure difference within the circulatory system also results in a change in the flow of blood. The arterioles are the major site of resistance in the vascular system and require a greater pressure to maintain blood flow. As the resistance decreases across the systemic vasculature, less pressure is necessary to maintain blood flow (Fig. 15.10). *Total peripheral resistance* refers to the resistance throughout the entire vascular system. It can be calculated on the basis of the pressure difference between the arteries and the veins. Clinically, SVR is used to specifically denote resistance peripheral to the heart and lungs. Because the primary determinant of SVR is the resistance vessels (arterioles), diseases or drug therapies that affect these vessels have the most profound impact on the SVR. Any condition that produces an increase in SVR, such as hypertension, requires more work for the heart to overcome the elevated resistance and eject its volume of blood (see Chapter 16). This increased workload means that the heart needs more oxygen and nutrients. When SVR is pathologically decreased, the blood is distributed over a larger area, and blood flow slows dramatically. Individual organs, such as the kidney and brain, may not obtain sufficient blood flow to meet metabolic needs. This is what occurs in distributive shock states (see Chapter 20).

Velocity and Laminar and Turbulent Flow

As previously discussed, blood flow is defined as the volume of blood that passes by a given point in a given unit of time. *Velocity* is a measure of the distance traveled in a given interval of time and is usually expressed in centimeters per second. Velocity is governed by the total cross-sectional area and varies inversely with it. An increase in the total cross-sectional area produces a decrease in velocity, whereas a decrease in the total cross-sectional area produces an increase in velocity. The total cross-sectional area of the aorta and vena cava is small, and they have the most rapid rate of flow, whereas the capillary beds combine to produce the greatest total cross-sectional area and have the slowest flow rate. The dividing and subdividing of vessels within the circulatory system results in greater velocity in the arterial and venous beds than in the capillary bed (Fig. 15.11). An understanding of the concept of velocity enhances discussion of laminar and turbulent flow.

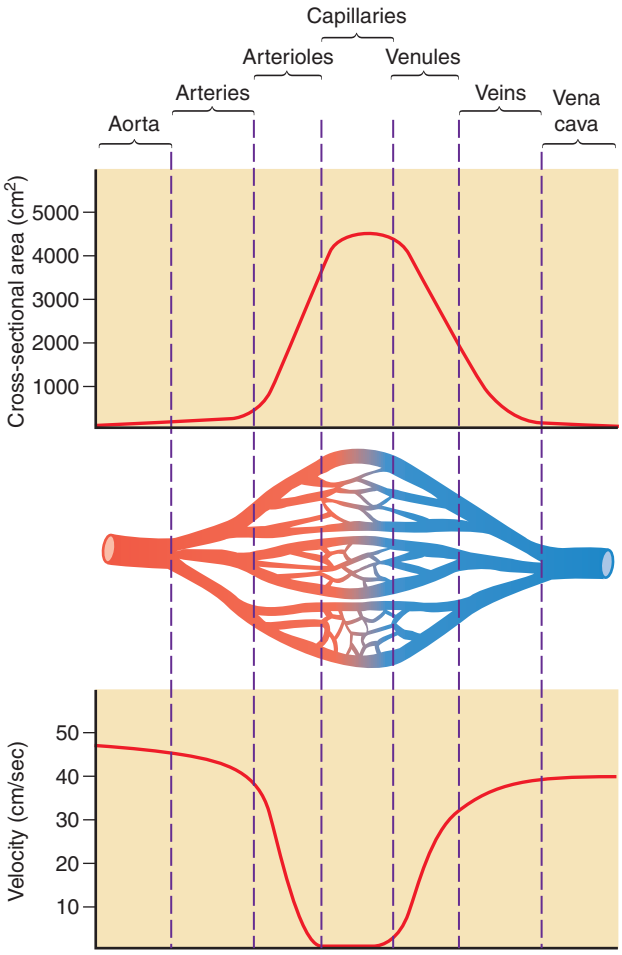


FIG 15.11 Effect of increasing cross-sectional area on the velocity of blood flow. Increased cross-sectional area in the capillary bed results in a significant decrease in velocity compared with the arterial and venous networks.

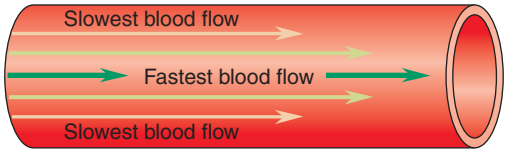


FIG 15.12 Parabolic profile of laminar blood flow.

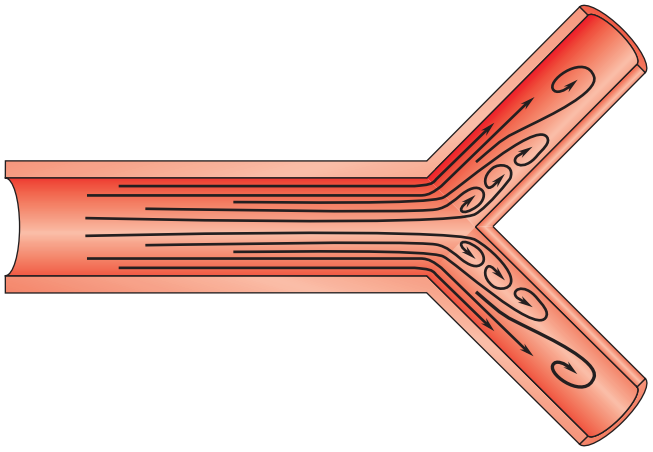


FIG 15.13 Turbulent flow generated at a blood vessel bifurcation.

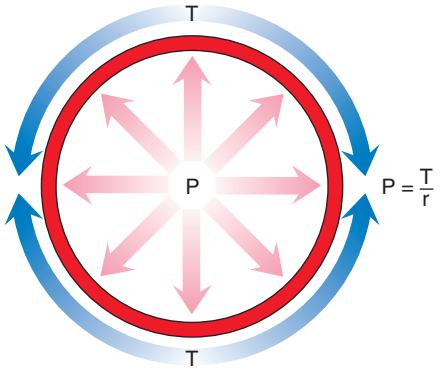


FIG 15.14 Law of Laplace as applied to a blood vessel. Distending pressure *(P)* is the difference between the pressures on either side of the vessel and is equal to the wall tension *(T)* divided by the radius of the blood vessel *(r)*.

When blood flows through a long, smooth-walled vessel, it does so in layers. The velocity of the layers varies, with blood in the center moving much faster than blood in the outer layers. The blood in the center layer moves the most quickly because it is in contact with blood only. The outermost layer is also in contact with the intima of the vessel wall, which exerts friction against the cellular components of the blood. Many blood cells stick to the intima; this layer may flow only minimally. Layers of blood between this outer layer and the central core of blood slide over one another with increasing velocity. This is referred to as the *parabolic profile of laminar flow* and is illustrated in Fig. 15.12.

The streamlined nature of laminar flow is disrupted by normal anatomy and by pathologic processes creating turbulent flow. Turbulent flow is an interruption in the forward current of blood flow by crosswise flow (Fig. 15.13). The propensity for turbulent flow increases with increasing velocity and increased vessel radius, so that some turbulence can be predicted at the aortic root and in the branches of major arteries. The same process can be seen in a river, where boulders interrupting the flow produce whirlpools and the characteristic roar of rapids. In the human body, turbulent flow through blood vessels can be auscultated as a bruit. Sometimes it can be palpated as well, and then it is called a *thrill*. This turbulence may be the result of a normal increase in velocity or be attributable to blood moving through vessels that branch at a sharp angle. Pathologically, turbulence results if blood flows around an obstruction in the vessel or over a roughened intimal surface. Regardless of cause, turbulent flow alters the parabolic profile seen with laminar flow, slowing velocity around the source of the turbulence. This slowing can cause cellular components of the blood to adhere to one another, to the turbulent focus itself, and to the intimal wall, promoting the formation of a blood clot (thrombus).

Wall Tension and Compliance

The relationship between distending pressure and wall tension is expressed by the *law of Laplace* and is illustrated in Fig. 15.14. This physical principle has broad applications in physiology; however, the present discussion focuses on its implications for blood vessels. The distending pressure *(P)* is the transmural pressure, or pressure on one side of the vessel wall minus the pressure on the other side of the blood vessel. It is equal to the wall tension *(T)* divided by the radius of the blood vessel *(r)*. In summary, an increase in radius or distending pressure results in increased wall tension.

When the pressure of the blood in the vessel begins to decline, wall tension forces exceed distending forces, the radius decreases, flow rate declines, and resistance increases. The distending pressure may fall to a point at which it is no longer possible to hold the blood vessel open. If the pressure reaches 20 mm Hg, a point called the *critical closing pressure,* blood flow ceases entirely.

The smaller the radius of the blood vessel, as in a capillary compared with an artery or a vein, the less tension is needed in the wall to equalize the distending pressure. Wall tensions decrease rapidly from 170,000 dynes/cm in the aorta to 16 dynes/cm in the capillaries, rising to 21,000 dynes/cm in the vena cava.

Wall tension is a product of the elasticity of the vessel and is a force that opposes the distending pressure. The manner in which wall tension in a given vessel responds to changes in distending pressures is based on its compliance. Compliance reflects the distensibility of a blood vessel—its ability to accept an increased volume of blood. The large quantity of muscle tissue in much of the arterial system limits its distensibility. Veins, however, are highly distensible and compliant, capable of holding a large quantity of blood at a low pressure. Because of this quality, veins are referred to as *capacitance vessels.* When the body is at rest, 75% of the total blood volume is found in the systemic venous system.

Dynamics in the Microcirculation: Capillaries and Lymphatics

The smallest vessels of the vascular system and the lymphatic vessels are commonly referred to as the *microcirculation.* The primary function of the capillary bed is essentially the essence of the entire circulatory system: the exchange of gases and nutrients. Blood flow in the capillary bed is largely laminar, with minimal turbulence at bifurcations. Within each organ or tissue in the body, capillary blood flow is related to the driving force, which is the difference between arterial and venous pressures, and inversely related to resistance.

The exchange of materials across the capillary endothelium through the interstitial space, either to or from the cells, occurs on an ongoing basis. Substances pass between tissue interstitial fluid and capillary blood by moving along a concentration gradient (diffusion), whereas fluid moves according to a pressure gradient (filtration). As fluid moves through the interstitial space, most of it returns to the capillary bed. Normally, approximately 10% of the fluid remains in the interstitium and is absorbed by the adjacent lymphatic system to be returned to the general circulation. Alteration in the pressure gradient responsible for filtration can allow an excessive amount of fluid to escape into the interstitial space. Increased fluid accumulation in the interstitial space also can occur when the lymphatic flow is impaired or when capillaries become more permeable and "leak" fluid. These are the pathophysiologic mechanisms that result in edema.

The pressure gradient between the capillary and the interstitium is produced and maintained in accord with the balance of four distinct forces or pressures: (1) capillary hydrostatic pressure (P_{cap}), (2) interstitial fluid colloid osmotic pressure (π_{tissue}), (3) plasma colloid osmotic pressure (π_{cap}), and (4) interstitial fluid pressure (P_{tissue}) (Fig. 15.15). This delicate balance of forces is summarized by *Starling's hypothesis,* which states that the net filtration is equal to the combined forces fostering filtration minus the combined forces opposing filtration. Using the normal values shown in Fig. 15.15, each component of the formula and the resulting net filtration pressure are shown:

$$\text{Pressures favoring filtration} = P_{cap} + \pi_{tissue}$$

$$(+17.3\ \text{mm Hg}) + (+8.0\ \text{mm Hg}) = 25.3\ \text{mm Hg}$$

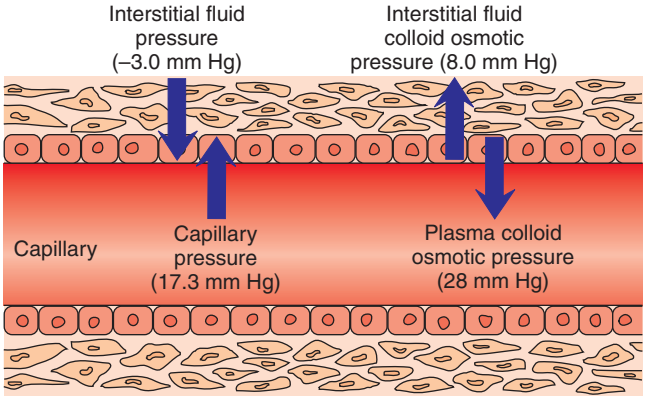


FIG 15.15 Components of the capillary pressure gradient. Filtration reflects the difference between the combined forces that push fluid out of the capillary (capillary pressure and interstitial fluid colloid osmotic pressure) and those that attempt to hold fluid in the capillary (plasma colloid osmotic pressure and interstitial fluid pressure).

$$\text{Pressures opposing filtration} = P_{tissue} + \pi_{cap}$$

$$(-3.0\ \text{mm Hg}) + (28\ \text{mm Hg}) = 25.0\ \text{mm Hg}$$

$$\text{Net filtration pressure} = (P_{cap} + \pi_{tissue}) - (P_{tissue} + \pi_{cap})$$

$$(+25.3\ \text{mm Hg}) - (+25.0\ \text{mm Hg}) = +0.3\ \text{mm Hg}$$

Clinically, capillary fluid pressure and plasma colloid osmotic pressure are the most important concepts to a discussion of pathophysiology. Capillary fluid pressure is the blood pressure in the capillary. It is the force pushing fluid from the capillary into the interstitium and is often called the *hydrostatic pressure.* The strength of this force depends on the blood pressure and the resistance within the arterial and venous systems. Pathologic conditions resulting in an increase in either the blood pressure or the resistance to flow can alter this force, increasing it and propelling more fluid into the interstitial space, resulting in the formation of edema.

Plasma proteins are responsible for the plasma colloid osmotic pressure, the primary force resulting in fluid remaining in the capillary. Most plasma proteins normally remain in the capillaries because they are such large molecules that they cannot move through the capillary spaces. The vast majority of plasma protein, by weight, is albumin. Although globulins and fibrinogen have greater molecular weight, albumin is present in plasma in greater quantity. The number of dissolved molecules in the plasma determines the *plasma colloid osmotic pressure.* The number of dissolved molecules in the interstitial space establishes the interstitial fluid colloid osmotic pressure. Plasma has nearly four times the concentration of proteins than does the interstitium. For that reason, plasma colloid osmotic pressure normally exceeds that in the interstitium, favoring fluids remaining in the capillaries.

As previously illustrated, the net filtration pressure in a typical capillary is 0.3 mm Hg. This pressure difference is responsible for producing the fluid excess in the interstitial space, which is then normally absorbed by the lymphatic system for eventual return to the systemic vascular circulation. If the pressures are altered, an even greater pressure gradient may be produced, and more fluid moves from the capillaries into the interstitial space. Likewise, a change in the permeability *(K)* of the capillary wall that allows plasma proteins to leak out or a reduction in lymphatic flow will allow fluid to collect in the interstitium. In each case, the result is edema, which can occur with many pathologic conditions. When the pathology is an impairment of lymphatic flow allowing

fluid to collect in the interstitium, it is more specifically termed *lymphedema*.

Once absorbed into the lymphatic system, interstitial fluid is referred to as *lymph*. It is similar in composition to interstitial fluid but has a lower concentration of protein. Molecules of fat and bacteria are also found in lymph. Lymph circulates throughout the body at a rate of approximately 3 L/day. Lymphatic flow can be increased in several ways: by increasing the capillary pressure, decreasing the plasma colloid osmotic pressure, increasing the interstitial fluid colloid osmotic pressure, or increasing the permeability of the capillaries. The interstitial fluid hydrostatic pressure increases (becomes less negative) when any of these factors changes, producing an increase in lymphatic flow.

KEY POINTS

- Physical laws govern the flow of blood through the circulatory system. Predictions regarding blood flow, blood pressure, and resistance to flow can be made using these laws. The important relationships may be summarized as follows:
  - Flow = pressure/resistance
  - Blood pressure = flow (cardiac output) × resistance
  - Resistance = pressure/flow
- The main factors affecting resistance to flow are the radius and length of the vessels and blood viscosity and turbulence. Usually, the radius of the vessel is the most important determinant of resistance. It affects resistance inversely and to the fourth power. A small decrease in radius results in a large increase in resistance.
- The velocity of blood flow varies inversely with the total cross-sectional area of the vascular bed. The capillaries have the greatest total cross-sectional area and, therefore, the slowest flow.
- Laplace's law describes the relationships among wall tension, distending pressure, and vessel radius *(P = T/r; T = Pr)*. An increase in radius or distending pressure results in increased wall tension. At critical closing pressure, wall tension overwhelms distending pressure and blood flow ceases.
- The transcapillary exchange of fluid and nutrients is accomplished by the processes of diffusion and filtration. Diffusion refers to movement of solute and is determined by capillary permeability and the size of the concentration gradient. Filtration refers to movement of fluid and is affected in the following ways:
  - Increased capillary fluid pressure and interstitial fluid colloid osmotic pressure enhance filtration.
  - Increased interstitial fluid pressure and plasma colloid osmotic pressure oppose filtration.
  - Increased permeability *(K)* enhances filtration.

CONTROL OF FLOW

Blood flow throughout the periphery is controlled by central mechanisms that are mediated by the autonomic nervous system, the venous and thoracic pumps, and intrinsic autoregulatory mechanisms. Lymphatic flow is controlled by increasing interstitial fluid colloid osmotic pressure and by the stimulation of the contractile fibers (often called *lymphatic pumps*) as they are stretched. In healthy people, these mechanisms of control respond to changes in the internal and external environments and compensate rapidly and efficiently; however, during states of illness these mechanisms may be inadequate to compensate for alterations in flow.

Control of Blood Flow

Extrinsic Mechanisms

The autonomic nervous system provides the primary extrinsic control of blood flow through the sympathetic nervous system (SNS). Although parasympathetic nervous system innervation is important to the regulation of the heart, it is not important to the regulation of peripheral resistance. Within the medulla, groups of neurons form the vasomotor center. This area plays a major role in the maintenance of blood pressure (see Chapter 16). The vasomotor center responds to direct stimulation and to afferent stimuli of both an excitatory and an inhibitory nature. A basal rate of discharge from the vasomotor center results in a continuous minimal level of contraction of vascular smooth muscle, referred to as *vasomotor tone*.

All blood vessels except the small venules and capillaries contain smooth muscle that is innervated by adrenergic fibers from the SNS. Because arteries have the most smooth muscle, they are most affected by SNS stimulation. Veins, by contrast, have little neural innervation, and venoconstriction has a minor role in controlling blood flow except in the skin and the splanchnic circulatory bed of the gut. In general, the release of norepinephrine, the SNS postganglionic neurotransmitter, results in arterial vasoconstriction via the α_1 receptors located on the vascular smooth muscle walls. Likewise, drugs that mimic the α_1-receptor response (α_1 agonists) produce vasoconstriction, increasing vasomotor tone and diastolic blood pressure. Administration of an α_1 antagonist results in the blockade of these receptors and results in vasodilation of the arterial bed, reducing blood pressure.

Although the β_2-adrenergic receptors located on blood vessels in skeletal muscle produce vasodilation when stimulated, they are only minimally affected by endogenous norepinephrine from the SNS. Epinephrine, the endogenous catecholamine released by the adrenal medulla, or its exogenous pharmacologic equivalent (adrenalin), stimulates these receptors, producing vasodilation. Therefore their major role is not so much to maintain vasomotor tone but to increase nutrient and oxygen supplies to skeletal muscles during periods of increased demand.

Blood flow through the venous system into the right side of the heart is maintained by the pressure gradient from the veins and by the venous and thoracic pumps. Blood is propelled through the circuit, pushed by the force of left ventricular contraction, and moves forward toward the low-pressure side of the pump on the right side of the heart. In the peripheral veins, what is known as the *venous pump* is activated by skeletal muscle activity. Folds in the intimal wall of the veins create valves. Contraction of the skeletal muscles bordering the veins compresses them, forcing the valves open and propelling venous blood back toward the heart. Instigation of the venous pump significantly facilitates venous return. Patients who are immobilized by bed rest lose this valuable mechanism, which results in a decrease in cardiac preload to the right heart and increased work of the heart to maintain the cardiac output. The *thoracic pump* acts to increase venous return to the heart as intrathoracic pressure changes with breathing. Inspiration increases negative intrathoracic pressure, resulting in more venous return (see Chapter 17).

Intrinsic Mechanisms

Autoregulation refers to the ability of blood vessels within organs to maintain a relatively constant blood flow, regardless of changes in arterial pressure. This flow is *relatively* constant because it does have limits; there is a range within which it is maintained, and the range varies slightly from organ to organ. Several processes contribute to the autoregulation of perfusion to meet the needs of individual organs within the body.

It is known that as vascular smooth muscle is stretched, it contracts. Therefore as arterial pressure rises and arterial walls stretch, contraction is stimulated, producing vasoconstriction. Resistance to flow is also increased with stretch by early closing of precapillary sphincters. This process certainly may contribute to autoregulation, but it is not the primary mechanism.

Over the past several decades, a great deal has been learned about the endothelium of blood vessels. The previous perceptions of the endothelium, as an inactive structure whose function was no more than acting as a barrier between the blood and the more functional layers of the vessel wall, have been proven incorrect. We now understand that the endothelium is a major participant in vascular tone and growth of vascular smooth muscle. The endothelium tissue plays an active role in the immune and inflammatory processes (see Chapter 9), platelet activity in normal coagulation and thrombus formation (see Chapter 14), and arteriosclerosis, discussed later in this chapter. In a discussion of autoregulation, the key is the endothelial role in modulating vascular smooth muscle to produce vasoconstriction or vasodilation. The endothelium is capable of sensing alterations of a chemical or physical nature within the vessel and responding to these stimuli directly or through the release of signals that initiate change. Almost all of the vasodilation occurs because of the effect of nitric oxide (NO). NO is a gas present in most body tissues and is produced by the cells of the vascular endothelium. As a gas, it diffuses from the endothelium to the smooth muscle cells, binding to intracellular receptors to affect cytoplasmic Ca^{++} concentration and produce vasodilation. Deficits of NO, or a decreased responsiveness, have been the focus of considerable recent research into the pathogenesis of hypertension (Chapter 16) and heart disease (Chapter 18). Other relaxing factors produced by the endothelium include prostacyclin and endothelium-derived hyperpolarizing factor. Angiotensin II (AII), endothelin, oxygen-derived free radicals, prostacyclin H_2, and thromboxane A_2 are among the constricting factors. The role of AII as a constricting factor has been the focus of considerable recent research. Because drugs that block the effects of AII are available and in widespread use, the ability of these drugs to improve peripheral vascular blood flow has been an important finding. In addition to substances produced by the endothelium itself, metabolic by-products (metabolites) or substrates have been found to exert a direct effect, altering blood flow to the area. Metabolites might include carbon dioxide or lactic acid. Histamine and prostaglandins are examples of metabolic substrates. Other substances, such as acetylcholine, bradykinin, histamine, and substance P, exert their effect by increasing the formation of NO or are themselves generated by NO. These various chemicals create a balance of forces in health. They may also be affected by aging, disease, or pharmacologic interventions.

A local increase in blood flow is referred to as *hyperemia*. The increase in local blood flow in response to increased metabolic demand is called *active* or *functional* hyperemia. *Reactive* hyperemia occurs when a temporary reduction in blood flow is reversed. The body responds by briefly increasing circulation to the area, resulting in the characteristic flushing seen, for instance, when a tourniquet is removed. The *tissue pressure hypothesis* of autoregulation postulates that an acute increase in the pressure within the arterial system causes an increase in interstitial volume and pressure. This increased tissue pressure, external to the vasculature, results in compression of small vessels, which increases resistance and reduces flow.

Control of Lymphatic Flow

The movement of lymph is expedited by lymphatic pumps. This is a general concept that encompasses the pumping action of the lymphatics themselves and the pumping effect on the lymphatic vessels produced by activity external to them. Like veins, lymphatic vessels have valves on their intimal surface that allow forward movement of fluid to join the venous return to the systemic circulation. Compression of lymphatic channels by adjacent skeletal muscles, the smooth muscle of organs, and the pulsatile movement of arteries force lymph forward. Intrathoracic pressure changes related to breathing increase lymphatic return as well as venous return. Lymphatic flow is therefore enhanced by increased physical activity, increased blood pressure, or increased respiratory rate. Lymphatic contractions are thought to be the primary factor in lymphatic flow. Lymph is propelled forward when lymphatic capillaries contract in response to being stretched. The rate of contractions increases as the volume of lymph increases.

KEY POINTS

- The blood flow through a particular vascular bed is regulated centrally by the autonomic nervous system and locally by the organ or tissue.
- In most vascular beds, the sympathetic nervous system (SNS) causes constriction, which increases resistance and reduces flow. Smooth muscle cells in these vascular beds have $\alpha 1$ receptors that bind the SNS neurotransmitter norepinephrine, causing contraction. There is no significant parasympathetic innervation of systemic vessels.
- Autoregulation refers to a tissue's ability to regulate its own flow. Autoregulation allows a tissue to maintain optimal flow despite changes in blood pressure or metabolic demands. In instances of high blood pressure or decreased metabolic demand, the arterioles and precapillary sphincters that control flow to the tissue constrict, reducing flow. In instances of low blood pressure or high demand, vessels dilate, increasing flow.
- Lymphatic vessels maintain flow by contracting when stretched with lymph. Intraluminal valves prevent backflow. External compression by contracting muscles enhances lymph flow.

GENERAL MECHANISMS THAT CAUSE ALTERED FLOW

A reduction in flow through the systemic vasculature results in the impaired ability to transport gases and nutrients to and from body tissues. Cells of the body vary in their oxygen demands. *Hypoxia,* an insufficient supply of oxygen, can occur for many reasons, such as a decrease in hemoglobin formation (see Chapter 13) or diminished oxygen transport in the lungs (see Chapter 21). When hypoxia is a result of a decrease in flow through the arterial system, it is called *ischemia.* Impairment in flow through the venous system interferes with the removal of metabolic waste products and causes fluid pressure to accumulate in the system, a condition known as *venous engorgement* or *venous obstruction.* When the lymphatic circulation is altered, the resulting fluid and pressure changes may be visible locally or systemically.

Blood Vessels: Obstructions

Pathologic processes affecting blood flow may involve impedance of the arterial or venous system. Some obstructions to flow are specific to either the arterial or the venous portion of the system, but most can occur in some form in both. Obstructions to flow that may interfere with arterial or venous flow are presented in detail in the following discussion. Those that are specifically related to one or the other are detailed more fully later in the chapter.

Thrombus

A thrombus is a stationary blood clot formed within a vessel or a chamber of the heart. Thrombosis is initiated by a change in the blood vessel resulting in localized stasis of flow. Inflammation of blood vessels may be the stimulus for thrombosis in either arteries or veins.

Etiology. Thrombosis refers to the pathologic formation of clots at these sites, to differentiate it from the clotting process that takes place as a homeostatic mechanism. Thrombi (blood clots) are composed of aggregated platelets, clotting factors, and fibrin that adhere to vessel

walls. Thrombi may form in the chambers of the heart in association with certain abnormal heart rhythms (see Chapter 19), following a myocardial infarction, or as a result of damage to heart valves or replacement of heart valves with artificial ones (see Chapter 18). More commonly, thrombi develop in either the arterial or the venous peripheral circulatory systems. Activation of the coagulation cascade within the vessel produces a hypercoagulable state resulting in thrombosis (see Chapter 14). Certain drugs, such as oral contraceptives, increase the tendency to form thrombi as well. Thrombosis is also more likely to occur when blood flow slows dramatically or becomes more turbulent, or if there is damage to intimal walls, creating a roughened surface.

Pathogenesis

Arterial. The significance of thrombosis rests in the ability of a clot within a blood vessel to reduce flow and increase turbulence, which enhances thrombus enlargement and the formation of more thrombi. The results of reduced blood flow vary depending on whether the arterial or venous system is involved. If the thrombus forms in the arterial system, decreased distal flow can result in ischemia. This is significant in several pathologic conditions, such as acute arterial occlusion (discussed in this chapter). Other examples of arterial thrombosis are explored elsewhere in this text (e.g., myocardial infarction, Chapter 18; stroke, Chapter 44).

Venous. In the venous system, thrombosis alters venous return, impairing removal of metabolic wastes and producing swelling (edema). When inflammation occurs in a vein (phlebitis) and is accompanied by the formation of a thrombus, it is called *thrombophlebitis*. The most common cause of thrombophlebitis is the inflammation produced by the presence of a needle or catheter used for intravenous therapy. Thrombosis may also be initiated by a generalized reduction in flow and the accompanying release of vasoactive substances that occur in shock states (see Chapter 20). Systemic derangement in coagulation takes place in disseminated intravascular coagulation, resulting in thrombosis in the microcirculation throughout the body (see Chapter 14). Risk factors associated with both arterial and venous thrombosis are listed in Box 15.1.

BOX 15.1 Risk Factors Commonly Associated With Thrombosis

General (Arterial and Venous)

Hypercoagulable conditions
- Polycythemia
- Dehydration
- Platelet aggregation

Pump failure
- Heart failure
- Shock

Dysrhythmias

Aging

Trauma, including surgery

Drugs
- Anesthetic agents
- Oral contraceptives
- Tobacco

Arterial

Arteriosclerosis/atherosclerosis

Venous

Immobilization/sedentary lifestyle

Clinical manifestations and treatment.

Arterial thrombosis is usually manifested by intermittent claudication (pain with activity) in the affected limb that improves with rest. The limb might also be cool to the touch and cyanotic. A late sign is a painful arterial ulcer found usually around one toe.

Symptoms for venous thrombosis may be absent or may be life threatening secondary to pulmonary embolism (PE). Other signs include calf or groin tenderness and swelling of the affected limb with associated increased skin temperature. Pain in the calf with dorsiflexion of the foot (Homan sign) appears in 10% of those with thrombophlebitis.

Interventions in the management of thrombus formation may be medical or surgical. Ideally, thrombosis is prevented in high-risk individuals through pharmacologic and other medical interventions. The prophylactic (preventive) interventions may include oral or parenteral anticoagulant therapy, or drugs to block platelet activation and/or aggregation. Risk factors must be addressed. Once a thrombus has formed, anticoagulant therapy at a therapeutic level is initiated to prevent the enlargement of the thrombus and formation of further thrombi. However, these drugs are not effective in dissolving an existing clot. These clots must be removed by the body's own fibrinolytic process, surgically removed, or dissolved by intravenous thrombolytic agents. Anticoagulant therapy is currently used for patients with thrombi in coronary and pulmonary arteries, peripheral arteries in the legs, and cerebral arteries. Their use must be closely supervised; patients receiving thrombolytic therapy are usually in critical care settings. Additional medical prophylactic interventions may include the use of antiembolic stockings or sequential compression devices for immobilized patients and initiation of ambulation as soon as possible. Surgical interventions, such as removal of the thrombus, may be an option.

Because thrombi partially or completely occlude flow through the involved vessel, they can produce ischemia distal to that point in an artery or congestion proximally in a vein. A thrombus that only partially occludes a vessel continues to be affected by the force of blood flow. Eventually, it may break free from the vessel wall and become an embolus.

Embolus

An embolus is a collection of material that forms a clot within the bloodstream. This traveling clot is propelled forward in the circulatory system by blood flow to a distant point, where it lodges to produce a new site of obstruction.

Etiology and pathogenesis.

An embolus is most often a blood clot, a thromboembolus, having begun as a thrombus that was subsequently dislodged from the vessel intima or from the valvular leaflets in the heart, or having formed within a chamber of the heart. Thromboemboli from the left side of the heart exit the aorta and most commonly lodge in a cerebral artery, resulting in a *stroke* (see Chapter 44). But most thromboemboli originate in the deep veins of the pelvis and lower extremities. They traverse the venous circulation and return to the right side of the heart, eventually lodging in the arterial side of the pulmonary vasculature and resulting in a PE (see Chapter 21). A thromboembolus from the right side of the heart will also result in a PE. Thromboemboli from the venous circulation are the most common cause of pulmonary emboli, but the cause may be nonthrombotic, as is the case for tumor, fat, air, amniotic fluid, or bacterial emboli; these are less frequent and are further examined later in this discussion.

Clinical manifestations.

An embolism exiting the left ventricle may lodge in the cerebral vasculature, leading to an ischemic stroke. Manifestations differ depending on the area of the brain affected. Symptoms include loss of cognitive function, motor changes, and different levels of sensory loss (see Chapter 44). More often, an embolism leaving the right ventricle lodges in the pulmonary vasculature. This PE may be asymptomatic or present with various signs and symptoms, many of

which are vague and nonspecific. Most common is the sudden onset of shortness of breath (dyspnea), increased respiratory rate, and chest pain. It may be a cause of sudden death.

Treatment. *Embolectomy,* the surgical removal of an embolus, is usually confined to thromboemboli. The use of this surgical technique is contingent on the location of the embolus. In patients who experience repeated emboli, usually originating from the peripheral venous system, a filter (e.g., Greenfield filter) may be surgically implanted in the inferior vena cava. As the blood passes through the filter, emboli are trapped and cannot progress into the pulmonary circulation. The body's own thrombolytic enzyme, plasmin, then destroys the trapped emboli.

Emboli produced by other causes. Various other materials, some totally foreign to the bloodstream, can also form emboli if present in sufficient quantity. Fat emboli are aggregates of fat molecules released into the blood after trauma or surgery involving bone. Most frequently the long bones of the legs are the source of these emboli. Increased pressure generated within the traumatized bone by the inflammatory response forces molecules of fat from the interior of the bone into the bloodstream. Malignant neoplasms can metastasize by various means, one of which is via the blood as tumor emboli. Collections of bacteria and infectious exudate may break free from a source within the circulation, such as the leaflets of the valves of the heart in bacterial endocarditis. Once in the bloodstream, the bacterial emboli continue to travel, eventually occluding circulation and becoming a new site of infection. Air from the external environment is a foreign material when found in the bloodstream as air emboli. Bubbles of air, having most likely entered the blood through an intravenous catheter, come to rest in small blood vessels and obstruct perfusion. It is difficult to identify the specific volume of air that can sufficiently obstruct flow to result in deleterious effects in humans. In animal studies, the quantity of air needed to produce death varies, partially affected by the speed with which it is injected. Under some circumstances, a 5-mL injection of air will result in death of animal models. At other times, a 100-mL bolus of air will not produce adverse effects.

Increased pressure in the abdomen generated during labor and delivery may force amniotic fluid into the bloodstream as emboli. Here the emboli cause a different set of problems. Amniotic fluid cannot perform the functions of the blood in carrying gases and nutrients, but as a fluid, it does not produce obstruction to flow. Instead, the proteins and cells in amniotic fluid act as antigens, initiating an immune response. There are no clinical signs or symptoms that are specific for PE. This nonspecific presentation is magnified during pregnancy due to an overlap between symptoms seen in patients with PE and those associated with the normal physiologic changes of pregnancy (e.g., dyspnea occurs in up to 70% of normal pregnancies).

Vasospasm

Vasospasm is a sudden constriction of arterial smooth muscle that results in an obstruction to flow. In some cases, vasospasm is sufficient to produce hypoxia distally, as in variant (Prinzmetal) angina (see Chapter 18) or vasospasm of cerebral vessels after a hemorrhagic stroke (see Chapter 44). Frequently, the cause of vasospasm is unknown. Certain individuals may be unusually sensitive to hormonal changes or food additives, which may result in vasospasm of cerebral arteries. The vasodilation after cerebral vasospasm is thought to contribute to migraine headaches. Vasospasm may also be mediated by environmental factors, such as exposure to cold or emotional stress, producing a localized response.

Inflammation

Vasculitis is inflammation of the intima of an artery. Inflammation of the lining of a vein is called *phlebitis*. If superficial, these inflammations may be visible as reddened, tender streaks on the skin. Of more significance is their potential to serve as foci for the thrombotic process.

Arteritis *(angiitis)* is a specific term that identifies an inflammatory process of autoimmune origin in arteries. The initiating stimulus is frequently an infectious process that is viral or bacterial (especially streptococcal), or an adverse response to drugs such as sulfonamides or phenothiazines.

Mechanical Compression

A variety of forces external to the vascular system may result in partial or complete obstruction of blood flow. Trauma may produce direct pressure on a blood vessel, resulting in occlusion. This same effect may result from constriction from casts or tight dressings. Swelling secondary to bleeding or edema within a fascial compartment created by fascial tissue surrounding groups of muscle, or external compression of the compartment by a tight cast, eventually compromises the circulation distally, producing *compartment syndrome* (see Chapter 51). Prolonged occlusion produces neurovascular alterations that can be assessed before the ischemia is irreversible. These alterations are identical to those of acute arterial occlusion, discussed later in this chapter. In an untreated patient, compartment syndrome can result in prolonged hypoxia, ischemia, and necrosis of tissues.

Blood Vessels: Structural Alterations

An assortment of conditions affecting blood vessel structure will produce alterations in blood flow. The structure of arteries or veins may be changed secondary to congenital anomalies or pathologic processes triggered later in life.

Types of Structural Alterations

Valvular incompetence. The intimal folds of veins that form the valves can be damaged, interfering with the effective flow of blood through a portion of the venous system (valvular incompetence). The subsequent pathologic processes may affect superficial veins (varicose veins) or deep veins (chronic venous insufficiency), resulting in severe tissue hypoxia and venous stasis ulcers.

Arteriosclerosis/atherosclerosis. *Arteriosclerosis* is a general term for the thickening and hardening of arteries. Atherosclerosis is a type of arteriosclerosis. Atherosclerosis, or hardening of the arteries, is a condition in which plaque builds up inside the arteries. Plaque is made of cholesterol, fatty substances, cellular waste products, calcium, and fibrin (a clotting material in the blood). Atherosclerotic changes are responsible for or contribute to many diseases throughout the body such as hypertension, renal failure, coronary artery disease (CAD), and cerebrovascular disease.

Aneurysms. An aneurysm is a localized dilation of an arterial wall. Aneurysms vary in the severity of their consequences, depending on their size, type, and location. All aneurysms produce an alteration in flow attributable to the changes in vessel diameter. More significant, however, is the fact that the aneurysm represents a weakened area in the artery that may eventually rupture.

Arteriovenous fistulas. An arteriovenous fistula (AVF) is an abnormal communication between arteries and veins. It is usually congenital in origin but may result from traumatic injury. Symptoms depend on the size and location of the fistula. Because AVFs provide a shortcut between the two vascular systems, they can result in alterations in oxygenation to the involved tissues and systemic hemodynamic changes. One of the most common and serious types of AVFs is an arteriovenous malformation (AVM). An AVM is a tangled knot of arteries and veins found most commonly within the brain vasculature. AVMs may be the underlying cause of such conditions as headaches, hemorrhagic stroke, dementia, or seizures (see Chapters 44 and 45).

Lymphatic Vessels

The lymphatic collection system may be overwhelmed when changes in capillary or interstitial oncotic pressures increase filtration into tissues. The result is edema, the collection of an excessive amount of fluid in the interstitial spaces. A wide variety of conditions can result in edema.

When lymphatic flow is altered because of impairment in the circulation of lymph itself, the condition is called *lymphedema*. The result is also an excessive quantity of fluid in the interstitium, but the underlying cause is an obstruction to flow.

KEY POINTS

- Altered blood flow results from obstructive processes. Obstruction results in reduced flow beyond the obstruction (downstream) and increased pressure before the obstruction (upstream).
- In the arterial system, obstruction manifests primarily as distal ischemia. In the venous system, obstruction manifests as edema.
- The causes of vessel obstruction include thrombi, emboli, vasospasm, external compression (e.g., compartment syndrome), and structural alterations (e.g., atherosclerotic plaques, aneurysms).
- Alterations in pressures within the circulatory system or interstitium produce edema, whereas an impairment of the lymphatic system results in lymphedema.

ALTERATIONS IN ARTERIAL FLOW

Alterations in arterial flow result from obstruction (arteriosclerosis/atherosclerosis, inflammation, vasospasm, thrombi, emboli, and acute occlusion) or mechanical alterations (AVFs and aneurysms).

Arteriosclerosis/Atherosclerosis

Etiology and Pathogenesis

Arteriosclerosis is a generic term meaning "hardening of the arteries" and broadly includes three pathologic processes: Mönckeberg sclerosis (medial calcific sclerosis), arteriolar sclerosis, and atherosclerosis. Mönckeberg sclerosis is a noninflammatory, degenerative disorder in which the media of small- and medium-size arteries becomes calcified. The disease is a risk factor for cardiovascular disease, but the pathology is independent of atherosclerosis. The intimal layer is not a part of the pathogenesis; although the vessel becomes increasingly thickened and rigid, it remains patent because of the changes in the medial layer. Arteriolar sclerosis is characterized by thickening and luminal narrowing of the small arteries that occurs in association with hypertension. However, because hypertension is primarily associated with atherosclerotic changes, this particular pathology is rarely addressed. Atherosclerosis, the most common arteriosclerotic process, affects intermediate-size and large arteries. Smooth muscle cells and lipids collect along the intimal surface, producing a narrowing of the luminal diameter and a reduction in flow. A clot may form and obstruct the lumen (Fig. 15.16).

Atherosclerosis is the dominant type of arteriosclerosis. The word is derived from two Greek words: *athero* (gruel or paste) and *sclerosis* (hardness). It is the pathologic origin for the vast majority of arterial disease that is ultimately the leading cause of death in the United States and western Europe, and is increasing in developing countries. Atherosclerosis tends to develop in large- and medium-size arteries, most frequently the coronary, cerebral, carotid, and femoral arteries and the aorta. Most of the mortality associated with atherosclerosis is the result of occlusion of coronary arteries (CAD), producing myocardial ischemia and infarction. The remainder of atherosclerosis-related deaths is secondary to thrombotic or hemorrhagic processes, primarily in the brain (stroke) and extremities, although other organ systems, including the kidneys, liver, and gastrointestinal tract, are also affected. When atherosclerosis involves the peripheral vascular system, it is most often the lower extremities, and the disease process may be called *atherosclerosis obliterans*. Box 15.2 lists the manifestations of this arterial peripheral vascular disease. Peripheral arterial disease (PAD) is a sign of systemic atherosclerosis. It affects 10% to 15% of the general population, and approximately 50% of PAD patients are asymptomatic, leading to underdiagnosis and undertreatment of the disease. Symptomatic PAD patients have a worse prognosis than patients presenting with CAD or cerebrovascular disease; unfortunately, their atherosclerosis risk factors may be less intensively treated. Both asymptomatic and symptomatic PAD patients have a high risk of death from cardiovascular disease (CVD); therefore early treatment reduces mortality. Patients with chronic peripheral arterial occlusion may complain of pain with activity (*intermittent claudication;* Fig. 15.17) and also have pain at rest with advanced disease. Peripheral pulses are diminished. Ulceration may occur (Fig. 15.18). Research

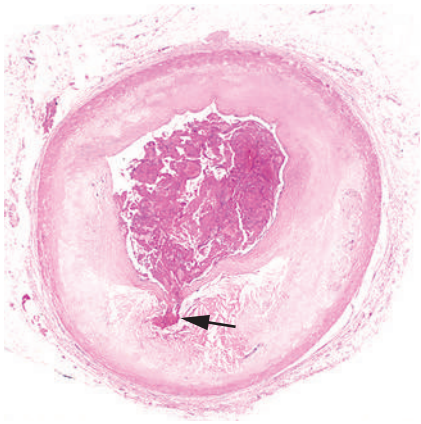


FIG 15.16 Acute coronary thrombosis superimposed on an atherosclerotic plaque with focal disruption of the fibrous cap, triggering a fatal myocardial infarction. An arrow points to the site of plaque rupture. (Reproduced from Schoen FJ: *Interventional and surgical cardiovascular pathology: clinical correlations and basic principles,* Philadelphia, 1989, Saunders, p 61.)

BOX 15.2 Clinical Manifestations of Arterial Peripheral Vascular Disease

Skin Assessment

- Cool or cold to touch
- Decreased or absent hair growth
- Dry, thin, glossy appearance
- Thickened nails
- Pallor when elevated, rubor when dependent
- Diminished or absent pulses

Pain Assessment

- Sharp and stabbing
- Intensified with activity
- Relieved by rest or dependency

Ulcer Assessment

- Severely painful
- Pale, gray base
- Well-defined edges
- Located on heels, lateral malleolus, between distal portions of phalanges, pretibial area

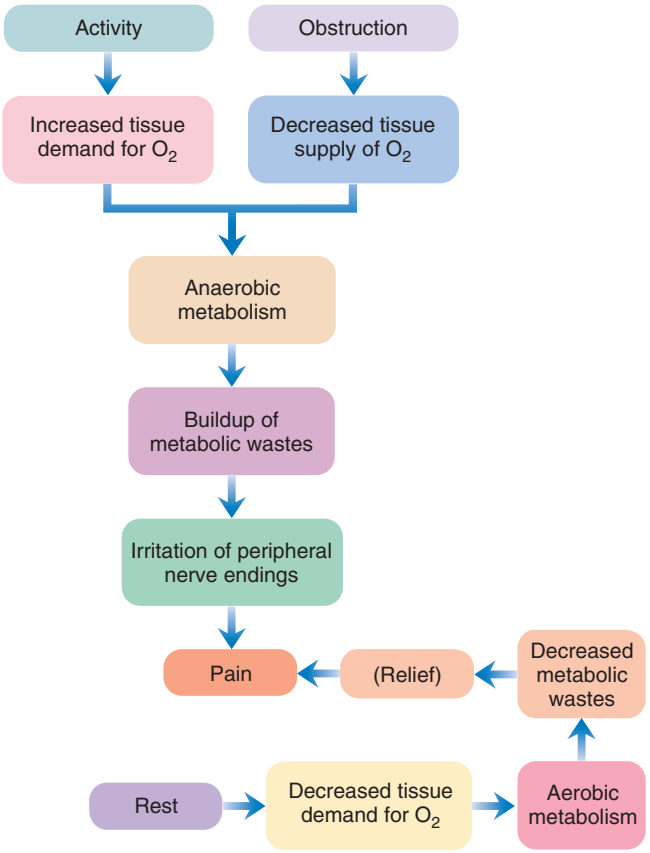


FIG 15.17 Pathophysiologic process of intermittent claudication and its relief.

FIG 15.18 Arterial ulcer. (From Mann DL et al: *Braunwald's heart disease: a textbook of cardiovascular medicine,* ed 10, Philadelphia, 2015, Saunders, p 1316.)

results from the classic Framingham Study identified the development of lower extremity arterial disease in 5% of the subjects over the first 24-year data collection interval.

Because of the breadth of diseases known to be associated with atherosclerosis, a great deal of research has been accomplished and an understanding of many of the significant aspects of its pathogenesis has evolved. These are summarized in Fig. 15.19. The process is initiated by damage to the endothelial surface of the arterial intima, initiating an inflammatory response and an increase in the vessel wall permeability. Many of the risk factors for atherosclerosis discussed later may be initiators of this vessel injury. The increased permeability of the vessel wall allows low-density serum lipoproteins to breach the intimal layer. Leukocytes also are drawn to the site, and along with the endothelial cells, they oxidize the lipids, producing further damage to the vessel wall. Simultaneously, platelets aggregate at the site of injury. They are activated, releasing platelet-derived growth factor, which stimulates growth of smooth muscle cells. Media smooth muscle cells, normally confined to the other tunicae, are drawn to the intima where they proliferate. The result is an atherosclerotic *plaque,* primarily composed of smooth muscle cells, lipoproteins, and inflammatory debris. While the plaques slowly enlarge, the orifice of the artery is decreased, and perfusion is diminished. In CAD, the plaque may also acutely rupture, initiating thrombus formation and acute loss of perfusion (see Chapter 18).

BOX 15.3 Risk Factors Associated With Atherosclerosis

Modifiable Risk Factors

- Smoking
- Elevated blood pressure
- Glucose intolerance
- Elevated cholesterol and low-density lipoproteins
- Decreased physical activity
- Obesity
- Weight fluctuations
- Ineffective stress management

Nonmodifiable Risk Factors

- Age
- Gender
- Ethnicity
- Heredity

Risk Factors

Risk factors for the development of atherosclerosis are categorized as modifiable or nonmodifiable, according to the degree to which they can be altered (Box 15.3). Historically, health care has focused on preventing atherosclerosis by the manipulation of predisposing modifiable factors. It often is difficult to isolate the effect of a single risk factor because they usually occur in combination.

The most frequently cited prospective research into atherosclerotic risk factors began in 1948 in Framingham, Massachusetts. Initially, 5209 men and women between the ages of 30 and 59 volunteered to be subjects in the study, the purpose of which was to identify factors associated with the development of atherosclerosis over time. The Framingham Study remains ongoing, with researchers now studying the children and grandchildren of the original participants. Much of the available information regarding atherosclerotic risk factors has its origins in the results of this research.

Modifiable risk factors. Fortunately, there are far more modifiable risk factors than nonmodifiable ones, and the changes individuals make have documented effects in risk reduction for CVD. Unfortunately, lifestyle changes are often difficult to make and to maintain.

Tobacco use in any form is atherogenic, but most of the research addresses cigarette smoking. Cigarette smoke contains more than 4000

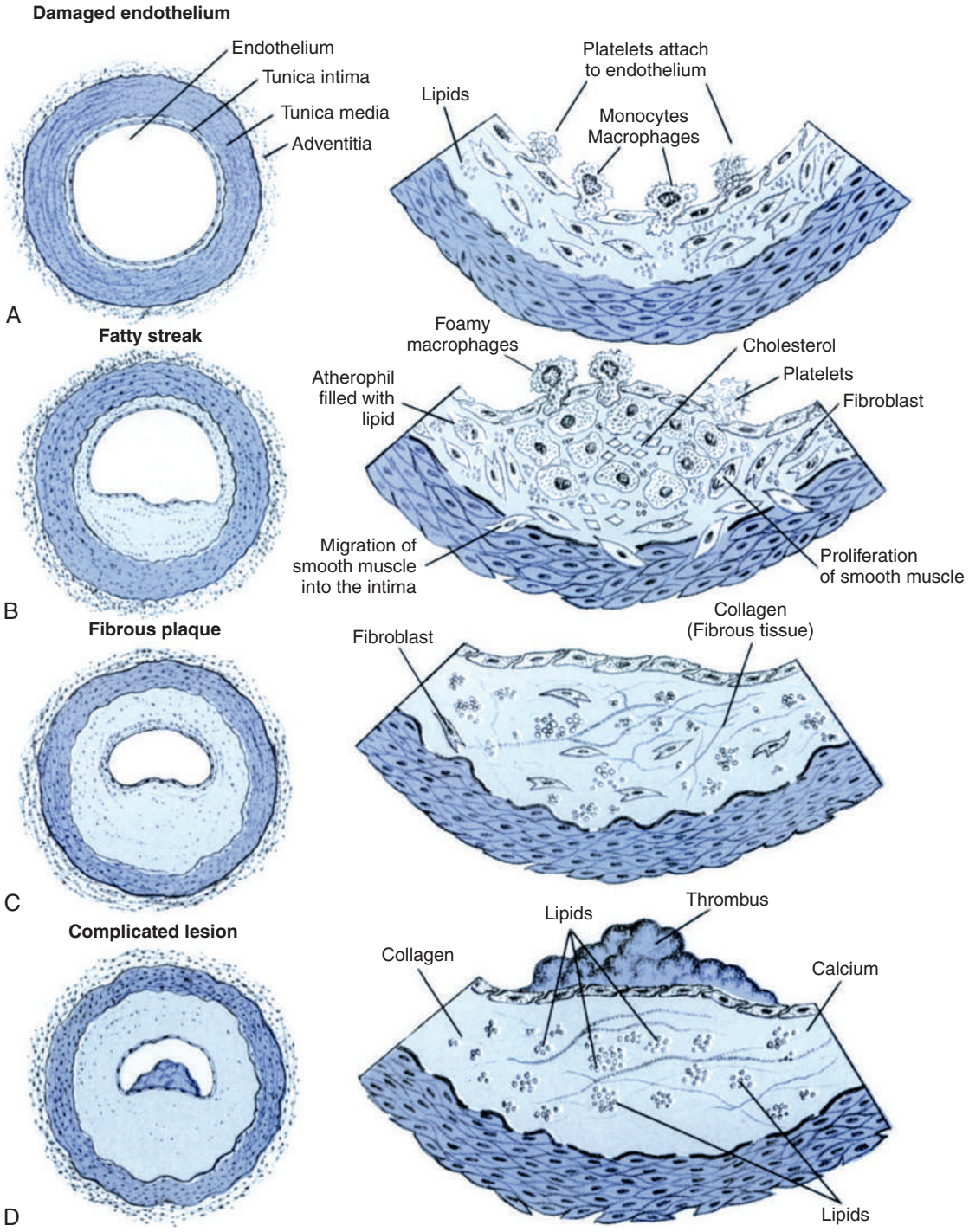


FIG 15.19 Pathogenesis of atherosclerosis. **A,** In response to trauma or irritation to the intima, injury stimulates platelet aggregation and the inflammatory response. **B,** Medial smooth muscle proliferates and migrates into the intima; LDL cholesterol leaks into the vessel wall. **C,** A fibrous cap forms over the plaque; the lesion slowly grows to decrease vessel diameter. **D,** Additional injury may cause rupture of the plaque, resulting in thrombus formation and manifestations of acute occlusion. (From Monahan FD et al: *Phipps' medical-surgical nursing: health and illness perspectives,* ed 8, Philadelphia, 2007, Mosby, p 750.)

elements; nicotine and carbon monoxide are the two with the most documented damage to blood vessels. Exposure to both active and passive smoke has been found to cause injury, although the precise mechanisms have not yet been elucidated. What is known is that cigarette smoking produces injury to the endothelium, generates superoxide anions, decreases both the production and the bioavailability of NO, and increases the production and release of endothelin. The result is dysfunction of the endothelium, increased tendency to form thromboses, and accelerated atherosclerosis. Nicotine also elevates levels of low-density lipoprotein (LDL) cholesterol and triglycerides, and decreases levels of high-density lipoproteins (HDLs). It produces vasospasm and increased platelet aggregation, which can decrease myocardial oxygen supply.

Endogenous catecholamines are released with smoking, increasing blood pressure and heart rate, which produce an increase in myocardial oxygen demand. Specifically, smoking increases the risk of coronary heart disease to two to four times that of normal. Even smoking only a few cigarettes per day is correlated with an increased risk. This risk is even greater if the individual has hypertension, hypercholesterolemia, glucose intolerance, or diabetes because these conditions have a synergistic effect with smoking. Death rates after a myocardial infarction are higher among smokers. Cessation of smoking results in a 50% risk reduction from coronary heart disease within the first year, and a risk equal to that in nonsmokers after 10 years.

Hypertension is both a risk factor for the development of atherosclerosis and an outcome of it. Increases in both systolic and diastolic blood pressure are associated with an increased incidence of atherosclerosis. Diastolic blood pressure elevations are probably more significant because they represent the status of the cardiovascular system when it is at rest. Control of hypertension reduces the injury it produces to the vessel walls and, at a minimum, decreases the rate of atherosclerotic formations. Hypertension is often found in the presence of other risk factors. (See Chapter 16 for a discussion of hypertension.)

Cholesterol, the lipoproteins, and triglycerides are important in the discussion of atherosclerosis. Cholesterol is a necessary component of cellular membranes and is used in the manufacture of steroids within the body. Approximately 40% to 50% of the body's cholesterol is absorbed; the remainder is synthesized by the liver utilizing dietary saturated fats. Cholesterol, like other fats in the bloodstream, is highly insoluble and is transported to and from the body's cells within lipoprotein shells. Although there are several forms of lipoproteins, LDL and HDL are most important in the discussion of atherosclerosis. Receptors on the surface of the LDL molecule bind with receptors on cell membranes, allowing the molecule to be absorbed into the cell. These receptors abound in the muscle cells of arteries. The protein coat is dissolved, and the cholesterol is then used to meet the body's cellular needs. The excess cholesterol that is not removed is stored and acts as a cellular irritant, participating in endothelial injury and plaque formation. Since the early 1990s, evidence of the cardiovascular benefit of controlling serum lipid levels has been mounting. The precise mechanism is unclear, but the correlation between high serum levels of LDL and cholesterol is clearly significant in the development of atherosclerosis. HDL seems to serve as a protective mechanism in the formation of atherosclerosis. It is postulated that HDL can remove cholesterol from formations in the arterial walls and transport it back to the liver. Consequently, serum lipid profiles are closely supervised. An acceptable total cholesterol level for an adult who has no coronary disease is less than 200 mg/dL. Levels of LDL are felt to be detrimental if greater than 160 mg/dL. In those with known coronary disease, an LDL value of less than 100 mg/dL is thought to be beneficial. Protective levels of HDL are those greater than 45 mg/dL for men and greater than 55 mg/dL for women. A major intervention related to atherosclerosis is encouraging the consumption of a moderate-fat diet, with those fats being primarily polyunsaturated (from vegetable sources as opposed to animal). Additionally, exercise and weight control are effective in improving lipid profiles. Pharmacologic management of elevated levels of serum lipids may be recommended.

Glucose intolerance/metabolic syndrome and diabetes mellitus have been found to be associated with elevations in LDL levels and reduced levels of HDL, hypertension, and atherosclerosis. Glucose intolerance is often a precursor of diabetes mellitus, a disease in which an absolute lack of or a significantly decreased response to insulin produces a derangement in metabolism (see Chapter 41). Atherosclerosis is highly correlated with glucose intolerance, probably because of the alterations in carbohydrate and fat metabolism and the direct damage to vessel basement membrane with elevated blood glucose levels. The incidence of atherosclerotic diseases is much higher among those with diabetes mellitus than in the general population. It does not appear to be related to the degree of hyperglycemia, yet glycemic control has been found to decrease the incidence of acute myocardial infarction.

Obesity, defined as a body weight 30% or greater than ideal, is thought to be a contributing risk factor for atherosclerosis in that it may accelerate the process. Abdominally distributed obesity is a greater risk than an increase in body mass index. A desirable waist circumference for men is less than 40 inches, and less than 35 inches for women. Weight gain is associated with increasing serum cholesterol and LDL levels, increasing systolic blood pressure, glucose intolerance, and a sedentary lifestyle.

Physical activity has been found to increase HDL levels, collateral circulation, and vessel size and to decrease total cholesterol levels, glucose intolerance, body weight, and blood pressure. Clearly, all these findings can retard the development and mitigate the severity of atherosclerosis. Research likewise substantiates physical inactivity as a risk factor for CVD.

Stress and mental health have historically received considerable attention as risk factors for atherosclerosis, and this continues to be the case. It is extremely difficult to isolate these factors and examine them quantitatively and qualitatively. Stress results in the release of endogenous catecholamines that contribute to the increased work of the cardiovascular system. Subjectively, the rushed, stressed person is less inclined to exercise and eat wisely and more inclined to smoke and be hypertensive. Recently depression has been identified as a cardiovascular risk factor. Both stress and depression also may contribute to a delay in seeking treatment.

The importance of managing modifiable risk factors in the reduction of cardiovascular risk cannot be overstated. Recent research continues to substantiate the interrelationship of these risk factors. Many factors interact; for example, exercise can reduce the perception of stress. Stress management is related both to decreased measurements for body weight, total cholesterol and triglycerides, and hemoglobin A_{1C} (an indicator of diabetic glycemic control) and to increased measurements of HDL. Clearly, not only is the reduction of a single risk factor meaningful, but also there is an interactive, additive effect among risk factors.

Nonmodifiable risk factors. Certain risk factors are not modifiable and cannot be manipulated for prevention or to decrease the severity of atherosclerosis and CVD. Age, gender, ethnicity, and genetics are interrelated, and their impact as risk factors can be modulated by lifestyle changes.

With aging, changes occur in the arterial walls (see the "Geriatric Considerations" box) that predispose to the development of atherosclerosis. Men have a higher incidence of atherosclerosis earlier in life than women, but after menopause, the incidence and prevalence of CVD equalize. Postmenopausal status is often seen as an independent risk factor for CVD, generating increased attention in research regarding the potential cardioprotective role of estrogen. At this time, studies indicate postmenopausal women taking either estrogen alone or in combination with progesterone have an increased risk of thrombotic events (myocardial infarction, deep vein thrombosis, ischemic stroke) and of breast cancer. Consequently, such interventions are not recommended for cardiovascular risk reduction.

A strong family history of CAD is an important predictor of its occurrence and subsequent prognosis. The specific mechanism is uncertain, but most likely it is a combination of genetic and environmental factors. Certain of the modifiable risk factors are also known to have a genetic component.

Studies of ethnicity as a nonmodifiable risk factor associated with atherosclerosis have predominantly focused on the increased incidence of CAD and hypertension among black Americans compared with white Americans. Degree of risk differs among Caucasians, African Americans,

Hispanics, and Asians. Prevalence of smoking, diabetes, and hypertension is greater for African Americans and Hispanics, whereas Caucasians are more likely to have abnormal serum lipid levels. After correction for age, gender, risk factors, and pharmacologic treatment of hyperlipidemia, one study reported the least amount of coronary artery calcification to be among Asian men and women. Historically, American Indians have had very low rates of CVD. However, it is now the leading cause of death in this population, with the majority of CVD cases in those with diabetes, which has a high incidence in this population. These disparities highlight the overlap of genetic and environmental factors and illustrate how ethnicity, as an isolated independent variable, is very difficult to evaluate.

Clinical Manifestations and Diagnosis

Disease manifestations vary with the tissues involved and the severity of altered flow. Atherosclerosis is an underlying pathologic condition for much of the hypertension, renal disease, cardiac disease, PAD, and stroke seen in health care practice.

Approaches to diagnosis and treatment of decreased organ or tissue function vary. Patient history and physical assessment provide significant information. Noninvasive tests such as Doppler flow studies may identify areas of occlusion or diminished flow. *Plethysmography* may be used to measure changes in the relative size of extremities associated with blood flow. Ankle pressures are obtained with a blood pressure cuff and Doppler ultrasonography and compared with brachial blood pressures in the ankle–brachial (A/B) index. A normal A/B index is greater than or equal to 1.0; an index less than 1.0 is indicative of diminished arterial flow in the lower extremities. Exercise or stress testing may be performed to evaluate the pain of arterial occlusive disease (IC). Angiography—the radiologic study of blood flow—is a frequently used diagnostic examination. (See Chapter 18 for a discussion of CAD, Chapter 16 for a discussion of hypertension, Chapter 28 for a discussion of renal failure, and Chapter 44 for a discussion of stroke.)

Treatment

Identification of and interventions directed toward modifiable risk factors are the major thrusts of treatment, regardless of the organs or tissues affected. Nonpharmacologic interventions, such as reduction of body weight, cessation of smoking, implementation of an exercise program, and consumption of a moderate-fat diet, are the first-line actions. Drug therapy to decrease hypercholesterolemia is considered when the nonpharmacologic approaches are found to be ineffective or inadequate, or the presence of additional risk factors indicates that the patient would benefit from such interventions.

A wide variety of additional interventions may be undertaken, depending on the specific disorder and organ involved. Balloon angioplasty, the surgical radiologic fragmentation of atherosclerotic plaques by inflation of a specially equipped catheter, is commonly performed on both coronary and peripheral arteries. Laser angioplasty has been combined with balloon angioplasty to create an opening in significantly obstructed peripheral vessels before the balloon is inflated. Balloon angioplasty with stent placement also is used. When balloon angioplasty of the coronary arteries is unacceptable or fails to result in satisfactory improvement, coronary artery bypass graft surgery may be performed. Peripheral arterial bypass grafts are common interventions for the lower extremities and are named for their sites of origin and termination (e.g., aortofemoral, femoropopliteal).

Thromboangiitis Obliterans (Buerger Disease)

Thromboangiitis obliterans (Buerger disease) is a rare inflammatory condition affecting both small- and medium-size arteries and veins of the upper and lower extremities, producing varying degrees of obstruction. Although both arteries and veins are involved, the signs and symptoms relate to obstruction of arterial flow (see Box 15.2). The rarity of this disease affects the availability of research funding; therefore progress in understanding the pathogenesis is limited. What is known is that this vasculitis is strongly associated with smoking and that smoking cessation is essential to effective treatment. New approaches to angiogenesis have resulted in the development of significant collateral circulation and marked clinical improvement. If patients do not abstain from smoking, the disease is progressive, and amputation may be necessary.

Raynaud Syndrome

An extreme vasoconstriction producing cessation of flow to the fingers and toes produces the characteristic signs and symptoms of Raynaud syndrome. Sometimes earlobes or the tip of the nose are also affected. Attacks are recurrent, usually beginning in adolescence. They are most often initiated by cold or emotional distress. The phenomenon is classically characterized by a series of color changes in the involved area, starting with white, corresponding to the vasoconstriction. The affected area then becomes cyanotic, with the desaturation of blood remaining in the area. Finally, the tissue flushes red (reactive hyperemia) with the resumption of perfusion. The attack also commonly affects local nerve function, causing pain and/or numbness.

Raynaud syndrome is more prevalent among women, presenting between puberty and menopause, and there appears to be a genetic predisposition. Women have an increase in sympathetic tone of the vessels of the skin, causing young women to have basal cutaneous flows that are half those of young men. Knowledge regarding the complex regulation of local perfusion has expanded over the years since the condition was identified by Maurice Raynaud in 1862. The pathogenesis is an interplay of intrinsic structural factors (e.g., inflammatory activation and damage), extrinsic neuroregulation, locally produced mediators, and soluble mediators (including estrogen).

Various treatment modalities have been used with differing degrees of success. Because the precise cause of Raynaud syndrome is unknown, interventions have been directed to enhancing the circulation. Biofeedback and relaxation techniques may be beneficial. The most widely used drugs are calcium channel blockers, which produce vasodilation by interfering with calcium influx into vascular smooth muscle cells. Sympatholytic drugs have been studied and found to be more effective than placebos. Prostaglandin therapy has proved helpful. Persons with the syndrome are urged to protect themselves from cold temperatures, vibration, and nicotine and to use stress reduction interventions.

Aneurysms

As described previously, aneurysms are localized arterial dilations. The arterial wall deteriorates until it is weakened sufficiently to bulge outward. The underlying cause may be atherosclerotic changes in the vessel; a congenital weakness; or a weakening induced by infection, inflammation, or traumatic injury. Aneurysms are most frequently found in the cerebral circulation (circle of Willis or posterior circulation) and in the thoracic and abdominal aorta. Estimates are that between 10 and 15 million individuals are treated for cerebral aneurysms annually, with rupture occurring in 30,000. Mortality and morbidity associated with ruptured cerebral aneurysms are high, so early diagnosis and treatment are most desirable.

Classifications

Aneurysms are classified as true or false, depending on the layers of the arterial wall involved (Fig. 15.20). In *true aneurysms,* all three tunicae are involved (intima, media, and adventitia), whereas in *false aneurysms,* at least one tunica is left unaffected. In a false aneurysm, the muscle

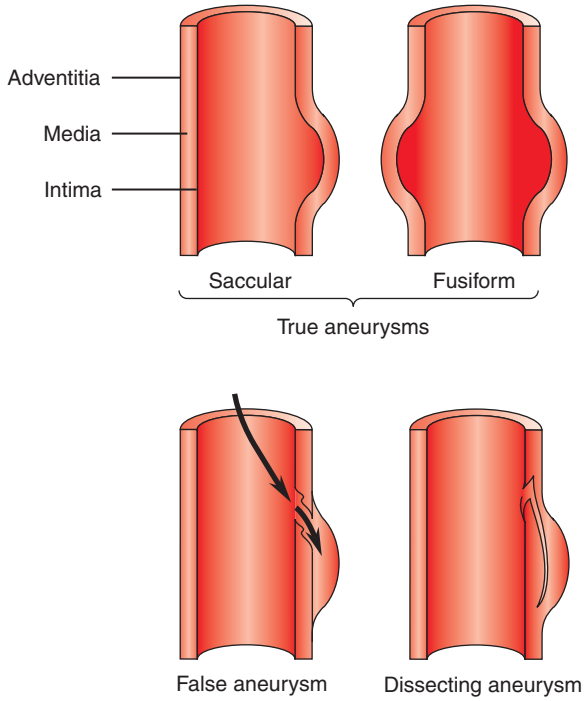


FIG 15.20 Classification of aneurysms. All three tunicae are involved in true aneurysms (fusiform and saccular). In false aneurysms, blood escapes between tunica layers and they separate. The muscle and fascia confine the leak; a thrombus forms and seals the leak. In a dissecting aneurysm, a tear in the intima creates a channel into which blood leaks, creating a hematoma. Continued expansion of the hematoma further separates the intima from the other layers, weakening the vessel.

tissue and fascia often confine the leaking blood, which enhances thrombus formation. False aneurysms are most often caused by trauma rather than vessel disease. True aneurysms are further divided by their shape and their size. In *saccular aneurysms,* the weakening is confined to one side of the vessel, producing a lateral ballooning. *Fusiform aneurysms* represent weakening on both sides of the vessel wall—a central ballooning. A *berry aneurysm* is the most common cerebral aneurysm; it is shaped like a berry, with a neck or stem.

All aneurysms can affect blood flow. Cerebral aneurysms are addressed in detail in Chapter 44. Of significant clinical concern is the *dissecting aortic aneurysm* (see Fig. 15.20). Here the tear in the arterial wall creates a channel for blood flow. The tear may be between the intima and media or between the media and adventitia. As more blood escapes into the space, the layers are separated from one another in both directions from the leak, and as the vessel becomes progressively weaker, it may rupture. Rupture can be explained by the law of Laplace—as the radius of the vessel increases, the tension in the wall increases. Rupture of a major vessel such as the aorta carries a high mortality.

Clinical Manifestations and Diagnosis

Signs and symptoms of a leaking or ruptured cerebral aneurysm are associated with increasing intracranial pressure and hemorrhagic stroke. Dissecting aortic aneurysms often present as sudden, severe, tearing pain that radiates into the back or abdomen. The patient may show signs and symptoms of shock. Renal blood flow or perfusion of the spinal nerves may be compromised if the descending abdominal aorta is affected. Renal failure or paraplegia may result. If the ascending aorta is affected, arterial blood flow to the head and upper extremities may be affected.

Diagnostic tests are somewhat dependent on location. Computed tomography (CT) and transesophageal echocardiography (TEE) are common diagnostic modalities for aortic aneurysms. TEE has proven to be reliable and is immediately available in an emergency setting. For cerebral aneurysms, CT, magnetic resonance imaging (MRI), and cerebral angiography are used.

Treatment

Dissecting aortic aneurysms are emergency situations and may be managed medically, surgically, or both. Medical intervention is directed at lowering the blood pressure to decrease the speed and severity of the dissection. Vasodilators are often administered parenterally. Surgical intervention involves resection (removal) of the aneurysm and insertion of a prosthetic graft. This procedure may require an open approach, but as with many surgeries, it may be done with a series of smaller incisions in an endovascular approach. If the aneurysm is extremely large, it may be inoperable.

A variety of interventions are used for cerebral aneurysms; these are briefly addressed in Chapter 44.

Acute Arterial Occlusion

Acute arterial occlusion is an emergency because it may result in such profound ischemia that the involved limb becomes gangrenous and sepsis may result. Acute arterial occlusion may be caused by a thrombus or embolus lodging in a major artery, or by external mechanical compression producing compartment syndrome. The result is an effective absence of arterial circulation to the extremity. Although it is usually attributable to a thrombus or an embolus, it may occur with vasospastic disease or trauma, as a complication of vascular surgery, or from swelling within a cast or tight dressing.

The classic signs and symptoms of acute arterial occlusion are known as the *six Ps. Pallor* occurs in the involved extremity. The patient may complain of *paresthesia,* and some degree of *paralysis* may be noted, due to the lack of oxygen to nerve cells. *Pain* is intense, continuous, and unrelated to activity. The skin is cold to touch *(polar)* and may be *pulseless* by palpation, although often a weak pulse may be noted by Doppler.

Perfusion must be restored or necrosis will develop because of ischemia to the extremity. The specific interventions vary with the etiology. Dressings may be loosened or casts cut if the cause is mechanical compression. If the cause is thrombotic or embolic, anticoagulant therapy may prevent enlargement and the formation of further thrombi. Bypass surgery or revascularization through thrombolytic therapy is usually attempted. Surgical removal of an embolism *(embolectomy)* may be necessary. If these approaches are not successful or ischemia is prolonged, amputation may be required.

KEY POINTS

- Common causes of arterial obstruction are atherosclerosis, inflammation, vasospasm, and aneurysms. Emboli are the usual cause of acute arterial occlusion.
- Atherosclerosis is the most common cause of chronic progressive arterial obstruction. Several risk factors for the development of atherosclerosis have been proposed, among them smoking, hyperlipidemia, male gender, advancing age, sedentary lifestyle, obesity, glucose intolerance, and a family history of cardiovascular disease (CVD).
- Acute arterial obstruction is accompanied by the classic manifestations known as the *six Ps:* pallor, paresthesia, paralysis, pain, pulselessness, and polar (cold to touch).

ALTERATIONS IN VENOUS FLOW

Pathologic venous conditions are the result of obstruction to flow (deep vein thrombosis) or structural alterations (valvular incompetence) and are primarily seen in the lower extremities.

Valvular Incompetence

Etiology and Pathogenesis

The intimal surface of veins periodically folds into valves to facilitate efficient flow (Fig. 15.21). When the valves are open, blood is propelled forward by the pressure changes exerted by the skeletal muscles and the intraabdominal and intrathoracic pumps. When this pressure decreases, backward flow of blood is prevented by proper closure of the valves (valvular competency). Valvular incompetence results in *venous insufficiency.* When the superficial veins are involved, the disorder is called *varicose veins.* Chronic venous insufficiency occurs when the deep veins are affected.

The cause of valvular incompetence is the overstretching of the valves as a result of excessive venous pressures. Veins are designed as low-pressure systems. After the blood leaves the high-pressure arterial bed, it passes into the fine capillary network, which slows flow and reduces the pressure. Blood flow through the veins is essentially accomplished by forces outside the veins—the skeletal muscle, intra-abdominal, and intrathoracic pumps. The highly distensible vein walls are capable of expanding to create a reservoir of blood. When the pressure against which the pumps must push is elevated for a prolonged period, the veins stretch and the valve cusps can no longer meet. Backflow results in further engorgement of the involved veins. The process is most frequently seen in people whose occupations require them to stand for long periods. The effect of gravity on venous flow accentuates the problem. Obesity and pregnancy also elevate venous pressure and may contribute to varicosity formation. As much as 10% of the European and North American population have valvular incompetence.

Clinical Manifestations and Treatment

Symptoms may include a feeling of heaviness or tension and pruritus. Thrombi can promote valve obstruction and further thrombus formation. In prolonged insufficiency, edema and stasis dermatitis (discoloration along the lower calf to ankle) may develop. Long-term insufficiency can lead to ulcer formation.

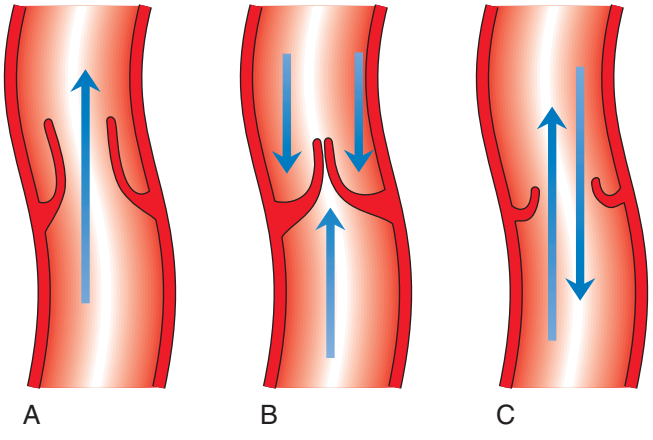


FIG 15.21 The venous valves. **A,** Open valves permit forward blood flow. **B,** Closed valves prevent backflow of blood. **C,** Incompetent valves, unable to close fully, allow blood to flow backward, producing venous insufficiency.

Prevention interventions include stopping smoking and beginning a walking program. Regular exercise has been shown to decrease future cardiovascular events. Drugs that interfere with platelet aggregation may be initiated, or revascularization procedures may be attempted.

Varicose Veins

Etiology and Pathogenesis

Varicosities are superficial, darkened, raised, and tortuous veins (Fig. 15.22). The greater saphenous vein is primarily affected, although varicosities may also develop in the lesser saphenous veins. Impaired venous return results in increased capillary pressure, and the involved limb may become edematous.

Clinical Manifestations and Treatment

Patients may complain of an aching, heavy discomfort, but they are primarily disturbed by the appearance of the varicosities. Venous ulcers may develop.

Many of the diagnostic tests used for the arterial system are used for the identification of venous disease. The patient history and physical assessment provide important baseline information. Doppler ultrasound and impedance plethysmography are among the most frequently used assessment tools.

Conservative medical interventions are designed to reduce venous pressure and enhance the venous pump, especially the skeletal muscle pump. Patients are encouraged to elevate their legs whenever possible and to avoid standing for long periods. Elastic stockings can facilitate venous return by enhancing the skeletal muscle pump. When sitting, patients are urged to not cross their knees or ankles. Exercise, particularly walking or swimming, is suggested. If appropriate, weight reduction is recommended.

More aggressive interventions may be initiated in severe cases, as illustrated in Fig. 15.22. *Sclerotherapy* involves the injection of a chemical that initiates an inflammatory process, and subsequent compression dressings force the lumen to collapse. The intima adheres to itself and heals, and the vein is obliterated. Collateral venous circulation meets

FIG 15.22 Varicose veins. Varicosities are best observed when the patient is standing because standing increases the pressure and causes the tortuous veins to become more visible. (From Black JM, Hawks J: *Medical-surgical nursing: clinical management for positive outcomes,* ed 8, Philadelphia, 2008, Saunders, p 1336.)

the need for venous return from the extremity. Surgical interventions include *vein stripping* and *vein ligation.* These are commonly performed as outpatient procedures and are often combined with sclerotherapy. Both of these treatment modalities depend on the presence of adequate deep venous structures to provide alternative routes for venous drainage.

Chronic Venous Insufficiency

Etiology and Pathogenesis

Chronic venous insufficiency results when valvular incompetence involves the deep veins (superficial femoral, anterior and posterior tibial, peroneal) of the legs. Communicating or perforating veins provide direct access between the superficial and deep veins. Because the pressure in the superficial veins remains elevated for a prolonged period, the deep veins are eventually affected. Individuals with chronic venous insufficiency often also have heart disease and a history of extremity trauma and phlebitis. Previous deep vein thrombosis is a risk factor.

Clinical Manifestations and Treatment

Venous stasis ulcers also develop as superficial veins rupture with the increased pressures associated with activity. The skin pigmentation becomes brown as small veins rupture, leaking RBCs, which are eventually broken down. Defining characteristics of chronic venous insufficiency are listed in Box 15.4.

Chronic venous insufficiency is primarily diagnosed clinically, but if necessary, ultrasound is considered the best method of evaluation. Many of the diagnostic techniques described for varicose veins may prove helpful, including Doppler ultrasound and impedance plethysmography. Treatment of venous ulcers is challenging and incorporates the use of compression as its foundation. Venous ulcers are prone to secondary infection, usually by *Staphylococcus aureus, Pseudomonas aeruginosa,* or β-hemolytic streptococci, requiring systemic antibiotic therapy. Other interventions previously described for superficial varicosities also are used for chronic venous insufficiency.

Deep Vein Thrombosis

Etiology and Pathogenesis

The pathophysiologic process of thrombus formation has been previously described. Acute venous obstruction is most frequently secondary to a thrombus in a deep vein of the lower extremities. Upper extremities are less frequently affected by deep vein thrombosis.

BOX 15.4 Defining Characteristics of Chronic Venous Insufficiency

Skin Assessment
- Warm, tough, and thickened to touch
- Pigmented areas, reddish brown
- Edema, especially at end of day
- Visible healed ulcers
- Evidence of varicose veins may be present

Pain Assessment
- Aching, cramping
- Sometimes decreases with ambulation
- Relieved by elevation

Ulcer Assessment
- Moderately painful
- Pink-red base
- Irregular, uneven edges
- Located on medial malleolus

Clinical Manifestations and Treatment

Deep vein thrombosis of the legs may be asymptomatic. Signs and symptoms, if present, typically include edema, manifestations of local inflammation, and possible dilated superficial veins secondary to the increased venous pressure. Pain may be present due to pressure on adjacent nerves and the inflammatory process.

Deep vein thrombosis is treated aggressively; deep vein thrombosis of the lower extremities and pelvic veins is the most frequent source of pulmonary emboli. Patients are often hospitalized so that intravenous anticoagulation therapy may be initiated; otherwise, they are treated on an outpatient basis with oral anticoagulants. Patients who have previously developed deep vein thromboses are at risk for further hypercoagulation and may undergo long-term prophylactic anticoagulation with antiplatelet therapy and parenteral anticoagulants with subsequent hospitalization for any reason.

KEY POINTS
- Common causes of venous obstruction are incompetent valves (as may occur with obesity, pregnancy, right heart failure, or prolonged standing), producing varicose veins and chronic venous insufficiency, and obstruction by deep vein thrombosis.
- Edema, venous stasis ulcers, and pain usually accompany chronic venous obstruction.
- Deep vein thrombosis is potentially life threatening because of the likelihood of embolization to the pulmonary circulation. It is treated aggressively with the administration of anticoagulants.

ALTERATIONS IN LYMPHATIC FLOW

Lymphedema

Etiology and Pathogenesis

Lymphedema occurs when the normal flow of lymph is obstructed or altered in some fashion (Fig. 15.23). This results in the collection of lymphatic fluid in the interstitium, initiating an inflammatory response, hypertrophy of subcutaneous adipose tissue, and fibrotic changes. Primary lymphedema is related to a congenital anomaly or dysfunction of the lymphatic system. Secondary lymphedema develops in association with a disease process or is iatrogenic (a consequence of medical intervention) in origin. Throughout the world, secondary lymphedema is most commonly caused by an infection by filarial worms that migrate to the nodes of the lymphatic system, producing an obstruction of flow. Infection by this nematode *(Wuchereria bancrofti)* affects more than 90 million people worldwide. In the United States secondary lymphedema is most frequently caused by the surgical removal of lymph nodes, as with some breast cancers, or by the destruction of the lymphatics from direct radiation to lymph nodes in the management of various malignancies.

Upper extremity lymphedema occurs in 15% to 28% of breast cancer survivors, is most common in those who had axillary lymph node dissection, and can present a few days or 6 to 8 weeks after surgery or radiation therapy. Lower extremity lymphedema occurs in as many as 80% of those who had lymph node dissection in the groin or those who have a compression of pelvic or inguinal lymph nodes.

Cancer-related lymphedema is a progressive and chronic syndrome of abnormal swelling and multiple symptoms resulting from the accumulation of lymph fluid from the obstruction or disruption of the lymphatic system. The most perplexing aspect of this late effect of

FIG 15.23 Types of lymphedema. **A,** Lymphedema of an arm secondary to surgical alterations in the lymphatic system associated with mastectomy. **B,** Lymphedema of a leg. (From Black JM, Hawks J: *Medical-surgical nursing: clinical management for positive outcomes,* ed 8, Philadelphia, 2008, Saunders, p 1339.)

cancer treatment is that lymphedema can occur even after less invasive surgical procedures, such as sentinel lymph node biopsy or partial mastectomy. In addition, lymphedema affects a large proportion of cancer survivors with a variety of malignancies, including gynecologic cancer (20%), melanoma (16%), genitourinary cancer (10%), and head and neck cancer (4%).

Clinical Manifestations, Diagnosis, and Treatment

Most often, lymphedema affects the extremities, but it may be found in the region of the head and neck, trunk, or genitalia. Initially it presents as a soft, pitting edema, but as it progresses the inflammatory response is activated and lipogenesis, fat deposition, and fibrotic changes occur. Without early diagnosis and treatment, lymphedema may progress from an acute problem to a chronic one, where congestion produces thick and roughened skin *(brawny edema)* and a large deformed limb.

Diagnosis is primarily one of exclusion, with other causes of localized edema being ruled out. A complete history and physical examination will allow the elimination of cardiac, renal, and peripheral vascular etiologies. The primary diagnostic test uses the injection of radioisotopes (lymphoscintigram) to assess the overall function of the lymphatic system. Ultrasound, duplex ultrasound, CT, and MRI are additional options and produce unique results.

There is no cure for lymphedema; interventions are both medical and surgical. Medical treatment includes use of external pneumatic compression devices, elastic stockings, and exercise. Decongestive lymphatic therapy or complete decongestive physiotherapy combines skin care, massage, compression dressings, and exercise. For many individuals, this therapy results in a 40% to 60% reduction in edema. At this time, no medication has been approved for lymphedema therapy by the Food and Drug Administration; although diuretics have been used with minimal effect, they are not recommended. Surgical interventions are limited to very select patients, as they may actually be harmful. Resections (debulking) remove subcutaneous tissue, and bypass procedures use lymphatic-venous anastomosis. A new approach involves the use of liposuction for the removal of subcutaneous fatty tissue.

KEY POINTS

- Obstruction of lymph flow is most commonly the result of surgical removal of, or radiation damage to, lymphatic vessels during treatment of cancer.
- Manifestations of lymphatic obstruction include regional edema and thickened subcutaneous tissue.

SUMMARY

The circulatory system is organized to facilitate its dual functions of oxygen and nutrient transport and metabolic waste product removal. The arrangement and unique structure of the circulatory vessels permit these functions to be accomplished.

An understanding of the principles and control of flow aids in the comprehension of the pathologic conditions that result in alterations in flow. Principles of flow, or the hemodynamics of the circulation, include concepts and physical laws relating to relationships of flow, pressure and

resistance, velocity, laminar and turbulent flow, and wall tension and compliance. Control of blood flow occurs through both extrinsic and intrinsic mechanisms. Lymphatic flow is controlled through the lymphatic pump system, governed by skeletal muscle and the smooth muscle of organs and arteries.

Pathophysiologic changes that result in alterations in blood flow can be classified as being caused by either obstruction (thrombosis, emboli, vasospasm, inflammation, mechanical compression) or structural alterations (valvular incompetence, arteriosclerosis/atherosclerosis, aneurysms, AVFs). Conditions that produce alterations in arterial or venous flow are the result of one of these primary processes. Pathology of the lymphatic system is essentially the result of disruption of the normal pressure relationships or an obstruction within the circulatory system; proper functioning of the lymphatic system depends on the appropriate functioning of the vascular system.

RESOURCES

Physiology of Blood Flow

Guyton AC, Hall JE: *Textbook of medical physiology*, ed 13, Philadelphia, 2016, Saunders.

Mensah GA: Healthy endothelium: the scientific basis for cardiovascular health promotion and chronic disease prevention. *Vascul Pharmacol* 46:310–314, 2007.

Nohria A, et al: Rho kinase inhibition improves endothelial function in human subjects with coronary artery disease. *Circ Res* 291:1426–1432, 2006.

Oka RK: Peripheral arterial disease in older adults: management of cardiovascular disease risk factors. *J Cardiovasc Nurs* 21:S15–S20, 2006.

Rajendran S, Chirkov YY: Platelet hyperaggregability: impaired responsiveness to nitric oxide ("platlet NO resistance") as a therapeutic target. *Cardiovasc Drugs Ther* 22(3):EPub, 2008.

Victor VM, et al: Regulation of oxygen distribution in tissues by endothelial nitric oxide. *Circ Res* 104:1178–1183, 2009.

Warnholtz A, et al: AT 1-receptor blockade with irbesartan improves peripheral but not coronary endothelial dysfunction in patients with stable coronary artery disease. *Atherosclerosis* 194:439–445, 2007.

Alterations in Blood Flow and Atherosclerosis

AHA Statistical Update: Executive summary: heart disease and stroke statistics—2017 update: a report from the American Heart Association. *Circulation* 135:e146–e603, 2017. http://circ.ahajournals.org/content/135/10/e146.

American Heart Association: Atherosclerosis: Available at www.heart.org/HEARTORG/Conditions/Cholesterol/WhyCholesterolMatters/Atherosclerosis_UCM_305564_Article.jsp: (Accessed 1 June 2016).

Au TB, Golledge J, Walker PJ, et al: Peripheral arterial disease - diagnosis and management in general practice. *Aust Fam Physician* 42(6):397–400 4p, 2013.

Baumhakel M, Böhm M: Recent achievements in the management of Raynaud's phenomenon. *Vasc Health Risk Manag* 6:207–210, 2010.

Budoff MJ, et al: Ethnic differences of the presence and severity of coronary atherosclerosis. *Atherosclerosis* 187:343–350, 2006.

Carson AP, et al: Ethnic differences in hypertension incidence among middle aged and older adults: the multi ethnic study of atherosclerosis. *Hypertension* 57:1101–1107, 2011.

Cheek D, Sherrod M, Tester J: Women and heart disease: what's new? *Nursing* 38:36–42, 2008.

Chobanian AV: Seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 complete report. *Hypertension* 42:1206–1252, 2003.

Cloutier LM: Diagnosis of pulmonary embolism. *Clin J Oncol Nurs* 11:343–348, 2007.

Cooke JP, Marshall JM: Mechanisms of Raynaud's disease. *Vasc Med* 10:293–307, 2005.

Couri CBE, et al: Mönckeberg's sclerosis—is the artery the only target of calcification? *BMC Cardiovasc Disord* 5:34, 2005. Available at: www.biomedcentral.com/1471-2261/5/342005. (Accessed 8 November 2011).

Daubenmier JJ, et al: The contribution of changes in diet, exercise, and stress management to changes in coronary risk in women and men in the multisite lifestyle intervention program. *Ann Behav Med* 33:57–68, 2007.

Dawber TR: *The Framingham Study: the epidemiology of atherosclerotic disease*, Cambridge, MA, 1980, Harvard University Press.

Donato G, et al: Abdominal aortic aneurysm repair in octogenarians: mith (sic) or reality? *J Cardiovasc Surg* 48:697–703, 2007.

Evered A: Understanding cholesterol and its role in heart disease. *Nurs Times* 103:28–29, 2007.

Ganong WF: *Review of medical physiology*, ed 23, New York, 2009, McGraw-Hill.

James PA, et al: Evidence-based guideline for the management of high blood pressure in adults: Report from the panel members appointed to the eight joint national committee (JNC 8). *JAMA* 311(5):507–520, 2014.

Keevil JG, et al: Implications of cardiac risk and low-density lipoprotein cholesterol distributions in the United States for the diagnosis and treatment of dyslipidemia: data from National Health and Nutrition Examination Survey 1999-2002. *Circulation* 115:1363–1370, 2007. Epub.

Klonaris C, et al: Changing patterns in the etiology of acute lower limb ischemia. *Int Angiol* 26:49–52, 2007.

Libby P: Inflammatory mechanisms: the molecular basis of inflammation and disease. *Nutr Rev* 65:S140–S146, 2007.

Marik PE, Plante LA: Venous thromboembolic disease and pregnancy. *N Engl J Med* 359(19):2025, 2008.

Mazzone P, et al: Pathophysiological impact of cigarette smoke exposure on the cerebrovascular system with a focus on the blood-brain barrier: expanding the awareness of smoking toxicity in an underappreciated area. *Int J Environ Res Public Health* 7:4111–4126, 2010.

McKenry L, Tessier E, Hogan MA: *Mosby's pharmacology in nursing*, ed 22, Philadelphia, 2006, Mosby.

Mercado C, Jaimes EA: Cigarette smoking as a risk factor for atherosclerosis and renal disease: novel pathogenic insights. *Curr Hypertens Rep* 9:66–72, 2007.

Monahan FD, et al: *Phipps' medical-surgical nursing: health and illness perspectives*, ed 8, Philadelphia, 2007, Mosby.

Nabel EG, Braunwald E: A tale of coronary artery disease and myocardial infarction. *N Engl J Med* 366:54–63, 2012.

Pagana KD, Pagana TJ: *Mosby's manual of diagnostic and laboratory tests*, ed 3, St Louis, 2008, Mosby.

Piazza G, Creager MA: Thromboangiitis obliterans. *Circulation* 121(16): 1851–1861, 2010.

Pottie A: Measuring cholesterol levels. *Nurs Stand* 21:42–47, 2007.

Puéchal X, Fiessinger JN: Thromboangiitis obliterans or Buerger's disease: challenges for the rheumatologist. *Rheumatology (Oxford)* 46:192–199, 2007.

Roger VL, et al: Heart disease and stroke statistics 2011 update: a report from the American Heart Association. *Circulation* 123(4):e18, e209, 2011. (Accessed 8 November 2011).

Rossouw J, et al: Postmenopausal hormone therapy and risk of cardiovascular disease by age and years since menopause. *JAMA* 297:1465–1477, 2007.

Stone NJ, Bilek S, Rosenbaum S: Recent National Cholesterol Education Program Adult Treatment Panel III Update: adjustments and options. *Am J Cardiol* 96(Suppl):53E–59E, 2005.

Wright I: Cerebral aneurysm—treatment and perioperative nursing care. *AORN J* 85:1172–1186, 2007.

Venous and Lymphatic Disorders

Cormier JN, Askew RL, Mungovan KS, et al: Lymphedema beyond breast cancer: a systematic review and meta-analysis of cancer-related secondary lymphedema. *Cancer* 116:5138–5149, 2010.

Etufugh CN, Phillips TJ: Venous ulcers. *Clin Dermatol* 25:121–130, 2007.
Fu MR, Deng J, Armer JM: PEP: Cancer-related lymphedema: evolving evidence for treatment and management from 2009-2014. *CJON* S3.18:68–79, 2014.
Grey JE, Enoch S, Harding KG: Venous and arterial leg ulcers. *BMJ* 332: 347–350, 2006.
McLaughlin SA, Wright MJ, Morris KT, et al: Prevalence of lymphedema in women with breast cancer 5 years after sentinel lymph node biopsy or axillary dissection: objective measurements. *J Clin Oncol* 26:5213–5219, 2008.
Warren AG, et al: Lymphedema: a comprehensive review. *Ann Plast Surg* 59:464–472, 2007.

16

Alterations in Blood Pressure

Benjamin J. Miller

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- How do changes in cardiac output and systemic vascular resistance affect blood pressure?
- How is blood pressure regulated on a short- and long-term basis?
- What are the risk factors for the development of primary hypertension?
- How is secondary hypertension defined, and what are the common etiologies?
- How is hypertension detected, classified, and managed?
- What are the end-organ consequences of inadequately controlled hypertension?
- What are the differences between hypertensive emergency and hypertensive urgency, and how are they managed?
- What are the risk factors for orthostatic hypotension, and how is the condition managed?

CHAPTER OUTLINE

Arterial Blood Pressure, 337
Determinants of Systemic Blood Pressure, 337
Measurement of Blood Pressure, 338
Components of Blood Pressure Measurement, 338
Direct Measurement of Blood Pressure, 339
Indirect Measurement of Blood Pressure, 339
Mechanisms of Blood Pressure Regulation, 341
Short-Term Regulation of Systemic Blood Pressure, 341
Long-Term Regulation of Systemic Blood Pressure, 341
Normal Fluctuations in Systemic Blood Pressure, 343
Hypertension, 343
Definition and Classification, 343
Primary Hypertension, 344
Subtypes, 344
Risk Factors, 344
Outcomes, 345
Treatment Interventions, 346
Secondary Hypertension, 348
Hypertensive Emergencies and Urgency, 349
Low Blood Pressure, 350

Meeting the needs of the body's tissues for oxygen and nutrients requires both adequate blood flow at the tissue level and sufficient perfusion pressure systemically to force that blood forward. The systemic arterial blood pressure provides the momentum, and the tissues depend on its preservation to ensure their metabolic needs are met. This maintenance requires a complex regulatory system. The body's organs can be damaged if the perfusion pressure is insufficient or if it is excessive.

ARTERIAL BLOOD PRESSURE

As described in Chapter 15, oxygenated blood is propelled from the left side of the heart into the arterial circulatory system, and following a pressure gradient, travels to the capillary beds of the body's tissues (Fig. 16.1). There, oxygen and nutrients are exchanged for metabolic wastes, and the blood then returns to the right side of the heart via the venous circulatory system, where it passes through the lungs to repeat the process. It is the pressure difference between the left and right sides of the heart that produces the gradient allowing this systemic movement of blood. The arterial blood pressure is produced by the force of the left ventricular contraction overcoming the resistance of the aorta to open the aortic valve, and is the pressure maintained in the arterial system throughout the cardiac cycle.

Determinants of Systemic Blood Pressure

The systemic arterial blood pressure is the physiologic result of the cardiac output and the resistance to the ejection of blood from the heart. Cardiac output (CO) is the product of two variables: stroke volume (SV) and heart rate (HR) (CO = SV × HR). SV is the specific volume of blood leaving the heart with each contraction, which itself is determined by the volume of blood in the heart before systole (end-diastolic volume) and the contractility of the myocardium. The end-diastolic volume is determined by the amount of blood returned to the heart between contractions, and is typically called the heart's *preload.* Stroke volume multiplied by the number of contractions of the heart per minute (heart rate) determines the amount of blood leaving the heart—the cardiac output, measured in liters per minute. The resistance to ejection into the arterial circulation is known as the *systemic vascular resistance* (SVR) and is determined by the radius of arteries and the

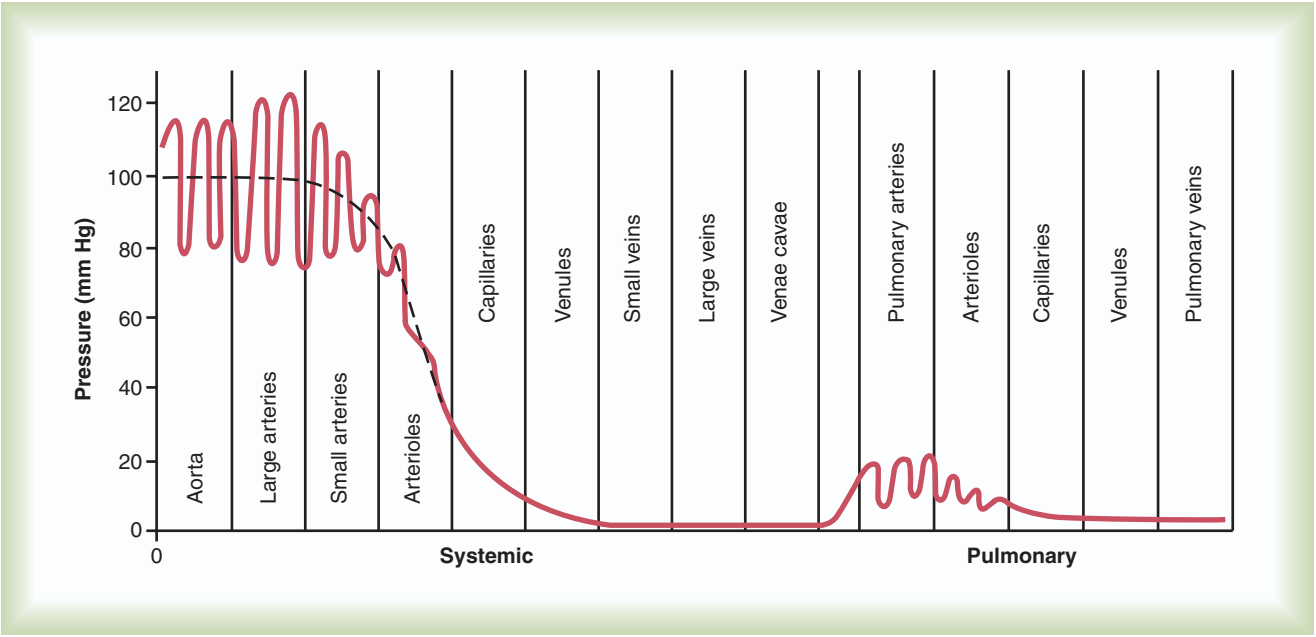


FIG 16.1 Normal pressures throughout the vascular system in the supine position. (From Hall JE: *Guyton and Hall textbook of medical physiology*, ed 13, Philadelphia, 2016, Saunders.)

degree of vessel compliance. SVR is synonymous with cardiac *afterload*, and can be altered by constricting or relaxing (dilating) arterial smooth muscle. It can be calculated by using a derivation of Poiseuille's law (see Chapter 15). This physical law states that in a tube with laminar flow, resistance is primarily determined by three factors: the radius of the tube, the length of the tube, and the viscosity of the fluid. Applied to SVR, because the viscosity of the blood and the total length of the arterial system are normally relatively constant, the radius of the arterioles becomes the major determinant of resistance. Therefore the formula for blood pressure is BP = CO × SVR. Alteration in any one of these variables will result in a change in blood pressure. This basic concept is important to normal physiologic function, disorders of blood pressure, and the therapeutic interventions undertaken to treat them. The pulmonary vascular bed contributes minimally to total systemic resistance and is seen as a separate resistance system, called *pulmonary vascular resistance*. It has its own pathology discussed in Chapter 21.

Measurement of Blood Pressure

Components of Blood Pressure Measurement

Arterial blood pressure is measured from its highest point during cardiac systole to its lowest during diastole. These are referred to as *systolic pressure* and *diastolic pressure*, respectively, and are measured in millimeters of mercury (mm Hg). During ventricular contraction, the pressure in the aorta rises to an average peak value of approximately 110 mm Hg in the adult (see Fig. 16.1). Whatever this peak pressure may be, it is referred to as the *systolic blood pressure*. The smooth muscle of the aorta passively recoils from this point, ejecting blood forward into the peripheral arteries at that given pressure. Stroke volume is the primary factor affecting systolic pressure; an increase or decrease in SV produces a corresponding change in systolic blood pressure. During ventricular diastole, the pressure in the arterial system falls to an average minimum value of 70 mm Hg in the adult. The value of this minimum pressure is called the *diastolic blood pressure*. SVR is the major determinant of diastolic blood pressure; an increase or decrease in diastolic pressure is the result of a corresponding increase or decrease in arterial resistance (SVR). The difference between systolic and diastolic blood pressure is termed the *pulse pressure*. Therefore the pulse pressure for a systolic pressure of 110 mm Hg and a diastolic pressure of 70 mm Hg would be 40 mm Hg.

TABLE 16.1 Blood Pressure Classification in Children and Adolescents

SBP and DBP <90th percentile*	Normal for children and adolescents
Average SBP and/or DBP ≥90th percentile but <95th percentile† or SBP ≥120 and/or DBP ≥80 mm Hg	Prehypertension in children or adolescents
SBP/DBP ≥95th percentile and <99th percentile plus 5 mm Hg	Stage I hypertension
SBP/DBP ≥99th percentile plus 5 mm Hg	Stage II hypertension

Selected data from Xi B, Zong XN, Kelishadi R, et al: Establishing international blood pressure references among nonoverweight children and adolescents aged 6 to 17 years. *Circulation, 133*(4):398-408, 2016.
DBP, Diastolic blood pressure; *SBP,* systolic blood pressure.
*For age, height, and gender.
†For age, height, and gender measured on at least three separate occasions.

Systolic and diastolic values are normed by age. Standards for the identification of normal blood pressure and levels of abnormal elevation have been established. The most precise standards for children are those based on height, age, and gender (Table 16.1). Standards for blood pressure have likewise been determined for the adult (Table 16.2).

Mean arterial pressure (MAP) is the calculated average pressure within the circulatory system throughout the cardiac cycle. Because more time is spent in diastole than in systole, MAP is not the arithmetic average of diastolic and systolic pressure, but rather reflects the relative time spent in each portion of the cardiac cycle. The calculation may be performed by computer during direct arterial blood pressure measurement, as described later, but is most conveniently determined by a

TABLE 16.2 Blood Pressure Classification in Adults

JNC 7 Blood Pressure Classification in Adults

Category	SBP (mm Hg)*	DBP (mm Hg)*
Normal	<120	<80
Prehypertension	120–139	80–89
Stage 1 hypertension	140–159	90–99
Stage 2 hypertension	≥160	≥100

From Chobanian AV: Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: The JNC 7 Complete Report, *Hypertension* 42:1206–1252, 2003.
DBP, Diastolic blood pressure; *SBP,* systolic blood pressure.
*Classification determined by the higher value.

ESH-ESC Blood Pressure Classification in Adults

Category	SBP (mm Hg)*	DBP (mm Hg)*
Optimal	<120	<80
Normal	120-129	80-84
High normal	130-139	85-89
Grade 1 hypertension	140–159	90–99
Grade 2 hypertension	≥160-179	100-109
Grade 3 hypertension	≥180	≥110
Isolated systolic hypertension	≥140	<90

From Mancia G, Fagard R, Narkiewicz K, et al: 2013 ESH/ESC practice guidelines for the management of arterial hypertension, *Blood Press* 23(1):3-16, 2014.

JNC 8 Blood Pressure Treatment Goals in Adults

Population	SBP (mm Hg)*	DBP (mm Hg)*
Age <60 all health states	< 140	< 90
Age > 60 with CKD or DM	< 140	< 90
Age ≥ 60 without CKD or DM	< 150	< 90

DBP, Diastolic blood pressure; *SBP,* systolic blood pressure; CKD, Chronic kidney disease; DM, diabetes mellitus.
From James PA, Oparil S, Carter BL, et al: 2014 evidence-based guideline for the management of high blood pressure in adults: Report from the panel members appointed to the eighth joint national committee (JNC 8). *JAMA, 311*(5):507-520, 2014.

simple formula using the values of blood pressure obtained indirectly. Several formulas are available, and they may use systolic, diastolic, or pulse pressures; the most common formula uses the systolic and diastolic pressures as follows:

$$\frac{(2 \times \text{diastolic pressure}) + \text{systolic pressure}}{3}$$

For a person with a systolic pressure of 110 mm Hg and a diastolic pressure of 70 mm Hg, the MAP would be:

$$(2 \times 70) + 110 = 250/3 \text{ or approximately } 83 \text{ mm Hg}$$

MAP is used clinically as part of cardiovascular assessment and in the incremental adjustment (titration) of parenterally administered vasoactive drugs.

Direct Measurement of Blood Pressure

Direct measurement of blood pressure is one aspect of hemodynamic monitoring and requires an intraarterial catheter and specialized equipment to transduce the arterial fluid pulsations into electrical signals. The catheter most often is placed in the radial artery. These signals are then displayed on a computer screen as waveforms, and the systolic, diastolic, and MAPs are digitally represented. This is the most accurate method of measuring blood pressure available, but is typically only performed in controlled settings, such as surgical or critical care units, and carries its own risk of measurement error. A detailed discussion of hemodynamic monitoring is beyond the scope of this text.

Indirect Measurement of Blood Pressure

Blood pressure is most commonly measured by indirect means at the brachial artery, using a sphygmomanometer and a stethoscope for auscultation or an automated oscillometric system such as Dinamap or the Welch Allyn Spot Vital Signs. Wrist or finger monitors are not recommended because of the inaccuracy of the values obtained compared with brachial measurements. Specific, evidence-based standards are available for the correct use of these noninvasive automated systems for adults and children, including scheduled calibration, and they are inherently less accurate if the blood pressure is significantly increased or decreased, or if there are cardiac dysrhythmias. Because the values in blood pressure references are based on the auscultatory method, and it is the easiest method and least stressful to patients, it is the preferred measurement technique. Although the brachial artery is typically used for convenience, certain assessment procedures require recording the blood pressure at other arterial sites (e.g., ankle–brachial index). Several studies have reported differences between the right and left arm pressures, but no pattern of differences is evident. Other studies report that in the absence of disease, systolic pressures do not differ significantly at a clinical or statistical level between the right and left arms. In practice, it is recommended that blood pressures be initially taken in both arms and the arm with the highest value be recorded. In situations such as a shock state, when systolic and diastolic pressures cannot be auscultated, the systolic pressure alone may be obtained by palpation or by amplification of the pulse using ultrasound technology (Doppler pressure).

Auscultated and oscillometric blood pressure measurements are burdened with the potential of measurement error, in both reliability and validity (Table 16.3). This dictates the need for careful technique, and in most cases enhances the value of trend data as opposed to individual readings. The individual patient's heart rate, degree of arterial compliance, and dynamics of blood flow may vary over time. Inappropriate blood pressure cuff size, arm position, and both the visual and auditory acuity of the clinician may affect the accuracy of individual readings. An additional source of error has been named the "white coat effect" for the elevation of blood pressure when taken in a clinic or office environment. First described in 1897 by Scipione Riva-Rocci, who was the first to document assessing the systolic pressure by palpating the brachial artery, these situational elevations in blood pressure are of concern because treatment may be initiated based on inaccurate data. This condition is most common in older individuals of either gender, but may occur at any age. Pickering and colleagues report that in approximately 15% to 20% of patients with stage 1 hypertension, elevated blood pressure may only be persistent under these circumstances. Significant pressure differences have been found using the automatic noninvasive technology between the supine, 45-degree elevation of the head of the bed and sitting position in the same patient, and between multiple body positions using the auscultatory method. Normal values are based on the subject being seated, with the back supported and the arm at heart level. Specific recommendations regarding all aspects of

TABLE 16.3 Intrinsic and Extrinsic Factors That Influence Indirect Blood Pressure Accuracy

Factor	Effect on Blood Pressure Measurement
Intrinsic Factors	
Heart rate	Elevated or decreased
Arterial compliance	Elevated or decreased
Alterations in flow dynamics	Elevated or decreased
Respiratory rate	Normal increase on inspiration
Extrinsic Factors	
Cuff	
Too small	Falsely elevated
Too large	Falsely decreased
Supine position	≈5 mm Hg lower DBP
Seated, back not supported	≈6 mm Hg increased DBP
Crossed legs	Increased SBP 2–8 mm Hg
Seated, arm position	
Above heart	Falsely decreased
Below heart	Falsely elevated
Inadequately supported	Falsely elevated
Excessive stethoscope pressure	Diastolic pressure falsely decreased
<1 min between measurements	Falsely elevated
Deflation rate >3 mm Hg/sec	Falsely decreased SBP and increased DBP
Exercise, eating, smoking, intake of caffeine ≤30 minutes before measurement	Falsely elevated
Talking during measurement	Falsely elevated
"White coat effect"/anxiety	Falsely elevated
Decrease in recorder auditory or visual acuity	Falsely elevated or decreased
Recorder bias	Falsely elevated or decreased

indirect measurement are provided by the American Heart Association, and sources of error within pediatric populations also have been documented.

The recommended approach for obtaining an auscultated blood pressure is a two-step approach, beginning with inflating the cuff to the point at which the pressure obliterates the palpated radial pulse (systolic pressure). The pressure is completely released, and after 15 to 30 seconds the cuff is reinflated to 30 mm Hg above that point and then gradually deflated while the clinician listens through the stethoscope with the diaphragm placed over the brachial artery and monitors the position of the mercury in the sphygmomanometer. The return of blood flow through the artery is signaled by the sounds produced by the turbulent flow through the partially occluded artery and named after the Russian physician who first described them in 1905 *(Korotkoff sounds)*. This sound is recorded as the systolic pressure. As the pressure continues to be released, sounds change in intensity until the point at which the Korotkoff sounds disappear, which is noted as the diastolic pressure (Table 16.4). Nurse researchers in Britain found statistically significantly lower diastolic values using this approach compared with a one-step approach. In this approach, the systolic pressure was estimated by palpating the brachial artery during cuff inflation, and inflation continued 30 mm Hg beyond that point before proceeding with deflation. It was postulated that the first inflation and occlusion produced a reactive vasodilation that could be responsible for this difference. Regardless, the auscultation of Korotkoff sounds results in systolic values that are

TABLE 16.4 Korotkoff Sounds

Phase	Description
I	Initiation of clear tapping sounds—systolic blood pressure
II	Murmuring or swishing sounds
III	Increase in intensity and crispness of sounds
IV	Muffling of sounds
V	Disappearance of sounds—diastolic blood pressure

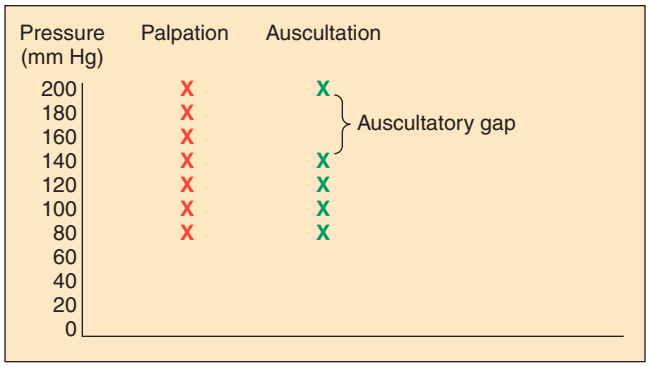


FIG 16.2 Auscultatory gap. Palpating the blood pressure (BP) before auscultation allows assessment of the true systolic BP. Palpated BP equals 200/P. The same result can often be obtained by elevating the arm overhead for 30 seconds before inflating the cuff. Auscultated BP when the cuff is inflated to only 180 mm Hg results in a falsely low value of 140/80 mm Hg.

lower than those obtained by direct, intraarterial blood pressure measurement. Older patients often have a period during measurement when the Korotkoff sounds disappear, returning 20 to 40 mm Hg later. This *auscultatory gap* may be attributed to intraarterial pressure fluctuations associated with hypertension (Fig. 16.2) and can often be eliminated by elevating the arm above the level of the head for 30 seconds before cuff inflation. This approach is postulated to enhance the audibility of Korotkoff sounds by increasing arterial flow following the increase in venous return.

Increasingly, self-monitoring of blood pressure is being performed at home. Potential sources of error as well as optimal schemes of measuring and recording have been identified. It has been found that the values documented in this setting are more accurate, if correctly obtained, because of the elimination of the white coat effect.

KEY POINTS

- Systemic arterial blood pressure varies with the cardiac cycle. The highest pressure (systolic) corresponds to ejection of blood from the left ventricle into the aorta. The lowest point in pressure (diastolic) occurs at the end of diastole, just before the next ventricular contraction.
- Blood pressure is the product of the cardiac output (CO) (HR × SV) and systemic vascular resistance (SVR). Changes in any of these variables will change blood pressure. The arterioles create most of the resistance in the vascular system; changes in the diameter of these vessels profoundly affect SVR and therefore blood pressure.
- The difference between the systolic and diastolic pressures is called the *pulse pressure*. The average pressure within the systemic arterial system is the mean arterial pressure (MAP), mathematically derived from the two pressure values.
- Blood pressure can be directly measured by placement of a catheter within an artery and utilization of specific computer software. More routinely it is

measured by auscultation. Systolic pressure is recorded as the onset of the Korotkoff sounds, and their disappearance is recorded as the diastolic pressure.

- Erroneous blood pressure values may be obtained because of a missed auscultatory gap, hydrostatic pressure changes associated with arm position, inappropriate cuff size, observer error, and other factors.

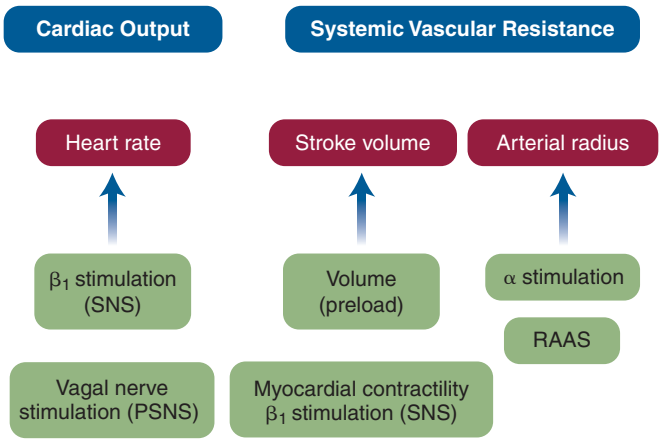


FIG 16.3 Systemic arterial blood pressure is controlled through influences on each of its variables: heart rate, stroke volume, and SVR. Some of these provide short-term adjustments, whereas others affect the long-term management of blood pressure. *PSNS,* Parasympathetic nervous system; *RAAS,* renin–angiotensin–aldosterone system; *SNS,* sympathetic nervous system.

MECHANISMS OF BLOOD PRESSURE REGULATION

Arterial blood pressure is physiologically controlled on both a short-term and a long-term basis. Regulation of blood pressure is achieved through changes in factors that affect the primary determinants of blood pressure: heart rate, stroke volume, and SVR (Fig. 16.3). These variables are affected by a complex interplay between neural, humoral, and renal factors to maintain stability in the face of ever-changing internal and external environmental demands. An understanding of these mechanisms is essential to exploring pathophysiologic alterations. Blood pressure normally fluctuates over the course of 24 hours due to physiologic changes associated with circadian rhythm.

Short-Term Regulation of Systemic Blood Pressure

Changes in blood pressure must occur quickly to accommodate behavioral changes (e.g., position changes, exercise), emotional changes (e.g., fear, anxiety), and physiologic changes (e.g., fever, volume depletion). Changes in physical activity require the most frequent alterations, and rapid adjustments are initiated in seconds so that the arterial blood pressure may be increased to twice the normal value within 5 to 10 seconds. This short-term regulation is mediated by the sympathetic branch of the autonomic nervous system (the sympathetic nervous system [SNS]). Activation of the SNS influences both heart rate and SVR. The force of contraction is primarily a factor of the circulating volume (preload) and affects long-term regulation of arterial blood pressure.

Modifications in systemic blood pressure are made by activation of the SNS directly or indirectly through stimulation of the baroreceptor reflex. (Autoregulatory changes in pressure at a local level, at the tissues of body organs, are discussed in more detail in Chapter 15.) These SNS activities related to the distribution and pressure of blood are directed through the vasomotor center in the medulla of the brainstem while the lower centers of the brain monitor the body's internal and external environments. The vasomotor center is directly activated by such stimuli as fever or external stressors to evoke increased activity and elevate systemic arterial blood pressure.

The autonomic nervous system maintains a basal level of arteriolar smooth muscle tone through the SNS and provides heart rate control through a balance of SNS and parasympathetic nervous system (PSNS) activity. Stimulation of the SNS results in the increased release of the neurotransmitters epinephrine and norepinephrine. At the smooth muscle of the arterial system, these neurotransmitters bind to α_1 receptors to initiate vasoconstriction and an increase in SVR. Stimulation of the PSNS has almost no effect on most systemic vessels, other than venodilation in localized areas such as the face, producing a blush. Receptors within the brain (α_2) provide negative feedback regulation, decreasing the central release of epinephrine and norepinephrine in response to stimulation. In the heart, the binding of these neurotransmitters to β_1 receptors results in an increase in the rate of firing at the sinoatrial node, increasing the heart rate in response to increased demands. The PSNS is responsible for maintaining a slower heart rate during periods of rest.

Indirectly, the vasomotor center is stimulated by a decreased rate of discharge by *baroreceptors.* Pressure-sensitive receptors (baroreceptors) are found in the vessel walls of nearly all large arteries in the thorax and neck, but are particularly plentiful in the sinuses of the carotid arteries and in the arch of the aorta. Signals from the aorta travel through cranial nerve X, and those from the carotids are transmitted through cranial nerve IX; both terminate in the vasomotor center of the medulla. These specialized receptors are sensitive to changes in MAP. They transmit impulses continuously, altering their rate of discharge in response to changes in MAP. Their response to these changes is very brisk, especially when pressure changes occur rapidly, which makes them the perfect mechanism to respond to variations in body position and minimize the gravity-induced decreases in pressure in the upper body. A decrease in sensed pressure induces a decrease in action potential formation by the baroreceptors. This causes the vasomotor center to increase SNS outflow to the heart and arterial bed and to decrease PSNS stimulation to the heart. The net result is an increase in both heart rate and SVR, producing an increase in blood pressure. An increase in sensed pressure results in an increased rate of firing by the baroreceptors and a negative feedback response, lowering systemic arterial pressure. The responsiveness of the baroreceptor reflex declines with age; age-related stiffening of the arterial walls has been implicated along with contributions from pathologic conditions such as hypertension and diabetes mellitus, which are more common in the older population. The results of animal studies indicate that the overall effect of the baroreceptor reflex is a reduction of the minute-to-minute fluctuations in arterial blood pressure by 33% of what it would be without this mechanism. There is abundant evidence that within 1 to 2 days of exposure to chronic elevations of blood pressure, baroreceptors reset to the new level and the rate of discharge begins to decrease and then slowly returns to the norm despite an elevated baseline pressure. This finding suggests that the baroreceptor reflex may contribute to long-term blood pressure regulation through the SNS stimulation of the kidneys discussed in the next section.

Receptors in the carotid and aortic arterials respond to chemical signals of hypoxia (H^+ and CO_2 level elevations) that occur when arterial pressure declines. These chemoreceptors stimulate the medullary vasomotor center to increase SNS activity. However, this mechanism responds significantly only when systolic pressures decrease below 80 mm Hg, so blood pressure can be prevented from falling even lower.

Long-Term Regulation of Systemic Blood Pressure

The regulation of arterial blood pressure on a long-term basis, week after week and month after month, is accomplished through the interplay

of neural, hormonal, and renal interaction and is intimately connected with the body's fluid volume homeostasis. The balance of the intake of water and sodium with their excretion by the kidney remains the central feature of long-term blood pressure maintenance. Historically, the role of the renin–angiotensin–aldosterone system (RAAS) has been seen as the primary contributor to this process, and although it continues to be a major determinant, mechanisms involving the baroreceptor reflex and the vasomotor center in the brainstem as well as localized renal systems are receiving increased attention in research.

An increase in extracellular fluid (ECF) volume, because of increased intake or decreased excretion, results in an increase in cardiac output; when combined with the volume-induced increase in SVR, this results in an elevation in the arterial blood pressure (Fig. 16.4). Body tissues initiate their local autoregulation mechanisms, constricting arterioles to protect against high-flow damage, which further contributes to the overall arterial resistance in the body. Unless fluid intake or renal functions are abnormal, this increase in SVR will not result in a prolonged elevation in arterial pressure. The kidneys will respond quickly, increasing excretion of sodium and water and normalizing pressure within a matter of hours. This physiologic regulatory response may be disrupted if the renal vasculature is constricted, as occurs in hypertension.

Because sodium is not as rapidly eliminated by the kidney as water, elevations in sodium intake are more likely to elevate arterial pressure. Excess sodium also adds to the body's fluid volume by several mechanisms. Sodium increases the osmolality of the ECF and activates the central thirst center, causing an increase in water intake. The increased serum osmolality will be sensed by the hypothalamus and posterior pituitary, causing the release of antidiuretic hormone (ADH) into the bloodstream. Once ADH arrives in the renal vasculature, it binds to receptors in the collecting ducts, resulting in the enhanced reabsorption of water in order to decrease osmolality (Chapter 26).

The physiologic mechanisms of the RAAS are tightly controlled and interdependent (Fig. 16.5). Prorenin, the inactive form of renin, is synthesized and stored by specialized smooth muscle cells located in the afferent arterioles of the kidney situated immediately proximal to the glomeruli. Known as the *juxtaglomerular cells,* these cells are stimulated by a decrease in arterial pressure to enzymatically cleave the precursor and release the activated renin enzyme into the vascular bed of the kidney. Most of the renin travels into the general circulation, where it acts on a circulating plasma protein called *angiotensinogen,* resulting in the release of angiotensin I, a peptide possessing minimal vasoconstrictive capacity. Angiotensin I continues to be created by renin for about 30 to 60 minutes, until renin is removed from the body. While the blood carrying angiotensin I circulates through the pulmonary vessels, an enzyme produced by the vascular endothelium (angiotensin-converting enzyme [ACE]) comes in contact with angiotensin I, and two amino acids are fragmented from angiotensin I to produce angiotensin II. Inactivated in minutes by angiotensinases, continued production of angiotensin II maintains the profound effects it initiates. Angiotensin II is an extremely potent vasoconstrictor, primarily of the arterial bed, but also slightly affecting the venous system. The SVR is therefore increased, raising blood pressure. The vasoconstrictive response to angiotensin II requires about 20 minutes to reach maximal capacity, but is capable of elevating arterial pressure to 50% of normal after severe hemorrhage. The enhanced venous return attributable to the elevated SVR improves cardiac function by increasing myocardial fiber stretch, producing increased contractility and therefore stroke volume. Angiotensin II also is an intermediary for an additional means of raising blood pressure—increasing circulating volume to significantly increase venous return to the heart and therefore stroke volume. Angiotensin

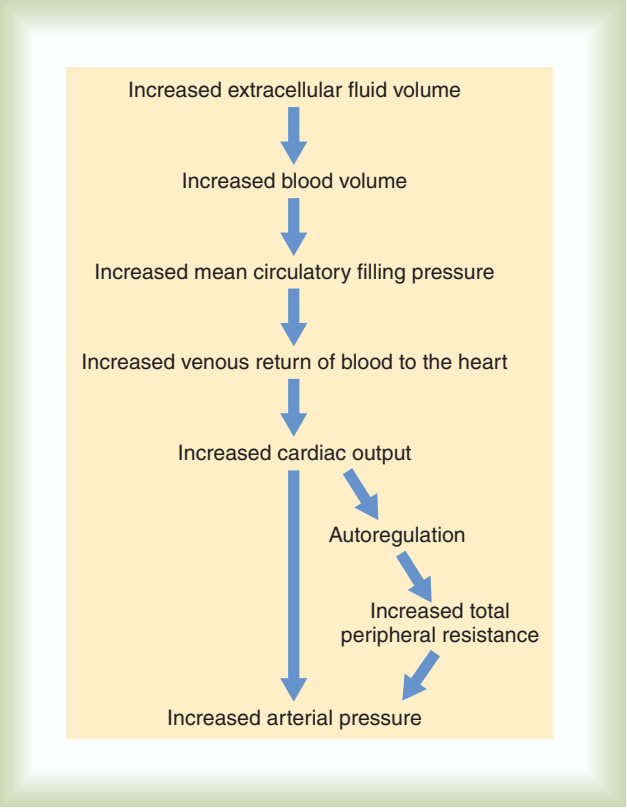


FIG 16.4 Mechanism by which an increase in extracellular fluid volume results in an increase in systemic arterial pressure. (From Hall JE: *Guyton and Hall textbook of medical physiology,* ed 13, Philadelphia, 2016, Saunders.)

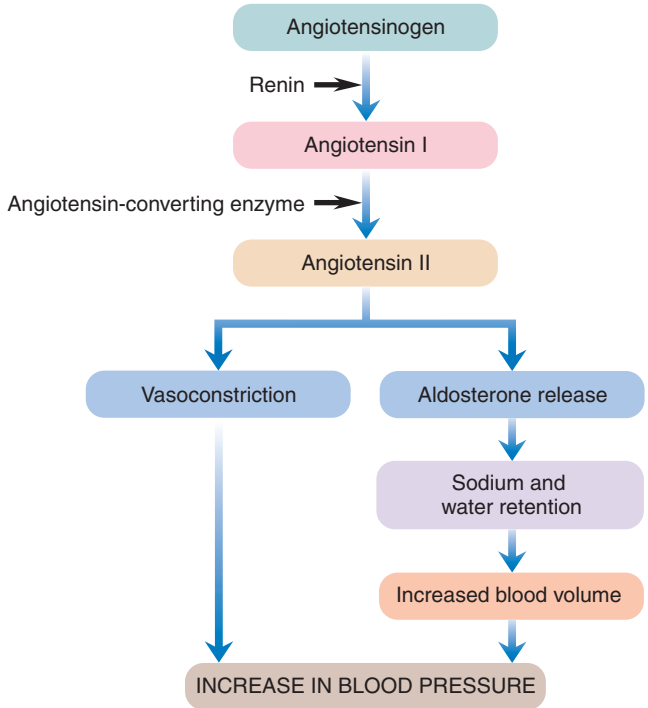


FIG 16.5 The renin–angiotensin–aldosterone system (RAAS) and its systemic effects.

II in the general circulation reaches the cortex of the adrenal glands, stimulating the release of the hormone aldosterone. Aldosterone circulates to the kidneys, where it binds to receptors in the renal tubules, causing the kidneys to reabsorb more sodium. Water follows the sodium back into the bloodstream. The result is an increase in blood volume and further elevation in blood pressure. Excessive amounts of angiotensin in the bloodstream have been found to effectively reset this mechanism of blood pressure control to a higher-than-normal level, potentially contributing to hypertension.

Some of the renin released by the juxtaglomerular cells exerts local effects within the kidney to elevate blood pressure. Renin receptors in the mesangium of the glomerulus and below the endothelial cells of the renal arteries are activated to enhance the conversion of angiotensinogen to angiotensin I. Angiotensinogen has been isolated in tubular and mesangial cells, and ACE is found both in vascular endothelium and in the epithelium of the tubular cells.

Other influences on long-term arterial blood pressure control include the activity of the SNS, levels of natriuretic peptides, and regulation of intrarenal mechanisms such as renal medullary endothelin production. Renin release is increased when neurotransmitters released by the sympathetic nervous system bind to β_1 receptors in the kidney. Additional local SNS effects include decreased glomerular filtration rate (GFR) as a result of renal arteriolar constriction and increased tubular reabsorption of sodium and water caused by increased quantities of angiotensin II and aldosterone. These effects contribute to the increased systemic blood pressure associated with severe prolonged stress. Increased SNS activity has been documented to be present in hypertension, and its role is confirmed because antihypertensive drugs that affect autonomic control of heart rate and SVR are so clearly effective in treatment. A number of natriuretic hormones play a role in arterial pressure through their effects on ECF volume regulation; most important of these is atrial natriuretic peptide (ANP). Increased volume in the atria of the heart triggers stretch receptors and stimulates the release of ANP into the bloodstream by cardiac muscle fibers. ANP causes the kidneys to increase water and sodium excretion by increasing GFR and decreasing sodium reabsorption so both sodium and water remain in the filtrate. This diuretic effect reduces circulating volume and therefore blood pressure. Endothelin-1 (ET-1) is a peptide produced in the renal medulla. ET-1 binds to receptors within the kidney, initiating an autocrine-induced vasodilatory response affecting renal perfusion, water and electrolyte movement, and release of renin. This makes ET-1 an important participant in normal systemic blood pressure control, and levels have been found to be abnormal in hypertension. Most likely, long-term blood pressure control is a reflection of the unified contributions of all the factors discussed here, and more are yet to be identified.

Normal Fluctuations in Systemic Blood Pressure

Many homeostatic mechanisms of the body undergo daily variations in their function governed by an area of the brain called the *suprachiasmatic nuclei*—the body's internal clock. Brain wave activity, cell regeneration, cortisol release, body temperature, heart rate, and blood pressure are only a few of the numerous circadian rhythms. In the case of blood pressure, it is known that it rises before awakening (morning surge), is highest in the middle of the morning, then begins to fall, and reaches its lowest level at night (nocturnal dip). In their review of the available research, Peixoto and White found these basic fluctuations to be primarily determined by internal neural and hormonal regulation, as well as by external environmental factors such as sodium intake and physical activity. Additional factors known to affect the normal rhythmic changes in blood pressure include lifestyle influences such as alcohol consumption and cigarette smoking, as well as cognitive activity and emotional state. Development of diabetic nephropathy, sudden cardiac death, myocardial infarction, and stroke are associated with circadian elevations in BP.

KEY POINTS

- Blood pressure is regulated on a short-term basis through the interaction of the carotid and aortic baroreceptors, the vasomotor center in the brainstem, and the activation of the sympathetic nervous system (SNS) and inhibition of the parasympathetic nervous system (PSNS) influences on the heart and smooth muscle in the arterioles. Short-term regulation primarily involves heart rate and systemic vascular resistance (SVR).
- Regulation of blood pressure on a long-term basis is complex, involving the influence of the nervous system, release of hormones, and responses of the kidneys to pressure changes. The vasomotor center and activation of α_1 receptors in the smooth muscle of the arterioles and the β_1 receptors of the heart continue to be involved when pressure changes are sensed by the baroreceptors.
- Secretion of antidiuretic hormone (ADH) in response to osmolality and of aldosterone from the activation of the renin–angiotensin–aldosterone system (RAAS) affects fluid balance, whereas angiotensin II produces an increase in SVR. Natriuretic peptides and intrarenal mechanisms contribute to the process of long-term blood pressure management. Long-term regulation involves all of the blood pressure variables: heart rate, stroke volume, and SVR.
- Normal fluctuations of blood pressure occur in a cyclic pattern attributable to changes in the body's internal and external environments.

HYPERTENSION

The current and projected global prevalence of hypertension is stunning. Hypertension is the most common primary diagnosis in the United States. About 32% or 80 million adults in the United States have high blood pressure. The prevalence of high blood pressure remains higher among non-Hispanic black adults (44.9% & 46.1%; men and women, respectively) compared with non-Hispanic white (32.9% & 30.1%; men and women respectively) and Mexican American adults (29.6% & 29.9%; men and women, respectively). Global estimates suggest more than 40% of adults 25 years of age or older are diagnosed with hypertension, affecting more than 1 billion people. Changes in the standard of living of those in developing countries mirror the trends in economically developed ones: increasing obesity, diabetes, and sedentary lifestyles. Increased consumption of alcohol, cigarette smoking, and diets deficient in fruits and vegetables contribute to the problem of escalating hypertension worldwide. Hypertension will affect nearly half of the adult population in the majority of the world. India and Asia have the lowest current and projected prevalence, whereas the former socialist republics, sub-Saharan Africa, the Caribbean, and Latin America have the highest rates. Given the risks to health with blood pressure elevations the future impact of hypertension is profound. Hypertension increases morbidity and mortality associated with heart disease, kidney disease, peripheral vascular disease, and stroke. It is responsible for a worldwide annual death rate of 7.6 million, and it is the most common risk factor for cardiovascular disease worldwide. An understanding of the types and causes of hypertension and the interventions associated with its treatment is essential to having an impact on the current and future effects of this disease.

Definition and Classification

The standard for the definition and classification of hypertension in adults continues to be drawn from the *Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (JNC 7),* published in 2003. For those individuals age 18 years and older, normal blood pressure is defined as <120 mm Hg

systolic and <80 mm Hg diastolic; stage 1 hypertension begins at a systolic pressure of 140 mm Hg or a diastolic pressure of 90 mm Hg (see Table 16.2). The range of pressures between normal blood pressure and stage 1 hypertension has been identified as *prehypertension* as part of efforts to initiate interventions early enough to prevent or at least slow the progression of the disease process. In 2014 the Eighth Joint National Committee released treatment goals based on age, but did not change the JNC 7 criteria for defining hypertension (see Table 16.2). These values differ from those established by the World Health Organization, International Society of Hypertension, and European Society of Hypertension/European Society of Cardiology in that those used in the United States are more conservative, identifying both normal and elevated levels at lower values. Standards for children and adolescents also have been established (see Table 16.1). Differing etiologies and risk factors have led to the differentiation of two major types of hypertension: primary and secondary.

Primary Hypertension

Primary hypertension, also called *essential hypertension,* does not have a clearly identifiable known etiology and is therefore an idiopathic disorder. This differentiates primary from secondary hypertension, in which blood pressure elevation occurs secondarily to another, identifiable cause. Primary hypertension is by far the most common form of the disease, representing somewhere between 90% and 95% of the known cases. Early diagnosis and intervention for adults with hypertension has been a major focus of health care for many decades; the more recent escalating incidence in children has generated alarm. Primary hypertension is increasing in prevalence among children and adolescents and is associated with positive family history of hypertension, obesity, and lifestyle factors. The prevalence of hypertension in children and adolescents is approximately 3.5%. Primary hypertension is rare before children reach the age of 10 years. Most of the hypertension diagnosed in preadolescents has a secondary etiology; by adolescence, 85% to 95% of the cases are primary hypertension.

Subtypes

Primary hypertension in adults less than 60 years of age has one of several presentations: isolated systolic hypertension (ISH) in which the systolic blood pressure is ≥140 mm Hg and the diastolic pressure remains <90 mm Hg; isolated diastolic hypertension (IDH) in which the diastolic pressure is ≥90 mm Hg with a systolic pressure of <140 mm Hg; and the combination of systolic and diastolic hypertension (SDH) occurring when both systolic and diastolic pressures exceed prehypertension values. The differing subtypes are more prevalent in specific populations, and researchers increasingly focus on subtypes in long-term outcome predictions and interventions. The evidence suggests that systolic blood pressure is the major risk for subsequent cardiovascular disease.

Risk Factors

In Western populations, there is a 90% lifetime risk for the development of hypertension. Many of the risk factors for hypertension have been known for decades and because so many are modifiable by lifestyle changes, targeted interventions are urged to address them. Other factors remain nonmodifiable, yet predictive of the development of hypertension. Ample data indicate that primary hypertension arises as a consequence of the interplay of several genes and environmental factors. Hypertension risk factors are listed in Table 16.5.

Increasing age is a nonmodifiable risk and an independent risk factor for hypertension beginning at midadulthood. Normal aging produces a rising systolic pressure over the course of a lifetime, whereas diastolic pressure increases for approximately 50 years, levels off during the sixth decade, and remains stable or declines thereafter. ISH is the dominant subtype of hypertension in those older than age 55. The level of the systolic pressure, MAP, and the difference between systolic and diastolic pressures (pulse pressure), among other factors, are used to guide pharmacologic interventions. Given this normal age-related development of hypertension, most early and subsequent data related to the increased risk of mortality and morbidity are based on this population, and systolic pressure elevation clearly affects risk more than diastolic, as noted in the earliest and most famous cardiovascular research in the Framingham Study.

TABLE 16.5 Risk Factors for the Development of Primary Hypertension

Nonmodifiable Risk Factors	Modifiable Risk Factors
Increasing age	Obesity
Family history	Sedentary lifestyle
	Metabolic syndrome
	Dietary factors
	• Increased fat intake
	• Increased sodium intake
	• Inadequate potassium intake
	• Inadequate calcium intake
	Tobacco use
	Laboratory data
	• Elevated blood glucose
	• Elevated total cholesterol
	• Elevated triglycerides
	• Decreased high-density lipids (HDL)
	• Elevated low-density lipids (LDL)

Age is not a risk factor for hypertension in childhood or adolescence. Hypertension does occur in these age groups, however, and the distribution of subtypes and the proposed bases and prognosis are worth noting. Determination of hypertension is based on the normal expectations for the child's age, gender, and height (see Table 16.1). Although IDH is more common among younger adults, ISH can occur. ISH in adolescents and young adults (<45 years of age) has been attributed to the increased elasticity of their arteries in the face of rapid growth; this produces an increase in brachial systolic pressure, although aortic pressure is unchanged. Others have found an increase in stroke volume with or without aortic stiffening to be the basis of ISH in this age group. IDH often seems to develop in prehypertensive young adults, and the prognosis remains open to debate. One study found that IDH does not appear to predict the development of ISH but is a strong predictor of the later development of SDH. The report of a review of the literature indicated that below the age of 50, diastolic pressure was a greater predictor of coronary artery disease (CAD), whereas CAD risk in those age 60 and older was greater with elevated systolic pressure. In children, hypertension is a risk both for adult hypertension and for subsequent development of adult cardiovascular disease.

Another significant nonmodifiable risk factor is ethnicity, which combines race with genetics. Adult African Americans have the highest risk, but there is controversy about this finding for the pediatric population. A concrete reason for this finding in adults continues to elude researchers, although increased salt sensitivity seems most likely. The number of individual genes and their signaling pathways and organizational arrangements that affect the control of blood pressure are vast and beyond the scope of this text. Despite the identification of the genetic role in hypertension, specifying the mechanisms themselves is a challenge. Genetics may be responsible for low renin levels and salt

sensitivity, heightened responses to angiotensin II, altered amounts or responses to local tissue factors such as endothelin and nitric oxide, and any number of mechanisms accounting for primary hypertension that have been proposed. At the same time, none of the currently identified genetic disorders have been demonstrated to be accountable for a noteworthy proportion of hypertension in the general population, either as individual genes or as several genes working in concert. A family history of hypertension is a risk factor in both adults and children, although these mechanisms are unclear.

Modifiable risk factors are often called *lifestyle factors* in acknowledgment of the role of individual choice in both their development and their control. Both weight gain and obesity are significant risk factors for all subtypes of primary hypertension at all ages. Obesity has reached epidemic proportions and represents between 65% and 75% of the overall risk for the development of hypertension. Diet and activity levels contribute to the development and continuation of obesity in all age groups. Diets high in fat and sodium and low in potassium and in fruits have been found to increase the risk of developing hypertension. Obesity not only increases the risk of hypertension, but also is a risk factor for hyperlipidemia, salt sensitivity, and insulin resistance. Known as *metabolic syndrome* and characterized by elevated circulating insulin and lipid levels, hypertension, and obesity, this condition was previously only identified as a risk factor for hypertension in adults but is now becoming more common in children. It has been estimated that the prevalence of metabolic syndrome is 7% in adolescents at risk for becoming overweight, 29% in overweight adolescents, and 50% in severely obese adolescents. Elevated blood glucose levels, diabetes mellitus, and elevated total cholesterol level, as well as smoking and excessive alcohol intake, are all implicated as risk factors for hypertension.

Recently the apparent increased incidence of childhood and adolescent hypertension has spawned research into predictors of the development of hypertension from the maternally provided intrauterine environment through the childhood and adolescent period. Maternal smoking, pregnancy-induced hypertension, and maternal dietary habits have been shown to influence the later development of hypertension. Low birth weight, followed later by rapid growth in both height and weight, seems to be more common in the history findings of patients with hypertension. Lower socioeconomic level of the mother and inadequate dietary calcium intake during pregnancy appear to increase the risk for later development of hypertension, whereas breast feeding seems to act as a protective factor against hypertension.

Outcomes

End-organ damage. The great concern for the prevention, early identification, and treatment of hypertension is because of the harm it may cause in body tissues and organs and the resulting significant morbidity and mortality (Fig. 16.6). This end-organ damage is a function of both the stage of hypertension and its duration. Unfortunately, early hypertension causes no overt clinical manifestations, and individuals may have considerable end-organ damage before the diagnosis is made. This has earned hypertension the moniker of the "silent killer" and explains the rationale for screening programs to encourage early diagnosis. Hypertension is an important independent risk factor for the development of renal failure, stroke, and CAD. CAD and hypertension heighten the risk of angina, myocardial infarction, and heart failure. As the systolic and diastolic pressures rise from normal levels, mortality from ischemic heart disease and stroke also increases linearly and progressively.

Cardiovascular disease is the most commonly recognized outcome of hypertension. For those older than age 50 years, systolic hypertension presents a far greater risk for the development of cardiovascular disease than does an elevated diastolic blood pressure. Risk for the occurrence

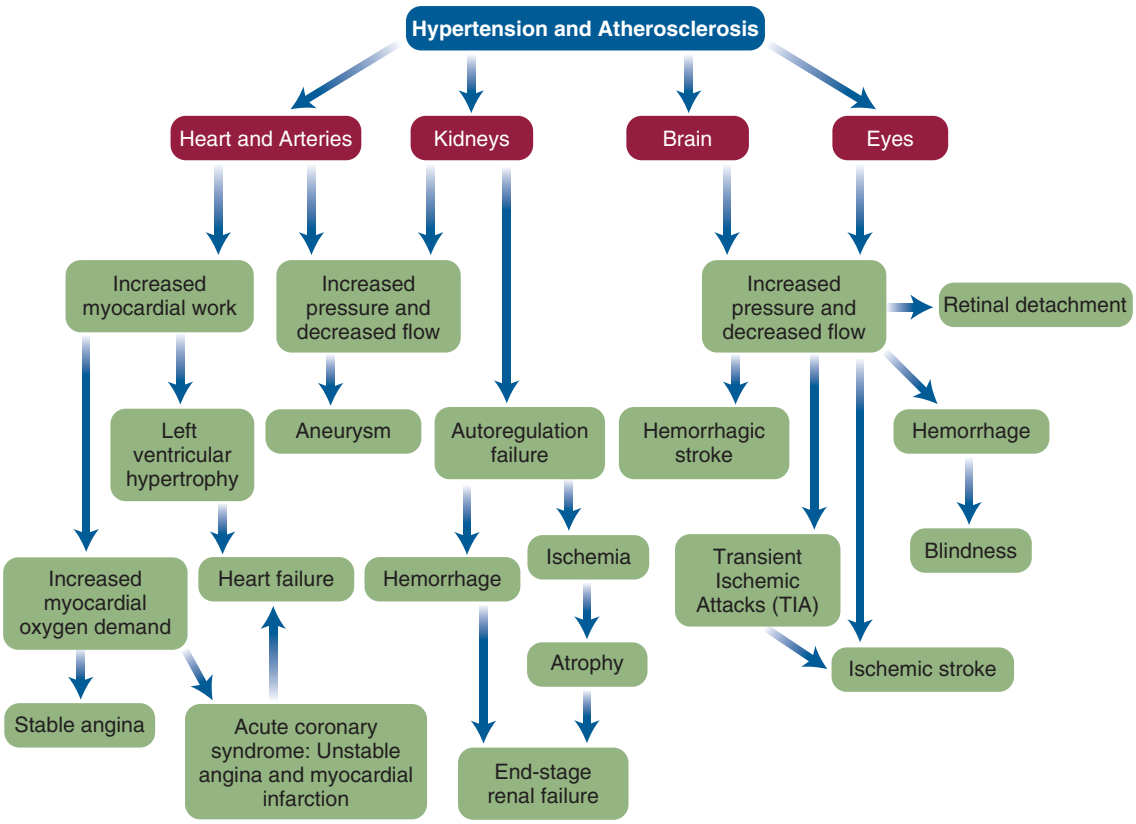


FIG 16.6 Effects of chronic hypertension and atherosclerosis on target end organs.

of cardiovascular disease doubles with each incremental increase of 20/10 mm Hg in BP.

Hypertension itself is directly harmful to the arterial system, but it also acts in concert with the other risk factors associated with the development and acceleration of atherosclerosis. Atherosclerosis is the underlying pathophysiologic basis of coronary artery disease (see Chapter 18). Evidence of atherosclerosis has been found in adolescents and very young children. The increased tension that high blood pressure generates on the walls of arteries precipitates an increase in the accumulation of collagen as well as reduction, fragmentation, and breakage of elastin fibers. An ongoing low level of inflammation occurs in arteries exposed to hypertension, and combined with the dyslipidemia commonly seen, the development of atherosclerotic plaques is escalated. CAD predisposes to stable angina and the acute coronary syndrome of unstable angina and myocardial infarction (see Chapter 18).

Hypertension reflects an elevation in SVR; rising afterload increases myocardial oxygen demand and overall cardiac workload. In an effort to compensate for this increased effort, the left ventricle hypertrophies. The development of left ventricular hypertrophy also has been noted in children and adolescents. The CAD typically found in association with hypertension limits the supply of oxygen to the heart, and this combination of increased demand and decreased supply predisposes the heart to ischemia. Ischemia may result in stable or unstable angina or myocardial infarction. Myocardial infarction and left ventricular hypertrophy increase the risk for the development of heart failure. Patients may seek health care intervention because of these conditions so that the presence of hypertension is discovered only secondarily.

The atherosclerotic process described previously with coronary artery disease (Chapter 18) is likely to be the basis for the damage to the microcirculation of the kidneys that develops with chronic hypertension. Within a proscribed MAP, healthy kidneys are able to autoregulate blood flow delivered to the glomerulus, but with prolonged or severe hypertension this regulatory ability is lost and glomerular damage ensues. Damage to the glomerulus allows large molecules not normally filtered out of the bloodstream to appear in the urine. The presence of microalbuminuria (proteinuria) is reflective of increased glomerular permeability and an early indicator of hypertensive renal injury. At this point, the patient is usually asymptomatic, but if interventions for blood pressure control are not initiated, renal impairment progresses, culminating in end-stage renal disease, which requires long-term renal dialysis or transplantation.

Identifiable damage to the kidneys is often preceded by changes in the microcirculation of the retina of the eye. Atherosclerosis also contributes to the retinal injury produced by hypertension. The result may be retinal detachment or hemorrhage, which can cause blindness.

Hypertension and the accelerated development of atherosclerosis affect arteries of all sizes throughout the body. Decreased flow or rupture of weakened blood vessels within the brain results in strokes. Ischemic strokes are associated with atherosclerosis, whereas hypertension is the major risk factor for hemorrhagic strokes. This type of stroke results in high morbidity and mortality. Hypertension is also the primary risk factor for the development and rupture of aortic aneurysms. The peripheral arteries of the lower extremities are common targets of atherosclerosis, and the resulting peripheral vascular arterial disease is the source of significant impairment of independence and mobility and potential amputation in the elderly.

Treatment Interventions

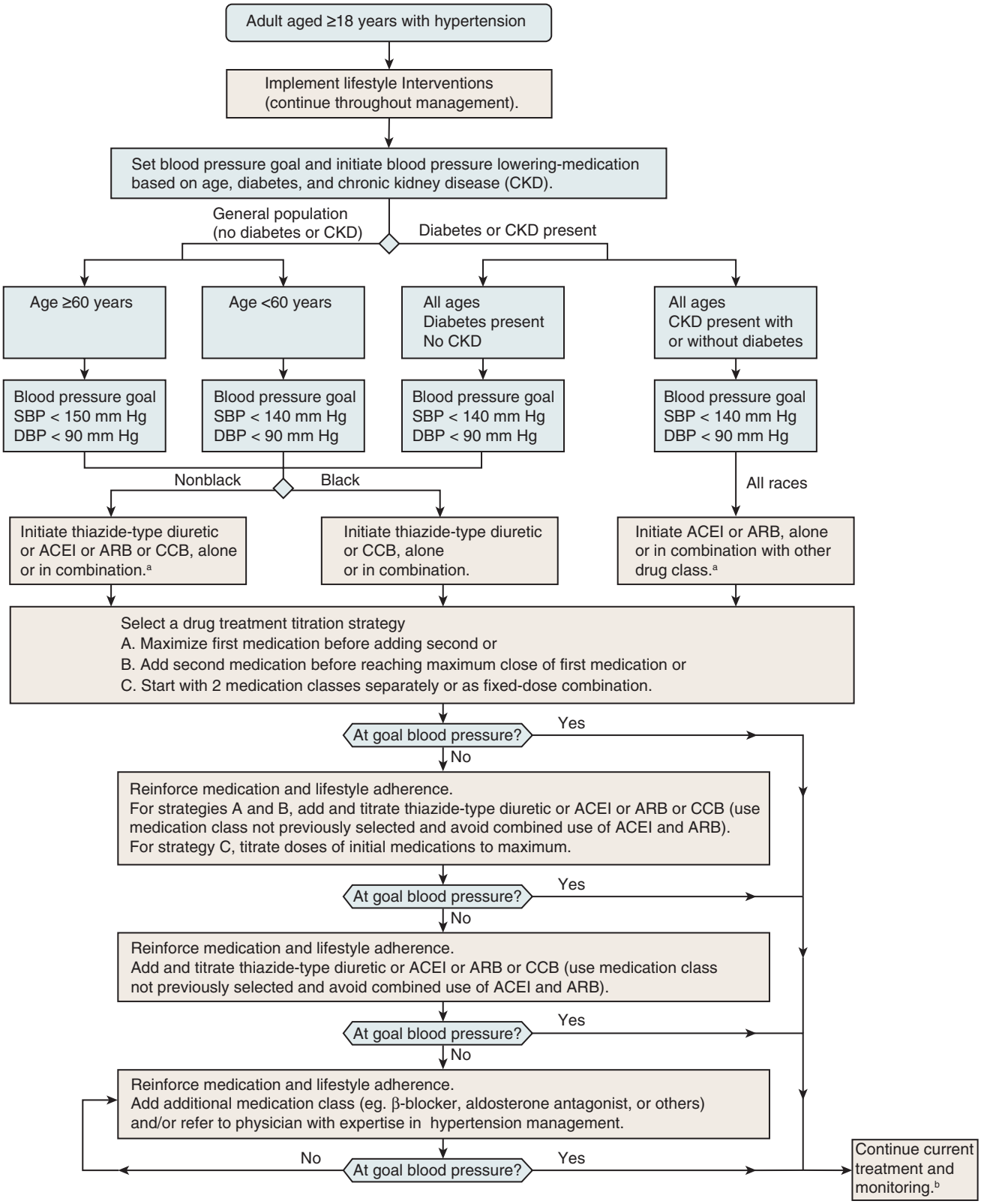
Effective treatment of hypertension results in decreased morbidity and mortality associated with cardiovascular, cerebrovascular, and renal disease. However, the goal of normal systolic and diastolic pressures is not attained for many patients with hypertension. The problem appears to be even greater in Europe, where only 5% to 12% achieve control compared with 27% in the United States. Cumulative data from clinical drug trials indicate the risk of stroke can be reduced 35% to 40% by decreasing blood pressure, myocardial infarction 20% to 25%, and heart failure by more than 50%. Approaches to treatment are affected by several factors, including the patient's age, stage of hypertension, identified risk factors, concomitant disorders, ethnicity, and medication history. Interventions fall into two categories: lifestyle alterations and pharmacologic interventions. Whatever the chosen therapy, it is important to get to the goal or target blood pressure for the individual. For those under 60 years of age, the goal is a systolic BP under 140 mm Hg and a diastolic BP under 90 mm Hg. One of the major changes of JNC 8 was to alter the target systolic BP to less than 150 mm Hg (instead of 140 mm Hg) for those 60 years old and older. The target diastolic BP was not changed and remains at under 90 mm Hg. JNC 8 also eliminated the various target BPs that were suggested in JNC 7 for those with conditions such as diabetes and heart disease. The overall approach to treatment is detailed in Fig. 16.7. Although an in-depth discussion of treatments is beyond the scope of this book, they are summarized in the following text.

Lifestyle alterations assume special importance because addressing modifiable risk factors has a documented effect in preventing hypertension from developing, as well as treating it in adults and children. Primary prevention of hypertension could have a profound influence on the morbidity and mortality associated with end-organ damage throughout the world and includes lifestyle changes and effective screening procedures to facilitate early diagnosis. These lifestyle changes are listed in Table 16.6. Weight loss is clearly an important intervention, with substantial evidence that it reduces cardiovascular mortality. The efficacy of exercise in blood pressure control is also well substantiated by research. Brisk

TABLE 16.6 Lifestyle Modifications to Prevent and Treat Primary Hypertension in Adults*

Modification	Recommendation	Range of Sbp Reduction†
Weight reduction	Attain and maintain BMI of 18.5–24.9 kg/m^2	5–20 mm Hg/10 kg
DASH diet	High in fruits and vegetables and low-fat dairy products with decreased total and saturated fat	8–14 mm Hg
Decreased sodium intake	No more than 100 mmol/day (2.4 gm sodium or 6 gm sodium chloride)	2–8 mm Hg
Exercise plan	Regular aerobic activity for at least 30 min/day most days of week	4–9 mm Hg
Moderate intake of alcohol	≤2 drinks/day for men ≤1 drink/day for women	2–4 mm Hg

Modified from Ozemek C, Phillips SA, Popovic D, et al: Nonpharmacologic management of hypertension: A multidisciplinary approach. *Curr Opin Cardiol,* March 17, 2017. [Epub ahead of print.]
BMI, Body mass index; *DASH,* Dietary Approaches to Stop Hypertension.
*Smoking cessation increases the overall reduction in cardiovascular risk.
†Results vary based on individual response, amount, and time of modification accomplished.



SBP indicates systolic blood pressure; DBP, diastolic blood pressure; ACEI, angiotensin-converting enzyme; ARB, angiotensin receptor blocker; and CCB, calcium channel blocker.

[a] ACEIs and ARBs should not be used in combination.
[b] If blood pressure fails to be maintained at goal, reenter the algorithm where appropriate based on the current individual therapeutic plan.

FIG 16.7 Treatment recommendations for primary hypertension. (From James PA, Oparil S, Carter BL, et al: 2014 Evidence-based guideline for the management of high blood pressure in adults: Report from the panel members appointed to the eighth joint national committee (JNC 8). *JAMA, 311*(5):507-520, 2014.) *ACEI,* Angiotensin-converting enzyme inhibitor; *ARB,* angiotensin receptor blocker; *BB,* β-blocker; *BP,* blood pressure; *CCB,* calcium channel blocker; *DBP,* diastolic blood pressure; *SBP,* systolic blood pressure.

exercise of at least 30 minutes most days of the week plus the acceptance of the Dietary Approaches to Stop Hypertension (DASH) diet address a number of hypertensive risk factors. Because they have been found to augment drug efficacy, these lifestyle adjustments also are included for those who require medication interventions for primary hypertension; however, it is unlikely that lifestyle interventions alone will be sufficient for those with stage 1 hypertension.

Drug therapy for hypertension addresses one or more of the variables responsible for blood pressure: heart rate, SVR, and stroke volume, which is primarily a function of the volume of blood returned to the heart during diastole. This is reflected in how the classifications of oral medications used in the treatment of hypertension are listed in Box 16.1. Combination drugs, taking advantage of the effects of more than one classification without increasing the total number of medications a patient is taking, are becoming increasingly popular. Because there is considerable variation in individual response to antihypertensive drug therapy, long-term monitoring is essential, and alterations in treatment may be necessary.

Secondary Hypertension

When hypertension is found to have a specific identifiable cause, it is termed *secondary hypertension*. The cause may be a specific pathology or condition that results in hypertension, or the development of high blood pressure may be the result of the ingestion of certain drugs, foods, or chemicals. Conditions associated with secondary hypertension are listed in Box 16.2. Some common substances that increase blood pressure are shown in Box 16.3.

In infants and preschool children, hypertension is usually of a secondary etiology, and primary hypertension is rare. In a study of 220 hypertensive children, 85% of the cases were found to be of a secondary etiology. The four variables independently associated with primary hypertension were absence of signs and symptoms, normal serum creatinine level, family history of hypertension, and elevated body weight. In the diagnostic assessment of adults, secondary etiologies should be ruled out, but from the age of 18 years, primary hypertension is far more common. Interventions for secondary hypertension are directed at removing the cause, if possible. Drug therapy may be indicated, using the same agents previously discussed in the treatment of primary hypertension. As with primary hypertension, both the severity of the elevation and its duration must be considered because they heighten the risks for end-organ damage. The most common etiologies of secondary hypertension are discussed next.

Renal disease may be the result of a disease process either involving the parenchyma of the kidney or involving its vascular system. Hypertension is a risk factor for the development of renal failure, but it can also develop secondary to renal pathologies. In adults with renal disease, hypertension is common and develops early, primarily as a result of heightened SNS activity. In children one of the most common causes of hypertension is renal disorders; however, by the age of 12 to 18, the major cause becomes idiopathic primary hypertension. As renal disease progresses, the kidneys' ability to excrete sodium effectively is lost and the RAAS, as well as the SNS, are inappropriately activated. Renal artery stenosis should be considered in the diagnostic evaluation of new-onset hypertension in patients younger than 30 or older than 55 years, and an abdominal bruit is often found on auscultation.

Another common cause of hypertension in children younger than age 6, along with renal disease, is coarctation of the aorta. Without treatment, median life expectancy is only 31 years, with death from

BOX 16.1 Drug Classifications Used to Treat Hypertension and the Variables They Affect

Reduce Stroke Volume
Thiazide diuretics
Loop diuretics
Potassium-sparing diuretics
Aldosterone receptor blockers
Angiotensin (ACE) inhibitors
Angiotensin II receptor blockers
Venodilators

Reduce Systemic Vascular Resistance
Combination α_1- and β-blockers
Angiotensin-converting enzyme (ACE) inhibitors
Angiotensin II receptor blockers
Calcium channel blockers
α_1-Blockers
Central α_2 agonists
Direct-acting vasodilators (arterial)

Decrease Heart Rate
β-Blockers
Combination α_1- and β-blockers

BOX 16.2 Common Pathologic Causes of Secondary Hypertension in Children and Adults

Renal (Parenchymal or Vascular)
Renal artery stenosis
Renal failure* (end-stage renal failure attributable to any etiology; acute renal failure)
Polycystic kidney disease
Glomerulonephritis*
Hypertensive nephrosclerosis

Cardiovascular
Coarctation of the aorta*

Tumors
Pheochromocytoma*
Neuroblastoma*
Wilms tumor*
Adrenal adenocarcinoma*

Endocrine
Hyperthyroidism*
Cushing disease*
Congenital adrenal hyperplasia*
Primary hyperaldosteronism*

Neurologic
Guillain-Barré syndrome*
Increased intracranial pressure*

Other
Systemic arteritis (e.g., Henoch–Schönlein purpura)
Sleep apnea*

*Also seen in children.

BOX 16.3 Substances Known to Contribute to High Blood Pressure

Over-the-Counter Drugs, Prescription Drugs, and Illicit Drugs

Sympathomimetic agents (e.g., decongestants, amphetamines)
Glucocorticoids
Cocaine
Calcineurin inhibitors (e.g., cyclosporine, tacrolimus)
Oral contraceptives, especially if high in estrogen
Nonsteroidal antiinflammatory drugs
Erythropoietin
Antidepressants
Phenylpropanolamine analogs (e.g., ma huang, "herbal ecstasy")
Nicotine (and withdrawal)
Anabolic steroids
Narcotic withdrawal
Ergotamine
St. John's wort

Foods

Foods containing tryptophan or tyramine
- Chicken liver
- Pickled herring
- Yeast extract
- Lima beans
- Aged cheeses
- Beer and wine

Caffeine
Sodium chloride
Alcohol
Licorice

Chemical Elements

Lead
Mercury
Lithium salts
Thallium and other heavy metals

hypertension-related systemic effects. It is for this reason that palpation of peripheral pulses and measurement of blood pressure in both arms are recommended both in routine pediatric physical examinations and in physical examinations in those whom hypertension is present. Early diagnosis facilitates early surgical intervention, reducing both premature mortality and end-organ damage. Hypertension persists after surgical repair in 20% to 30% of patients, but the pathophysiologic basis of this finding has yet to be determined. Unfortunately, this means that these individuals remain at risk for the long-term effects and increased mortality associated with hypertension.

Hypertension arises in 5% to 12% of all pregnancies. Hypertension during pregnancy is of concern because of increased risk of maternal, fetal, and neonatal morbidity and mortality. Preterm labor, abruptio placentae, disseminated intravascular coagulation, hemorrhagic stroke, liver failure, and acute renal failure are all potential outcomes of hypertension during pregnancy. When hypertension is diagnosed during pregnancy, it is classified into one of four categories: chronic hypertension (preexisting), preeclampsia, chronic hypertension with superimposed preeclampsia, or gestational hypertension. Pharmacologic interventions are used cautiously, and lifestyle interventions such as limiting salt intake and avoiding the use of alcohol or tobacco may be sufficient.

Obstructive sleep apnea (OSA) is closely associated with obesity; it is found in 2% to 4% of adults, and hypertension is present in 45% to 60% of those diagnosed with OSA. Historically, there has been debate as to whether OSA itself was an etiologic factor in hypertension or whether obesity simply increased the risk of both. Researchers now take the position that the potential causality between hypertension and OSA entails both an independent role of OSA in chronic blood pressure elevation and the obesity–hypertension linkage. Certain molecular mechanisms, including increased vasomotor activity mediated by angiotensin II, endothelin, and nitric oxide, may occur in both. The severity of OSA has a direct relationship to the level of blood pressure elevation, and when untreated, mortality and morbidity resulting from cardiovascular pathologies is increased. A milder form of OSA is found in children, and evidence is increasing that it, too, is associated with discernible cardiovascular abnormalities, including hypertension, decreased arterial distensibility, and left ventricular hypertrophy.

Pheochromocytoma is a catecholamine-secreting tumor of the adrenal medulla that generates hypertension on either a short-term or a long-term basis. The condition is rare, although well recognized; it can result in angina, myocardial infarction, acute heart failure, dilated cardiomyopathy, cerebral ischemia or hemorrhagic stroke, and cardiac dysrhythmias. Treatment involves control of blood pressure pharmacologically and then surgical removal of the tumor.

Hypertension is a predictable finding in primary hyperaldosteronism. Most frequently it is caused by a hypersecreting benign adenoma of the adrenal cortex or either unilateral or bilateral idiopathic adrenal hyperplasia. Although evidence exists that aldosterone is produced by other body tissues, hormone from the adrenal gland represents by far the majority of circulating aldosterone. The ratio of aldosterone to renin may be genetically influenced, but this has not been consistently documented. For decades, hyperaldosteronism was thought to be a rare cause of hypertension; it is now known to be the most common form of secondary hypertension, responsible for at least 12% of all cases, and it is believed that this number would be higher with improved screening. Screening for hyperaldosteronism is recommended for hypertensive patients with decreased potassium levels or those found to be refractive to three or more antihypertensive agents. Diagnosis requires measurement of serum aldosterone and renin levels. Surgical removal of the involved adrenal gland results in a cure for 30% to 60% of cases and improved blood pressure levels in the remainder. Removal of one or both adrenal glands in bilateral disease rarely appears curative, so bilateral disease is treated medically with agents that block aldosterone's binding sites in the kidney.

Hypertensive Emergencies and Urgency

Acute rises in blood pressure are identified by several names, complicating discussion of the condition. *Hypertensive crisis (HTN-C)* was the term introduced to replace the initial term *malignant hypertension,* which originated as early as 1914. Eighty million Americans have been diagnosed with hypertension, and about 7% of these will experience a hypertensive crisis during their lifetime. The 1 year mortality of untreated HTN-C is 79% with a median survivial of 10 months. Most cases of HTN-C seem to be caused by secondary hypertension or poorly or uncontrolled primary hypertension. Some other notable etiologies for hypertensive crisis include autonomic dysfunction, as is seen in Guillain-Barré syndrome, and autonomic dysreflexia, which can manifest in patients with high spinal cord injuries as well as in patients discontinuing certain drugs, such as β-blockers. The contemporary use of *hypertensive crisis* is associated with two differentiated sub groups: *hypertensive emergency* and *hypertensive urgency.* In both cases, the diastolic blood pressure is usually >120 mm Hg. Hypertensive emergencies are situations characterized by a sudden increase in either or both systolic and diastolic pressures

accompanied by evidence of acute end-organ damage. These same references use the term *hypertensive urgency* to describe similar blood pressure elevations, but without the end-organ damage. The differentiation is necessary because it is the presence of end-organ damage and not the level of the blood pressure that usually determines the treatment.

Hypertensive emergencies can occur in the previously undiagnosed patient or the patient with chronic hypertension; these emergencies are twice as common in males as in females, and elderly African Americans have the highest incidence. Hypertensive emergencies can occur at any age and are estimated to be responsible for more than 25% of emergency department visits. Of all the end-organ damage with which hypertensive emergencies are associated, abnormalities of the central nervous system are the most frequent. These include ischemic stroke, encephalopathy, and subarachnoid or intracerebral hemorrhages. Acute heart failure, including acute pulmonary edema, myocardial infarction, and aortic dissection, is among the common cardiovascular complications, and retinopathy is a frequent finding. From a pathophysiologic standpoint, hypertensive emergencies are the result of multiple factors, including an abrupt release of catecholamines, mechanical stress producing endothelial damage, inappropriate activation of the RAAS, and oxidative stress. These changes overwhelm the normal autoregulatory mechanisms and result in a sudden and significant increase in systemic vascular resistance, initiating an inflammatory response. Because of the evidence of end-organ damage, recommendations are for the rapid but controlled reduction of blood pressure using primarily parenteral antihypertensive agents closely monitored in a critical care setting. The standard goal is to decrease the diastolic blood pressure to 100 to 110 mm Hg or about a 25% decrease in the MAP.

Hypertensive urgency is actually more common than hypertensive emergency. The approach to treatment of hypertensive urgency, when end-organ damage is not evident, is quite different. Once measurement error is eliminated, other sources of rapid-onset reactive hypertension should be ruled out. These may include anxiety, pain, abrupt withdrawal of alcohol or antihypertensive medications, postoperative hypertension especially after cardiac and vascular surgery, and full bladder. In some of these cases, interventions other than antihypertensive drugs are indicated. In patients with hypertensive urgency, rapidly decreasing blood pressure has been associated with a substantial mortality. Blood pressure in these patients is usually brought under control over 24 to 48 hours through the use of oral medications, although there are indications for more aggressive intervention with parenteral agents.

KEY POINTS

- Primary hypertension has no identifiable etiology, but risk factors include age; dietary factors, including excess sodium and obesity; ethnicity and family history; sedentary lifestyle; and tobacco use.
- In adults, a normal blood pressure is <120 mm Hg systolic and <80 mm Hg diastolic pressure. Stage 1 hypertension begins with a systolic pressure of 140 mm Hg or a diastolic pressure of 90 mm Hg. Between these values, the individual is said to have prehypertension, and interventions related to lifestyle changes should be initiated.
- Treatment of primary hypertension includes lifestyle modifications and drug therapy. Lifestyle changes address the modifiable risk factors. Drug therapy targets one or more of the variables of blood pressure: heart rate, stroke volume, and systemic vascular resistance (SVR).
- In secondary hypertension, the elevated blood pressure is the result of identifiable pathologic conditions or certain drugs or foods. It is less common in adults, but is the major cause of hypertension in children. The underlying cause must be treated; drug interventions may also be necessary.
- Hypertension is usually asymptomatic until there is significant damage to vulnerable organs or tissues. This process is augmented by atherosclerosis in the coronary, renal, and cerebral arteries. Ultimately, hypertension increases the risk of stroke, angina, myocardial infarction, heart failure, renal failure, and blindness caused by retinopathy.
- Extreme and rapidly developing hypertension is divided into two groups: emergency, where there is evidence of end-organ damage; and urgency, where there is not. Urgencies are treated more slowly and with oral medications; emergencies require hospitalization and more rapid-acting interventions.

LOW BLOOD PRESSURE

The mechanism for short-term maintenance of blood pressure described previously is designed to respond rapidly to changes in both internal and external environments. Over the course of the day, this system of increased autonomic activity usually accommodates changes in activity, especially changes in position. Recall that when moving from a supine position to sitting or standing, gravity pulls blood away from the upper body and stimulates the baroreceptors in the carotid arteries and aortic arch; 500 to 1000 mL of a person's circulating blood volume pools in the venous system of the lower extremities. Messages transmitted from these receptors to the vasomotor center of the brain result in SNS activation, increasing both heart rate and arterial smooth muscle tone. The effect of these SNS-mediated responses to position change is the rapid increase in blood pressure and improved perfusion to the upper body, especially the brain. When this mechanism fails to produce this response in a timely fashion, the drop in blood pressure with position change is called *orthostatic hypotension (OH),* and may have serious consequences.

OH (postural) is a widespread but often unrecognized disorder with potentially serious consequences. It has been reported to occur in 6% to 30% of healthy elderly persons with normal blood pressures. When perfusion is not rapidly returned to the brain, dizziness, blurred vision, fainting (syncope), and injury from falls are familiar outcomes. But OH can have even more serious consequences. It has been demonstrated to be associated with cardiovascular disease, and research results indicate it may predict stroke, cognitive impairment, and death.

The standard definition of orthostatic hypotension is a decrease in systolic blood pressure of ≥20 mm Hg or a decrease in diastolic pressure that is ≥10 mm Hg within 3 minutes of moving to an upright position. An excessive increase in heart rate, by 20 to 30 beats/minute, is also diagnostic. Some researchers have found that the response can be delayed well beyond that timeframe in the elderly, perhaps up to 10 minutes or more.

An ineffective response to position change may be associated with problems within the nervous system resulting from a number of pathologies, a vasovagal reaction, depletion in circulating volume, or cardiac dysrhythmias. It can also occur as an adverse effect of drug therapy, most frequently antihypertensive, tricyclic antidepressant, and pain medications. The elderly are at special risk of OH when they are taking these medications. They also commonly have an inadequate fluid intake, age-related decreases in autonomic nervous system function, and disorders such as Parkinson disease and diabetes with which postural hypotension is associated. OH also is more common in the presence of arterial stiffness, which may be caused by an alteration in baroreceptor sensitivity. Alcohol ingestion and exposure to heat will also cause vasodilation and may precipitate OH.

A normal response to the SNS activation by the baroreceptors depends on effective functioning of all components of the system. Damage to

the vasomotor center or neurons within the central or peripheral nervous system may be responsible for a lack of sufficient response. This may be caused by disease or blunted responses associated with normal aging, prolonged bed rest, or medications. The prevalence of OH in Parkinson disease has been reported to be 37% to 58%. Direct damage to nerve fibers by elevated blood glucose levels in diabetes or an autoimmune injury as in multiple sclerosis can blunt the response, as well as impaired transmission resulting from spinal cord injury. Altered sensitivity of the baroreceptors has been well documented. The vasovagal response is a paradoxical increase in parasympathetic activity and a decrease in sympathetic activity resulting in bradycardia and vasodilation rather than an increase in heart rate and vasoconstriction. This contradictory response can be triggered by other stimuli such as stress, painful or unpleasant events, and activities such as coughing that increase intraabdominal or intrathoracic pressures. Dysrhythmias that impair cardiac output or an inadequate volume in the vascular space to respond to vasoconstriction signals will both also produce OH. Volume depletion as occurs in hemorrhage, burns, or severe diarrhea may reach a point where normal compensatory responses to position changes are inadequate; this happens after about a 30% volume loss. The development of postural changes in vital signs is a useful clinical indicator of inadequate circulating volume.

Because OH is often caused by physiologic conditions that are not amenable to modification, patients must be taught how to make changes to avoid initiating the response or to reduce its impact. In addition to changing positions slowly to reduce the initial drop in blood pressure, patients are encouraged to avoid hot environments (baths or saunas), because of their vasodilating effects, and large or carbohydrate-heavy meals, because postprandial hypotension can result from the increased blood volume drawn to the splanchnic bed. When symptoms begin, before fainting, actions can be taken to prevent the progression of OH. Squatting, bending forward to lower the head, or crossing the legs while tightening calf, thigh, and buttocks muscles may counter the effects. Elastic compression stockings and abdominal binders have documented value, as does elevating the head of the bed. Unless contraindicated, liberal intake of both salt and fluids is encouraged. Medication history should be carefully reviewed; a thorough history and physical examination, chemistry panel and blood count, and 12-lead electrocardiogram should be carried out. Elderly patients may find use of a cane with a folded tripod seat or a walker equipped with a seat helpful to reduced OH-related falls.

KEY POINTS

- Orthostatic hypotension (OH) is an extreme response to the change from supine to upright position, where the activation of the short-term control mechanisms is slow or inadequate in its response. Heart rate and diastolic and systolic blood pressures are more affected by gravitational effects of position change than is normally expected.
- OH results in dizziness, blurred vision, confusion, and possible syncope, which may cause injuries secondary to falls. OH is associated with cardiovascular disease and is a risk factor for stroke, cognitive impairment, and death.
- OH may be the result of a number of pathologies involving the baroreceptor response, damage to the vasomotor center or the peripheral nervous system, a vasovagal reaction, or cardiac dysrhythmias, or it may be an adverse drug effect. Most often it occurs because of insufficient circulating volume.
- Nonpharmaceutical interventions may be used if the cause cannot be ameliorated.

SUMMARY

Adequate perfusion of body organs and tissues depends on the maintenance of arterial blood pressure. This is accomplished through the highly orchestrated interaction of multiple systems on both a short-term and a long-term basis.

Blood pressure may be elevated secondary to other pathologic conditions or to food or drug ingestion. Secondary hypertension is treated by managing the causative factors, although medication also may be necessary. More commonly the etiology is not discernible, although risk factors are identified, and primary hypertension is diagnosed. Primary hypertension affects millions of Americans and is a public health concern worldwide. Once identified, lifestyle modifications and pharmaceutical interventions are initiated to avoid the significant pathologic outcomes to body organs.

When the mechanism for short-term blood pressure regulation fails to adequately respond to position changes, the resulting OH can cause syncope and potential injury. If the cause cannot be identified and treated, accommodations can be used to decrease its occurrence and minimize risks.

RESOURCES

Measuring Blood Pressure and Definitions of Hypertension

American Association of Critical-Care Nurses: Practice alert: noninvasive blood pressure monitoring. *AACN News* 23(6):4–5, 2006.

Argarwal R, et al: Role of home blood pressure monitoring in overcoming therapeutic inertia and improving hypertension control: a systematic review and meta-analysis. *Hypertension* 57(1):29–38, 2011.

Cicolini G, Gagliardi G, Ballone E: Effects of Fowlers body position on blood pressure measurement. *J Clin Nurs* 19(23-24):3581–3583, 2010.

Eguchi K, et al: Consistency of blood pressure differences between the left and right arms. *Arch Intern Med* 167:388–393, 2007.

Eşer I, et al: Issues in clinical nursing: the effect of different body positions on blood pressure. *J Clin Nurs* 16:137–140, 2006.

Godwin M, et al: A primary care pragmatic cluster randomized trial of the use of home blood pressure monitoring on blood pressure levels in hypertensive patients with above target blood pressure. *Fam Pract* 27:135–142, 2010.

Hall JE: *Guyton and Hall textbook of medical physiology*, ed 13, Philadelphia, 2016, Saunders.

Headley JM: Arterial pressure-based technologies: a new trend in cardiac output monitoring. *Crit Care Nurs Clin North Am* 18(2):179–187, 2006.

James PA, et al: 2014 Evidence-based guideline for the management of high blood pressure in Adults: Report from the panel members appointed to the eight joint national committee (JNC 8). *JAMA* 311(5):507–520, 2014.

Jones S, Simpson H, Ahmed H: A comparison of two methods of blood pressure measurement. *Br J Nurs* 15(17):948–951, 2006.

O'Rourke MF, Seward JB: Central arterial pressure and arterial pressure pulse: new views entering the second century after Korotkov. *Mayo Clin Proc* 81(8):1057–1068, 2006.

Parati G, Mancia G: Assessing the white-coat effect: which blood pressure measurement should be considered? *J Hypertens* 24:29–31, 2006.

Pickering TG, et al: Recommendations for blood pressure measurement in humans and experimental animals, I: blood pressure measurement in humans: a statement for professionals from the Subcommittee of Professional and Public Education of the American Heart Association Council on High Blood Pressure Research. *Hypertension* 45:142–161, 2005.

Podoll A, et al: Inaccuracy in pediatric outpatient blood pressure measurement. *Pediatrics* 119(3):e538–e543, 2007. Available at: www.pediatrics.org/cgi/content/full/119/3/e538.

Schell KA: Evidence-based practice: noninvasive blood pressure measurement in children. *Pediatr Nurs* 32(3):263–267, 2006.

Shaw J, et al: Are stroke patients' reports of home blood pressure readings reliable? Cross sectional study. *Fam Pract* 28:118–122, 2011.

Urbina E, et al: Ambulatory blood pressure monitoring in children and adolescents: recommendations for standard assessment: a scientific statement from the American Heart Association Atherosclerosis, Hypertension and Obesity Youth Committee of the Council on Cardiac Disease in the Young and the Council for High Blood Pressure Research. *Hypertension* 52:433–451, 2008.

U.S. Department of Health and Human Services, National Institutes of Health, National Heart, Lung, and Blood Institute: *The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure*, 2004, NIH Pub No. 04–5230. Available at www.nhlbi.nih.gov/guidelines/hypertension/jnc7full.pdf.

Xi B, Zong XN, Kelishadi R, et al: Establishing International Blood Pressure References Among Nonoverweight Children and Adolescents Aged 6 to 17 Years. *Circulation* 133(4):398–408, 2016.

Mechanisms, Risks and Treatment of High Blood Pressure

Adrogué H, Madias NE: Sodium and potassium in the pathogenesis of hypertension. *N Engl J Med* 356(19):1966–1978, 2007.

Aronow W, et al: ACCF/AHA 2011 expert consensus document on hypertension in the elderly: a report of the American College of Cardiology foundation task force on clinical expert consensus documents. *Circulation* 123:2434–2506, 2011.

Bergel E, Barros A: Effect of maternal calcium intake during pregnancy on children's blood pressure: a systematic review of the literature. *BMC Pediatr* 7(Article 15):2007. Available at: www.biomedcentral.com/1471-2431/7/15.

Bouvet CB, et al: Arterial stiffness as a therapeutic target for isolated systolic hypertension: focus on vascular calcifications and fibrosis. *Curr Hypertens Rev* 6(1):20–31, 2010.

Charkoudian N, Rabbitts J: Sympathetic neural mechanisms in human cardiovascular health and disease. *Mayo Clin Proc* 84(9):822–830, 2009.

Chioler A, et al: Has high blood pressure increased in children in response to the obesity epidemic? *Pediatrics* 119:544–553, 2007.

Fagard RH, Cornelissen VA: Effect of exercise on blood pressure control in hypertensive patients. *Eur J Cardiovasc Prev Rehabil* 14:12–17, 2007.

Falkner B: Hypertension in children and adolescents: epidemiology and natural history. *Pediatr Nephrol* 25:1219–1224, 2010.

Franklin SS, et al: Predictors of new-onset diastolic and systolic hypertension: the Framingham Heart Study. *Circulation* 111(9):1121–1127, 2005.

Gomez R, et al: Primary versus secondary hypertension in children followed up at an outpatient tertiary unit. *Pediatr Nephrol* 26:441–447, 2011.

Grassi G, et al: Baroreflex function in hypertension: consequences of antihypertensive therapy. *Prog Cardiovasc Dis* 48(6):407–415, 2006.

Guariguata L, Whiting DR, Hambleton I, et al: Global estimates of diabetes prevalence for 2013 and projections for 2035. *Diabetes Res Clin Pract* 103(2):137–149, 2014.

Ingelfinger JR: The molecular basis of pediatric hypertension. *Pediatr Clin North Am* 53:1011–1028, 2006.

Kennedy S: Clinical update: essential hypertension—recent changes in management. *Community Pract* 79(1):23–24, 2006.

Kohan DE: Endothelin, hypertension and chronic kidney disease: new insights. *Curr Opin Nephrol Hypertens* 19:134–139, 2010.

Lawlor DA, Smith GD: Early life determinants of adult blood pressure. *Curr Opin Nephrol Hypertens* 14:259–264, 2005.

Materson BJ: Variability in response to antihypertensive drugs. *Am J Med* 120(4A):S10–S20, 2007.

Mathew B: Obesity-hypertension: emerging concepts in pathophysiology and treatment. *Am J Med Sci* 334(1):23–30, 2007.

Mattace-Raso F, et al: Arterial stiffness, cardiovagal baroreflex sensitivity in older adults: the Rotterdam study. *J Hypertens* 25(7):1421–1426, 2007.

McCrindle BW: Assessment and management of hypertension in children and adolescents. *Nat Rev Cardiol* 7:155–163, 2010.

McEniery CM, et al: Increased stroke volume and aortic stiffness contribute to isolated systolic hypertension in young adults. *Hypertension* 46:221–226, 2005.

Mozaffarian D, Benjamin EJ, Go AS, et al: *Heart Disease and Stroke Statistics—2016 Update*, 2015. A Report From the American Heart Association.

Narchi H: Assessment and management of hypertension in children and adolescents: part B—investigation and management. *J Med Sci* 4(1):14–24, 2011.

Peixoto AJ, White WB: Circadian blood pressure: clinical implications based on the pathophysiology of its variability. *Kidney Int* 71(9):855–860, 2007.

Public Library of Science (June 18): Circadian rhythms dominate all life functions, *Science Daily*, 2007. Available at www.sciencedaily.com/releases/2007/06/070615075550.htm.

Rosendorff C, et al: Treatment of hypertension in the prevention and management of ischemic heart disease. *Circulation* 115:2761–2788, 2007.

Siragy HM: Angiotensin II compartmentalization within the kidney: effects of salt diet and blood pressure alterations. *Curr Opin Nephrol Hypertens* 15:50–53, 2006.

Spiotta RT, Luma GB: Evaluating obesity and cardiovascular risk factors in children and adolescents. *Am Fam Physician* 78(9):1052–1058, 2008.

Tanemoto M: Regulatory mechanism of "K^+ recycling" for Na^+ reabsorption in renal tubules. *Clin Exp Nephrol* 11:1–6, 2007.

Vehaskari VM: Developmental origins of adult hypertension: new insights into the role of the kidney. *Pediatr Nephrol* 22:490–495, 2007.

Yoon EY, et al: Medical management of children with primary hypertension by pediatric subspecialists. *Pediatr Nephrol* 24:147–153, 2009.

Yoon S, Ostchega Y, Louis T: *Recent trends in the prevalence of high blood pressure and its treatment and control, 1999-2008, NCHS Data Brief No. 48*, Hyattsville, MD, 2010, National Center for Health Statistics.

Secondary Hypertension

Cay S, Metin F, Korkmaz S: A common cause of secondary hypertension: coarctation of the aorta. *Heart* 92:734, 2006.

Chan DK, Chow AS, Kwok K: Childhood sleep-disordered breathing and its implications for cardiac and vascular diseases. *J Paediatr Child Health* 41:640–646, 2005.

Chandar J, Zilleruelo G: Hypertensive crisis in children. *Pediatr Nephrol* 27(5):741–751, 2012.

Chow CK, Teo KK, Rangarajan S, et al: Prevalence, awareness, treatment, and control of hypertension in rural and urban communities in high-, middle-, and low-income countries. *JAMA* 310(9):959–968, 2013.

De Caro E, et al: Aortic arch geometry and exercise-induced hypertension in aortic coarctation. *Am J Cardiol* 99:1284–1287, 2007.

Doi S, et al: Optimal use and interpretation of the aldosterone renin ratio to detect aldosterone excess in hypertension. *J Hum Hypertens* 20(7):482–489, 2006.

Driscoll DM, et al: Acute cardiovascular changes with obstructive events in children with sleep disorder breathing. *Sleep* 32(10):1265–1271, 2009.

Feldstein C: Management of hypertensive crises. *Am J Ther* 14:135–139, 2007.

Fernandes GH, et al: Delayed diagnosis of pheochromocytoma associated with chronic kidney disease. *Indian J Nephrol* 20(3):166–167, 2010.

Frishman WH, et al: Pathophysiology and medical management of systemic hypertension in pregnancy. *Cardiol Rev* 13(6):274–284, 2005.

Hernandez-Vila E: A Review of the JNC 8 Blood Pressure Guideline. *Tex Heart Inst J* 42(3):226–228, 2015.

Ipek E, Oktay AA, Krim SR: Hypertensive crisis: an update on clinical approach and management. *Curr Opin Cardiol* 2017. March 16, 2017. [Epub ahead of print].

James PA, Oparil S, Carter BL, et al: 2014 evidence-based guideline for the management of high blood pressure in adults: Report from the panel members appointed to the eighth joint national committee (JNC 8). *JAMA* 311(5):507–520, 2014.
Lin PC, et al: Pheochromocytoma underlying hypertension, stroke, and dilated cardiomyopathy. *Tex Heart Inst J* 34:244–246, 2007.
Mancia G, Fagard R, Narkiewicz K, et al: 2013 ESH/ESC Guidelines for the management of arterial hypertension. *Blood Press* 22(4):193–278, 2013.
Newton-Cheh C, et al: Clinical and genetic correlates of aldosterone-to-renin ratio and relations to blood pressure in a community sample. *Hypertension* 49:846–856, 2007.
Polson JW, et al: Evidence for cardiovascular autonomic dysfunction in neonates with coarctation of the aorta. *Circulation* 113:2844–2850, 2006.
Schulenburg M: Management of hypertensive emergencies: implications for the critical care nurse. *Crit Care Nurs Q* 30(2):86–93, 2007.
Smith ML, Pacchia CF: Sleep apnoea and hypertension: role of chemoreflexes in humans. *Exp Physiol* 92(1):45–50, 2007.
Sowers JR, Whaley-Connell A, Epstein M, et al: Narrative review: the emerging clinical implications of the role of aldosterone in the metabolic syndrome and resistant hypertension. *Ann Intern Med* 150:776–783, 2009.
Touyz RM, Dominiczak AF: Hypertension Guidelines. Is it time to reappraise blood pressure thresholds and targets? *Hypertension* 67(4):688–689, 2016, 2016.
Vamvakis A, Gkaliagkousi E, Troamtafyllou A, Douma S: Beneficial effects of nonpharmacological interventions in the management of essential hypertension. *JRSM Cardiovasc Dis* 6:1–6, 2017.
Weiss JW, Liu Y, Huang J: Physiological basis for a causal relationship of obstructive sleep apnoea to hypertension. *Exp Physiol* 92:21–28, 2007.
Young WF: Primary aldosteronism: renaissance of a syndrome. *Clin Endocrinol* 66:607–618, 2007.
Yu S, et al: Effect of revised UPPP surgery on ambulatory blood pressure in sleep apnea patients with hypertension and oropharyngeal obstruction. *Clin Exp Hypertens* 32:49–53, 2010.

Low Blood Pressure

Cooke J, et al: The changing face of orthostatic and neurocardiogenic syncope with age. *QJM* 104(8):689–695, 2011.
Sathyapalan T, Atkin SL: Postural hypotension. *Br Med J* 342:1–3, 2011.
Task Force for the Diagnosis and Management of Syncope of the European Society of Cardiology: Guidelines for the diagnosis and management of syncope (version 2009). *Eur Heart J* 30(21):2631–2671, 2009.
Thomson P, Wright J, Chakravarthi R: Non-pharmacological treatments for orthostatic hypotension. *Age Ageing* 40:292–293, 2011.
Zesiewicz TA, et al: Practice parameter: treatment of nonmotor symptoms of Parkinson disease. *Am Acad Neurol* 74:924–931, 2010.

17

Cardiac Function

Jacquelyn L. Banasik

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- How are events of the cardiac cycle reflected in pressure and volume changes within the cardiac chambers?
- What factors affect the blood supply to myocardial tissue?
- How does sarcomere cross-bridge formation lead to muscle cell contraction?
- What is the process of excitation-contraction coupling in heart muscle cells?
- How are action potentials generated and conducted in myocardial and pacemaker cells?
- How does the electrocardiogram relate to impulse conduction through the heart?
- How do heart rate, preload, afterload, and contractility affect cardiac output and cardiac workload?
- What diagnostic tests are used to evaluate cardiac structure and function?

CHAPTER OUTLINE

Cardiovascular Anatomy, 355
Heart, 355
Circulatory System, 356
Cardiac Cycle, 358
Isovolumic Contraction, 359
Ventricular Ejection, 360
Isovolumic Relaxation, 360
Atrial Events, 360
Aortic and Pulmonary Artery Events, 360
Coronary Circulation, 360
Anatomy of the Coronary Vessels, 360
Regulation of Coronary Blood Flow, 361
Cardiac Myocytes, 363
Myocyte Structure, 363
Structure of the Contractile Apparatus, 363
Characteristics of Contractile Filaments, 365
Molecular Basis of Contraction, 366
Overview of Contraction, 366
Sliding Filament/Cross-Bridge Theory of Muscle Contraction, 366
Role of Calcium in Muscle Contraction, 367
Energy of Muscle Relaxation, 367
Cardiac Energy Metabolism, 368
Oxygen Utilization, 368
Substrate Utilization, 369
Cardiac Electrophysiology, 369
Cardiac Resting Potential, 369
Cardiac Action Potential, 369
Phase 0, 369
Phase 1, 370
Phase 2, 370
Phase 3, 370
Phase 4, 370
Rhythmicity of Myocardial Cells, 370
Specialized Conduction System of the Heart, 371
Autonomic Regulation of Rhythmicity, 372
Electrocardiography, 372
Determinants of Cardiac Output, 374
Determinants of Heart Rate, 374
Determinants of Stroke Volume, 375
Volume of Blood in the Heart (Preload), 375
Contractile Capabilities of the Heart (Contractility), 375
Impedance to Ejection From the Ventricle (Afterload), 376
Cardiac Workload, 376

Endocrine Function of the Heart, 376
Tests of Cardiac Function, 376
Electrocardiography, 376
Magnetic Resonance Imaging and Computed Tomography, 378
Echocardiography, 378
Nuclear Cardiography, 379
Cardiac Catheterization/Coronary Angiography, 379

The primary function of the heart is to produce the driving force that propels blood through the vessels of the circulatory system. Along with the lungs, the heart works to distribute oxygenated blood and nutrients to tissues and organs of the body. Complex regulatory mechanisms function to match the cardiac output with the metabolic needs of the tissues. Cardiac dysfunction can lead to abnormal function or death of cells in tissues throughout the body. Cardiovascular disease is the leading cause of mortality in the United States, and a significant proportion of the population suffers from physical limitations associated with impaired cardiac function. Familiarity with cardiac anatomy and physiology is requisite to understanding cardiac diseases and therapy.

CARDIOVASCULAR ANATOMY

Heart

The heart is located in the mediastinum**,** suspended between the lungs; behind the sternum; and in front of the vertebral column, thoracic aorta, and esophagus (Fig. 17.1). When viewed from the front, the heart appears to be rotated to the left so that the right atrium and right ventricle are most anterior. The base of the heart protrudes somewhat into the right side of the chest and is relatively fixed in place by its attachments to the great vessels. The apex of the heart lies primarily in the left side of the chest and is directed forward toward the anterior chest wall. With each heartbeat, a characteristic thrust, or point of maximal impulse (PMI), is generated and can be palpated where the apex strikes against the chest. The PMI is normally located on the left side of the chest where the fifth intercostal space and midclavicular line intersect. Variations in heart size and position within the chest may be related to age, body size, shape, weight, or pathologic conditions of the heart and other nearby structures.

Functionally important cardiac tissues include connective tissues, which form the fibrous skeleton and valves; cardiac muscle, which produces the contractile force; and epithelial tissue, which lines the cardiac chambers and covers the outer surfaces of the heart. The fibrous skeleton includes an extensive network of matrix that supports cardiac cells and four rings that provide a firm scaffold for attachment of the cardiac valves (Fig. 17.2). Four cardiac valves control the direction of blood flow through the heart (Fig. 17.3). The mitral valve (bicuspid) directs blood flow from the left atrium to the left ventricle, whereas the tricuspid valve directs blood from the right atrium to the right ventricle. The edges of these atrioventricular (AV) valves are attached to rings formed by the fibrous skeleton. Valve leaflets are tethered to papillary muscles of the ventricular chambers by connective tissues called chordae tendineae**.** Papillary muscles attach to ventricular walls and help prevent the valve leaflets from bending backward into the atria during ventricular contraction (Fig. 17.4). The AV valves open passively during diastole when the pressure of blood in the atria exceeds that in the ventricles. Ventricular contraction reverses the pressure gradient and causes AV valves to snap shut, preventing blood from flowing backward into the atria.

Two semilunar valves are located in the ventricular outflow tracts. The pulmonic valve lies between the right ventricle and pulmonary artery, and the aortic valve lies between the left ventricle and aorta. Compared with the AV valves, the semilunar valves are thicker and are not supported by fibrous cords. They open and close passively according to pressure gradients, just as the AV valves do. When intraventricular pressures exceed pulmonary and aortic pressures, the semilunar valves remain open and then close when ventricular pressures fall below aortic and pulmonary artery pressures.

The cardiac muscle layer (myocardium) produces the contractile force that pushes blood through the circulatory system. Heart muscle is organized into four separate chambers of varying muscular wall thickness, reflecting the degree of pressure each chamber must generate to pump blood. Atria serve primarily as conduits and have a thinner

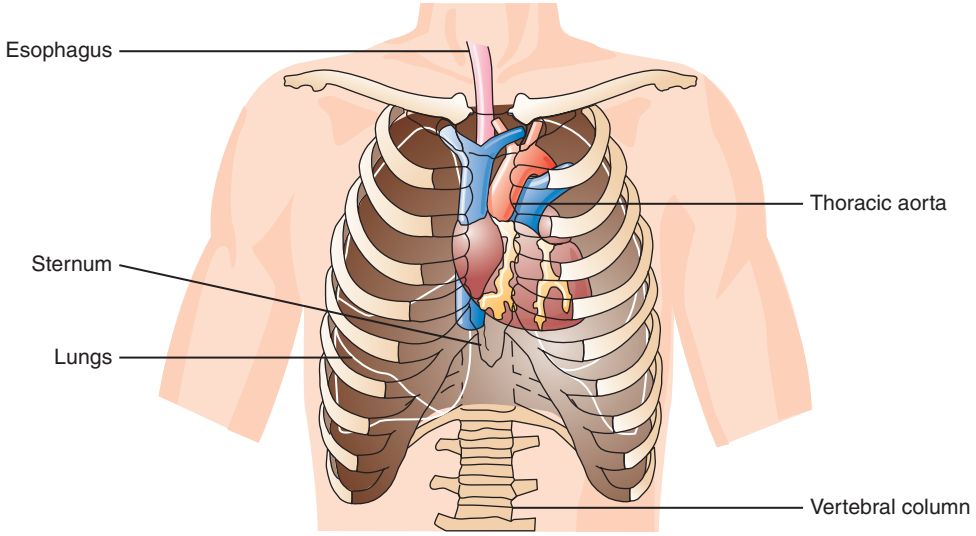


FIG 17.1 Position of the heart in the mediastinum. The base of the heart protrudes into the right side of the chest, whereas the apex lies in the lower left side of the chest.

layer of muscle than the ventricles. The left ventricular muscle is two to three times thicker than that of the right ventricle because higher pressures are required to eject blood into the systemic circulation than into the pulmonic system. Normal chamber pressures are shown in Table 17.1. Alterations in chamber pressures may reflect pathologic cardiovascular changes such as valvular disorders, blood volume abnormalities, and heart failure (see Chapters 18 and 19).

Cardiac chambers and valves are lined by a layer of squamous epithelial cells called the endocardium**.** The endocardial layer provides a smooth surface that prevents clotting and minimizes trauma to red blood cells. The endocardium is continuous with the endothelium of the vascular system. Outer surfaces of the heart are also covered by a layer of epithelial cells called the epicardium**,** which is part of a protective covering called the *pericardium*. The pericardium is composed of two layers that envelop the heart like a sac (Fig. 17.5). The inner layer (visceral pericardium or epicardium) is attached directly to the heart's outer surface, whereas an outer layer (parietal pericardium) forms a sac around the heart. The parietal pericardium is composed of an epithelial layer and a tough fibrous layer.

Visceral and parietal pericardial layers are separated by a thin, fluid-filled space (pericardial space) that usually contains 10 to 30 mL of serous fluid. This fluid lubricates pericardial surfaces and reduces friction while the layers slide against one another during cardiac contraction. Accumulations of fluid in the pericardial space or inflammation of the pericardial sac can restrict cardiac filling and impair cardiac output.

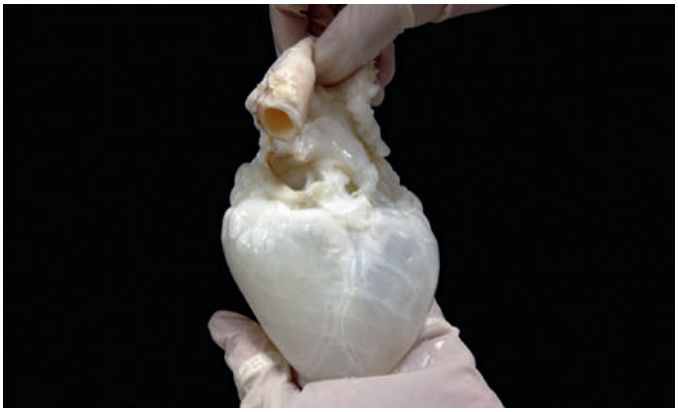


FIG 17.2 Connective tissue extracellular matrix of the heart with all cells removed developed by Dr. Doris Taylor's lab. (Image provided by Regenerative Medicine Research Labs at Texas Heart Institute.)

Circulatory System

The circulatory systems of the lungs and body can be viewed as two separate but interdependent systems (Fig. 17.6). The left-sided heart chambers produce the force to propel blood through the vessels of the systemic (body) circulation. The left atrium receives oxygenated blood from the lungs by way of the pulmonary veins and delivers it to the

TABLE 17.1 Normal Pressures in the Heart

Location	Pressure (mm Hg)*
Right atrium	0–8
Right ventricle	15–28/0–8
Pulmonary artery	15–28/4–12
Left atrium	4–12
Left ventricle	100–120/4–12
Aorta	100–120/60–80

*Right and left atrial pressures listed as means; other pressures written as systolic/diastolic.

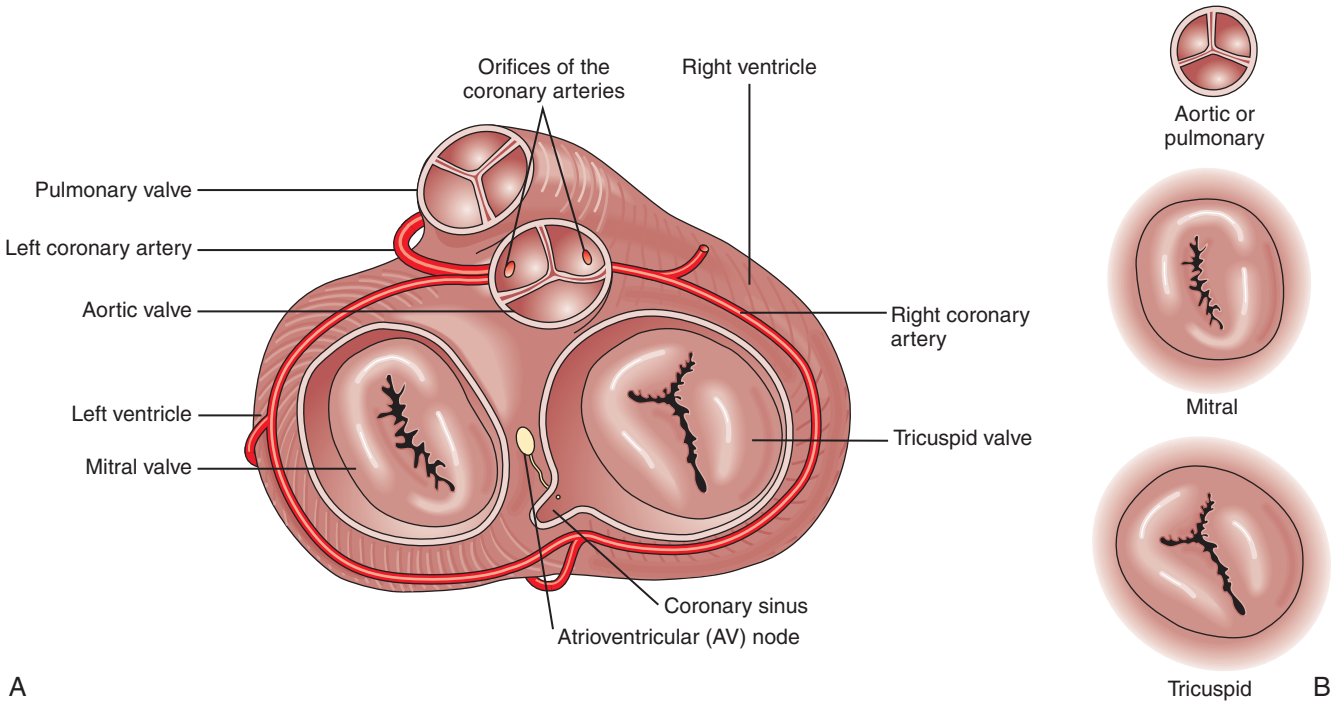


FIG 17.3 A, Position of the heart valves as viewed from above. **B,** Configuration of the heart valves showing the two cusps of the mitral valve and the three cusps of the tricuspid valve. The pulmonary and aortic valves have three leaflets.

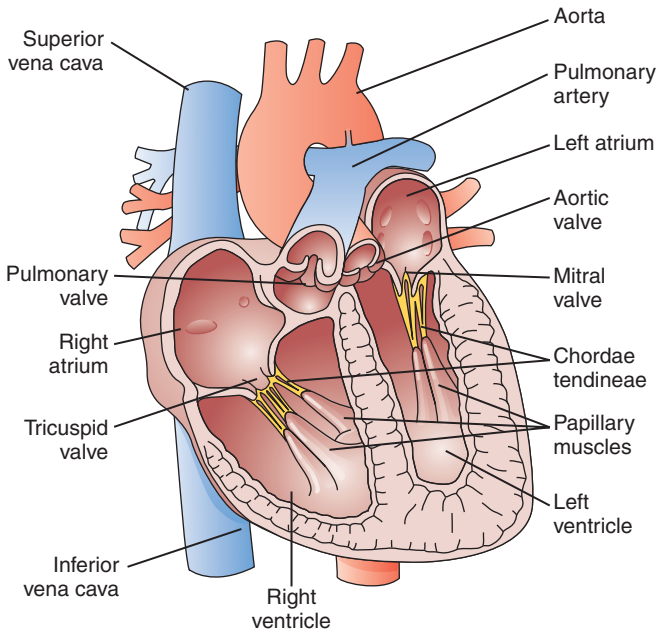


FIG 17.4 Chordae tendineae and papillary muscles attach the mitral and tricuspid valve leaflets to the ventricular myocardium.

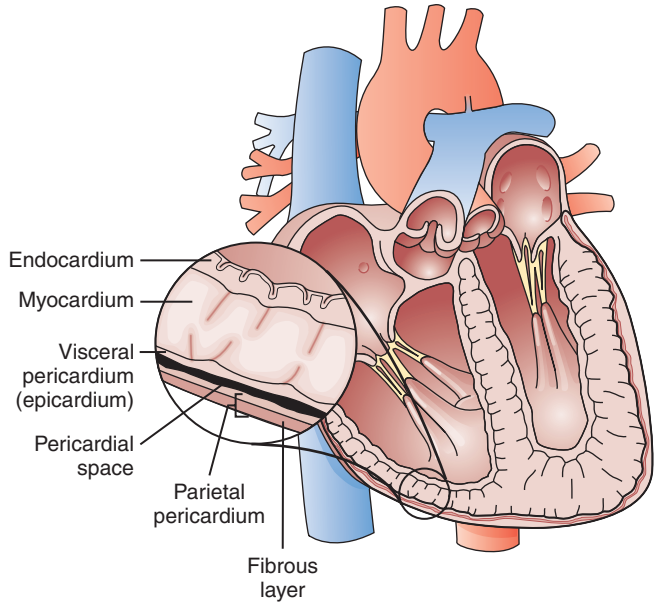


FIG 17.5 Pericardial sac is composed of two layers separated by a narrow, fluid-filled space. The visceral pericardium (epicardium) is attached directly to the heart's surface, whereas the parietal pericardium forms the outer layer of the sac.

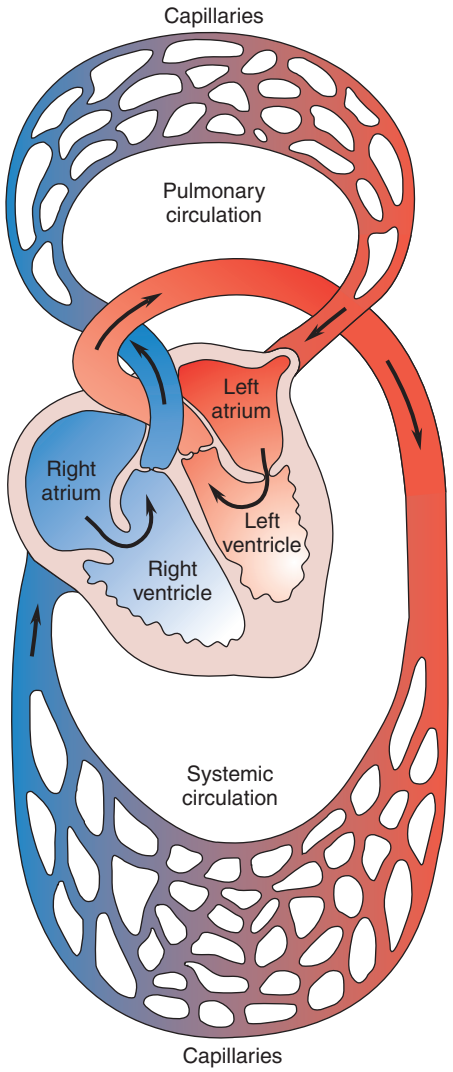


FIG 17.6 Systemic and pulmonary circulations viewed as separate but interdependent systems. The right ventricle pumps blood through the pulmonary vasculature, whereas the left ventricle pumps blood through the systemic circulation.

left ventricle. This oxygenated blood is pumped by the left ventricle into the aorta, which supplies the arteries of the systemic circulation. Venous blood is collected from capillary networks of the body and returned to the right atrium by way of the vena cavae. Blood from the head returns to the right atrium through the superior vena cava; blood from the body returns via the inferior vena cava. There are no valves between the vena cavae and the right atrium, and the atrial pressure waves that are generated during the cardiac cycle cause characteristic visible pulsations in the jugular veins. An increased right atrial pressure may be observed as distention within the jugular veins.

The right side of the heart receives deoxygenated blood from the systemic circulation and pumps it through the lungs by way of the pulmonary artery. The pulmonary artery divides into left and right branches, which subdivide to supply blood to pulmonary capillary beds. Exchange of respiratory gases occurs at the pulmonary capillaries so that blood delivered to the left atrium by the pulmonary veins is well oxygenated.

Blood flow through the left and right heart chambers is connected in series such that the output of one becomes the input of the other. Thus the functions of the right and left sides of the heart are interdependent. Failure of one side of the heart to pump efficiently soon leads to dysfunction of the other side.

Characteristic changes in the anatomy and physiologic functioning of the heart and circulatory systems occur with aging (see Geriatric Considerations: Changes in the Heart). In general, these changes result in a decreased cardiac reserve and a greater predisposition to cardiac muscle ischemia.

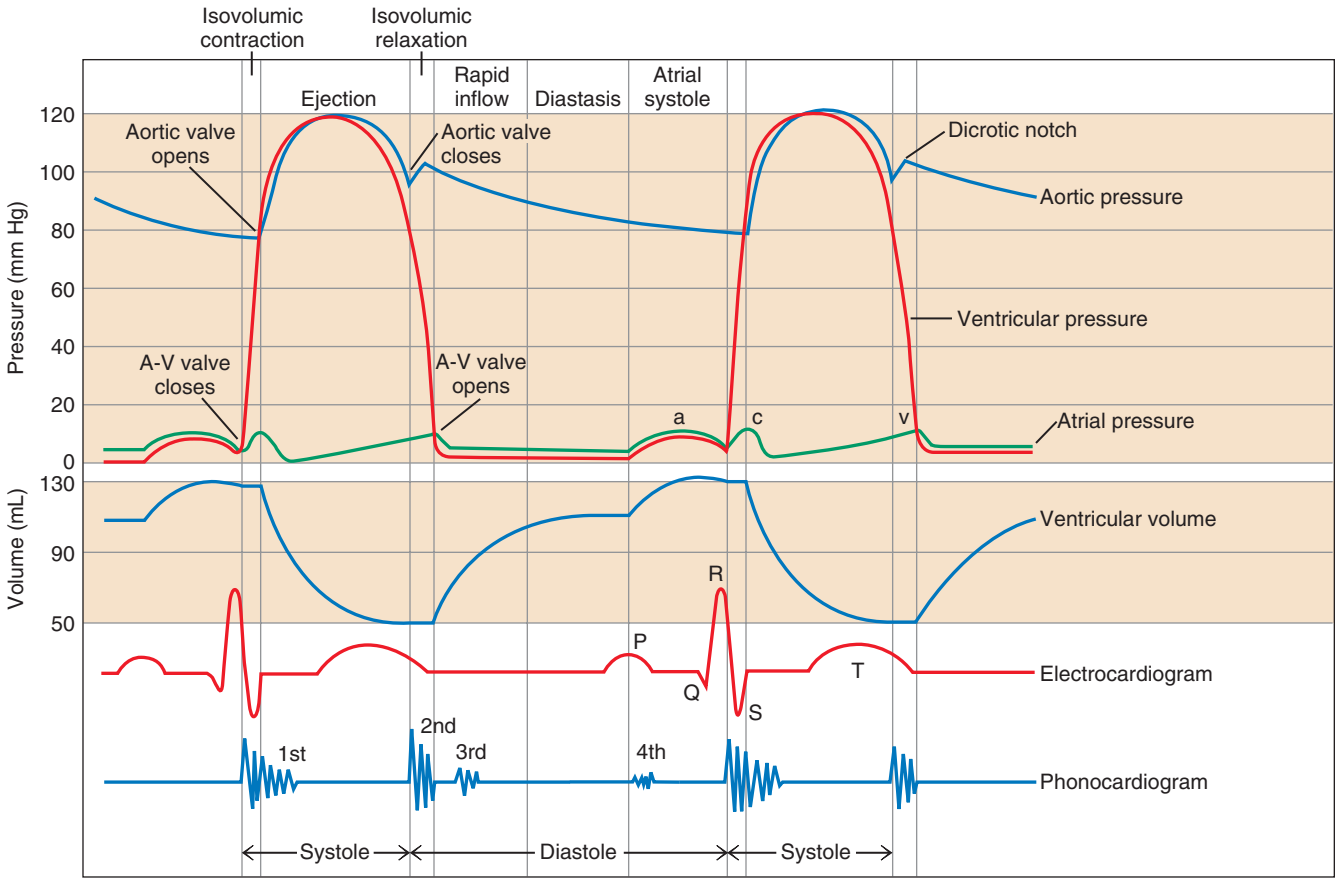


FIG 17.7 Events of the cardiac cycle showing relationships among left atrial and ventricular pressures, ventricular volume, and aortic pressure. An identical set of events occurs on the right side of the heart, although pressures are lower.

KEY POINTS

- Blood flows from the right atrium to the right ventricle through the tricuspid valve. The pulmonic valve lies between the right ventricle and the pulmonary artery. Blood flows from the left atrium to the left ventricle through the mitral valve. The aortic valve lies between the left ventricle and the aorta.
- Heart muscle (myocardium) is lined with endothelium on the inner surface and covered with epicardium on the outer surface.
- The pericardial sac envelops and protects the heart from friction.
- The right-sided heart chambers pump deoxygenated (venous) blood through the lungs. The left-sided heart chambers pump oxygenated blood through the systemic circulation.

CARDIAC CYCLE

Each heartbeat is composed of a period of ventricular contraction (systole) followed by a period of relaxation (diastole). The interval from one heartbeat to the next is called a cardiac cycle and includes ventricular, atrial, and aortic (or pulmonic) events. Each of these events is associated with characteristic pressure changes within the cardiac chambers. Pressure changes result in valvular opening and closing and unidirectional movement of blood through the heart. The various events of the cardiac cycle are illustrated as a function of time in Fig. 17.7. Another method of graphing ventricular function is the pressure–volume loop (Fig. 17.8). Pressure–volume loops are useful for assessing the relationships between pressure and volume at various points in the cardiac cycle to evaluate left ventricular function. Abnormalities in these waveforms may occur with diseases of the cardiac valves, changes in blood volume, or changes in pumping capacity of the heart (see Chapter 18). These waveforms are commonly monitored with specialized cardiac catheters in patients with cardiac or hemodynamic disorders.

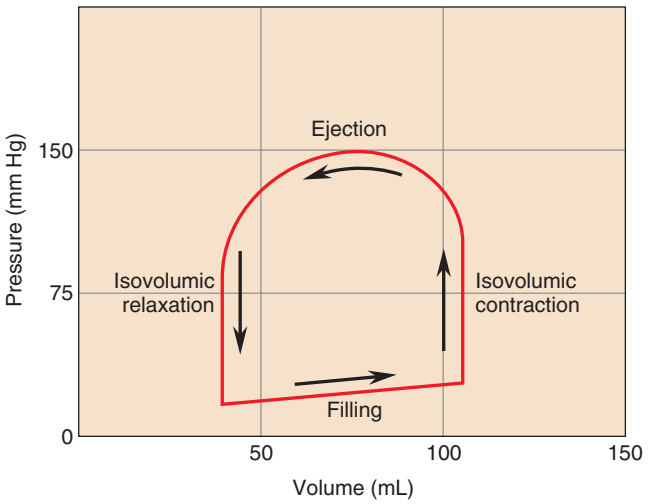


FIG 17.8 Pressure–volume loop showing changes in left ventricular volume and pressure during the cardiac cycle.

The cardiac cycle can be described sequentially, beginning with ventricular filling. During diastole the ventricles are relaxed and blood

flows in from the atria through open AV valves. Initially, ventricular filling occurs passively because of a pressure gradient between the atria and ventricles. Toward the end of ventricular diastole, the atria contract, squeezing more blood through the AV valves into the ventricles. The "atrial kick" provided by atrial contraction is particularly important during fast heart rates, when the time for ventricular filling is shortened; the atrial contraction helps load the ventricle quickly to prevent a reduction in stroke volume. Ventricular events include isovolumic contraction, ejection, and isovolumic relaxation. Each of these cycle events is further described in the following sections.

Isovolumic Contraction

Immediately after atrial systole the ventricles begin to contract, causing intraventricular pressure to rise and the AV valves to close. AV valve closure produces a sound that can be heard at the chest wall and is termed S_1. Ventricular pressure rises rapidly during isovolumic contraction because all four cardiac valves are closed, and the volume of blood within the ventricular chamber is forcefully compressed by the powerful ventricular myocardium (see Fig. 17.7, *red tracing*). Volume remains constant during this phase. The rate of rise in pressure is an indication of the

GERIATRIC CONSIDERATIONS

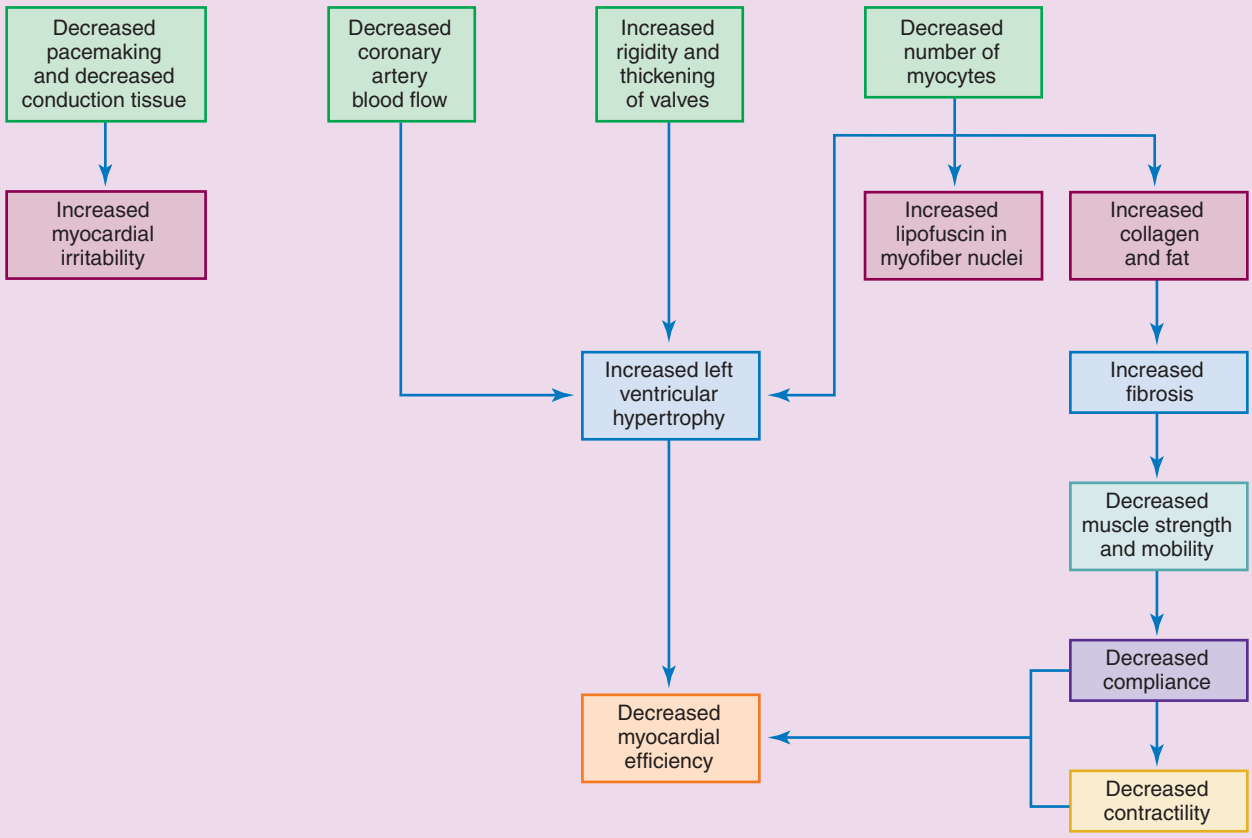
Changes in the Heart

With aging, there is a decrease in the number of myocytes, but normally the heart size does not change appreciably. With the loss of overall cardiac muscle tissue, a corresponding expansion occurs in myocardial collagen and fat. The left ventricular muscle wall becomes thicker, with a resulting increase in oxygen demand. The endocardium becomes fibrotic and sclerosed. Cross-linking of the collagen tissue within the heart muscle increases myocardial stiffening, which causes decreased compliance. The decrease in compliance produces a decline in cardiac contractility, which reduces the heart's pumping ability. The rate of ventricular relaxation decreases.

Fibrotic changes in cardiac valves result from a combination of hemodynamic stress and generalized thickening. There is also a decrease in coronary artery blood flow to the myocardium, which affects myocardial oxygen and nutrient supply. The myocardial cells increase in size, with increased lipofuscin pigment and lipid deposition.

Within the specialized electrical conduction tissue, there is loss of myocytes and fibrosis of conduction pathways, especially in the sinoatrial (SA) node, AV node, and bundle of His. The number of pacemaker cells in the SA node decreases, resulting in less responsiveness of that node to adrenergic stimulation. Myocardial cell irritability increases. On the ECG, the P wave may be notched or slurred. The PR interval is longer, and the QRS amplitude decreases. The axis may shift left as a result of left ventricular muscle thickening (hypertrophy). The T wave may be notched, and the amplitude may decrease.

The changes previously noted affect cardiac function. The resting heart rate in the elderly is unchanged. During stress or exercise, the aging heart is unable to respond quickly with an elevated rate, and the maximal heart rate elevation is reduced. Once the heart rate is elevated, it takes a much longer time for the heart rate to return to the resting level. The cardiac stroke volume and cardiac output generally decrease with age. Oxygen consumption in the myocardium is reduced, resulting in less efficient function when stressed and an overall decreased cardiac reserve.



contractile state of the heart. The greater the change in pressure per unit time (dP/dt), the higher the contractile state. Sympathetic nervous system activation increases dP/dt, whereas conditions such as heart failure are characterized by a slower rate of pressure development. The term *inotropy* is commonly used interchangeably with contractility and is reflected by the velocity and degree of cardiac muscle shortening during systole.

Ventricular Ejection

Ventricular contraction results in a rapid rise in ventricular pressure. As ventricular pressure exceeds aortic pressure (or pulmonic), the valve is forced open and a period of rapid ejection of blood from the ventricle follows. The rapid ejection phase is followed by a period of reduced ejection as aortic (or pulmonic) pressure rises and ventricular pressures and volumes fall. The amount of blood ejected with each contraction of the ventricle is called the *stroke volume* (SV), the volume of blood in the ventricle before ejection is the *end-diastolic volume* (EDV), and the amount of blood that remains in the ventricle after ejection is the *end-systolic volume* (ESV). Thus stroke volume equals EDV minus ESV. An important and commonly used index of pumping effectiveness is the ejection fraction (EF), which is calculated by dividing SV by EDV. A normal EF is 60% to 80%; patients with systolic heart failure often have an EF of less than 40%.

Isovolumic Relaxation

The isovolumic relaxation phase begins with semilunar valve closure in response to falling ventricular pressures and ends when the AV valves open to allow ventricular filling. Ventricular blood volume remains constant during this period because all four cardiac valves remain closed. Closure of the semilunar valves causes the second heart sound, S_2. Opening of the AV valves signals the beginning of rapid ventricular filling and the start of another cardiac cycle. The rate of ventricular relaxation is indicated by the drop in ventricular pressure per unit time and is called the $-dP/dt$. The rate and degree of ventricular relaxation is called *lusitropy* and is an energy-requiring process that reflects the efficiency of calcium removal from the cytoplasm. Rapid relaxation is necessary to allow the ventricle to fill quickly and at a low pressure before the next systole. Impaired relaxation (lusitropic dysfunction) is a common finding in patients with heart failure and contributes to the symptoms of congestion (see Chapter 19). Because relaxation of the ventricle is an energy-requiring process, it may become impaired when blood flow and oxygen delivery to the heart are inadequate.

Atrial Events

Atrial pressure waves have three characteristic curves: *a, c,* and *v* (see Fig. 17.7, *green tracing*). The *a* wave corresponds to atrial contraction, which immediately precedes AV valve closure. The *c* wave occurs early in ventricular systole and is thought to represent bulging of AV valves into the atrial chambers. The *v* waves have a gradual incline, which represents filling of the atrium as blood returns from the circulation. The *v* wave drops rapidly as atrial pressure is relieved by AV valve opening. A large *v* wave is often associated with inadequate closure of the AV valve, resulting in regurgitation of ventricular blood back into the atrium during ventricular systole. The mean right atrial pressure, also called the *central venous pressure,* is commonly measured as an indicator of the blood volume in the heart, which is dependent in part on the amount of blood being returned from the systemic circulation.

Aortic and Pulmonary Artery Events

Aortic and pulmonary artery pressures rise and fall in relation to the cardiac cycle. Arterial pressures fall to their lowest value just before semilunar valve opening. This lowest pressure is called *diastolic blood pressure.* Arterial pressure reaches its maximum during ventricular ejection and is called *systolic blood pressure.* A characteristic notch *(dicrotic notch)* in the arterial pressure curve may be seen as the semilunar valves close (see Fig. 17.7, *blue tracing*).

The difference in aortic pressure between systole and diastole is partly dependent on the aorta's elastic characteristics. During systole, the aorta stretches to accommodate blood ejected by the ventricle. The stretched aorta has "stored," or potential, energy that is released during diastole to maintain driving pressure and to keep blood flowing continuously through the circulation. Aortic stiffening, as occurs with aging or arteriosclerosis, may result in higher systolic and lower diastolic blood pressures attributable to loss of aortic elastic properties. When aortic or pulmonic pressures are chronically elevated, the ventricles must generate more pressure to open the semilunar valves and eject the stroke volume. Over time this extra effort required to increase the pressure can damage the heart muscle and lead to hypertrophy or failure.

KEY POINTS

- Characteristic pressure wave changes that occur during the cardiac cycle may be useful in diagnosing cardiac disease and volume status.
- The atria have three characteristic waves: *a, c,* and *v.* The *a* wave corresponds to atrial contraction, the *c* wave corresponds to the atrioventricular (AV) valve bulging during ventricular contraction, and the *v* wave corresponds to atrial filling.
- The ventricles have four important phases: isovolumic contraction, ejection, isovolumic relaxation, and diastolic filling. The rate and amplitude of these pressures reflect chamber volume, contractility, and valvular function. Left ventricular pressure–volume relationships are frequently used to diagnose heart failure.
- Pressure changes in the aorta during a cardiac cycle are partly dependent on the elasticity of the aorta. Differences between systolic and diastolic pressures are less with a compliant aorta. Aortic stiffness results in higher systolic and lower diastolic pressures.

CORONARY CIRCULATION

Anatomy of the Coronary Vessels

The blood supply to heart muscle is provided by the coronary arteries (Fig. 17.9). Right and left coronary artery openings are located in the sinuses of Valsalva, in the aortic root, just beyond the aortic valve. The right coronary artery originates near the aortic valve's anterior cusp and passes diagonally toward the right ventricle in the AV groove. In approximately 50% of the population, the right coronary artery gives rise to a posterior descending vessel that supplies blood to the heart's posterior aspect. In 20% of the population, the left coronary artery is dominant in supplying blood to the ventricles, and in 30% of the population the right and left coronary arteries deliver about the same amount of blood and neither is dominant. The *left main coronary artery* arises near the aortic posterior cusp and travels a short distance anteriorly before dividing into the *left anterior descending* and *circumflex* branches. The anterior descending branch supplies septal, anterior, and apical areas of the left ventricle, whereas the circumflex artery supplies the lateral and posterior left ventricle. The three major coronary arteries give rise to a number of smaller branches that penetrate the myocardium and branch into small arterioles and capillaries. Regular exercise and stable atherosclerotic plaques in the coronary arteries are thought to stimulate the development of more extensive collateral circulation in the heart. Collateral vessels may help limit infarct size in patients suffering acute coronary occlusions (see Chapter 18). Areas supplied by divisions of the coronary arteries are listed in Table 17.2. Most of the heart's capillary beds drain into the coronary veins, which then empty into the right atrium through the coronary sinus (Fig. 17.10).

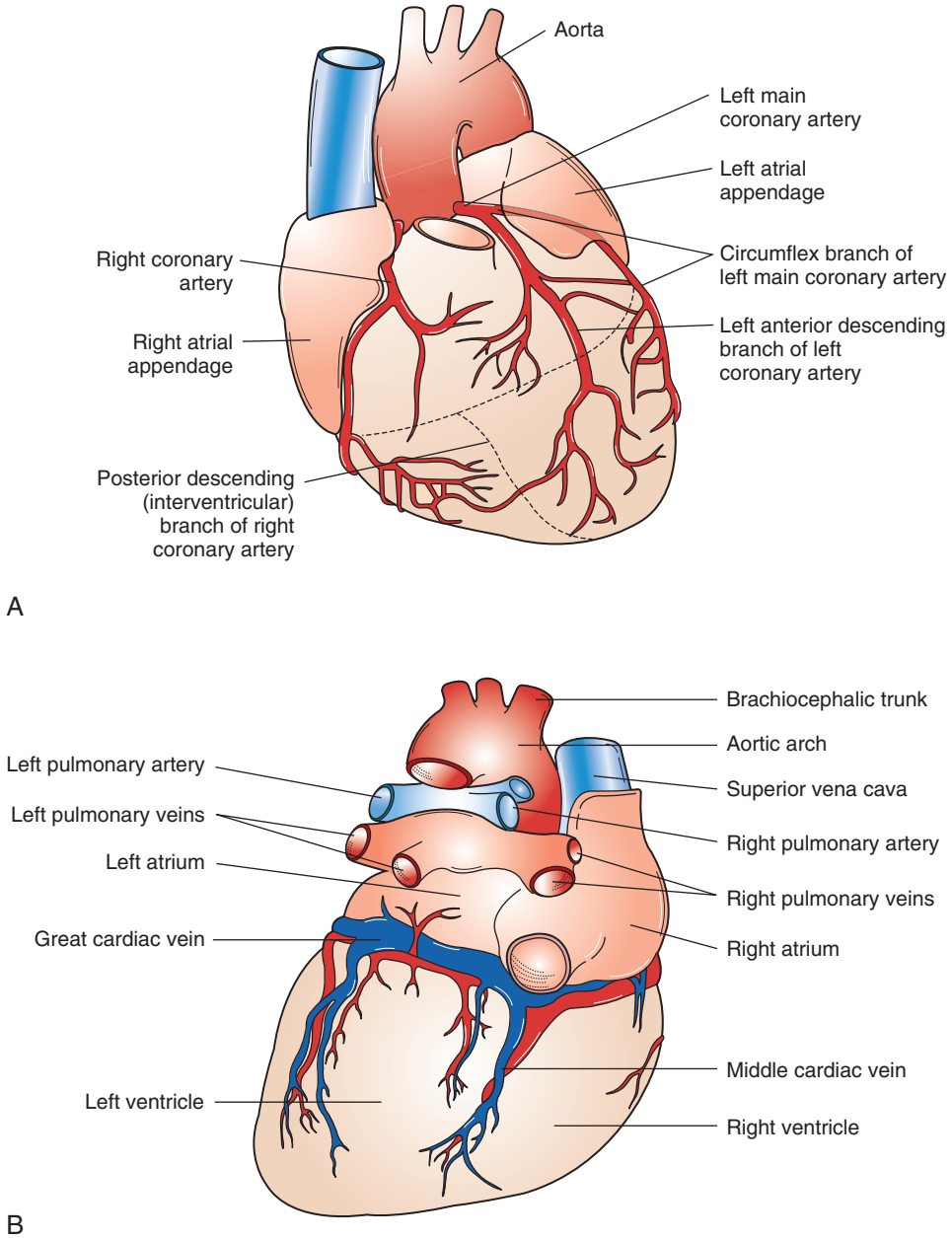


FIG 17.9 Coronary arteries supplying the heart. The right coronary artery supplies the right atrium, ventricle, and posterior aspect of the left ventricle in most individuals. The left coronary artery divides into the left anterior descending and circumflex arteries, which perfuse the left ventricle. **A,** Anterior view. **B,** Posterior view.

Regulation of Coronary Blood Flow

Blood flow through coronary vessels is determined by the same physical principles that govern flow through other vessels of the body, namely driving pressure and vascular resistance to flow. According to Ohm's law, an increase in driving pressure *(P)* increases blood flow *(Q)*, whereas an increase in resistance *(R)* reduces blood flow: $Q = P/R$ (see Chapter 15). Driving pressure through the coronary arteries is determined by aortic blood pressure and right atrial pressure. This relationship can be expressed in the following equation:

$$\text{Coronary driving pressure } (P) = \text{ABP} - \text{RAP}$$

where ABP is aortic blood pressure and RAP is right atrial pressure. Thus an increase in aortic pressure enhances coronary blood flow, whereas an increase in right atrial pressure opposes coronary flow.

Coronary vascular resistance *(R)* has two major determinants: (1) coronary artery diameter and (2) the varying degrees of external compression attributable to myocardial contraction and relaxation. Coronary artery diameter is continuously adjusted to maintain blood flow at a level adequate for myocardial demands. *Autoregulation* is the term used to describe the intrinsic ability of the arteries to adjust blood flow according to tissue needs. Vessel dilation *(vasodilation)* occurs in response to increased tissue metabolism or reduced driving pressure, whereas decreased metabolic activity or increased driving pressure results in a decreased vessel diameter *(vasoconstriction)*.

The mechanism of autoregulation can be explained by the metabolic hypothesis, which proposes that increased metabolism, reduced oxygen concentration, or decreased blood flow results in a buildup of vasodilatory chemicals in the vessel. Smooth muscle encircling the vessel relaxes in response to the presence of the chemicals, increasing vessel diameter.

TABLE 17.2 Areas Supplied by the Coronary Arteries

Artery	Area Supplied
Right coronary	Right atrium (55% of persons)
	Right ventricle
	Intraventricular septum
	Sinus node (55% of persons)
	Atrioventricular node
	Bundle of His
Left anterior descending	Right atrium (45% of persons)
	Right ventricle (minor)
	Left ventricle (anterior, apex)
	Anterior papillary muscles
	Right and left bundle branches
	Intraventricular septum
Left circumflex	Left atrium
	Left ventricle (posterior, anterior)
	Sinus node (45% of persons)

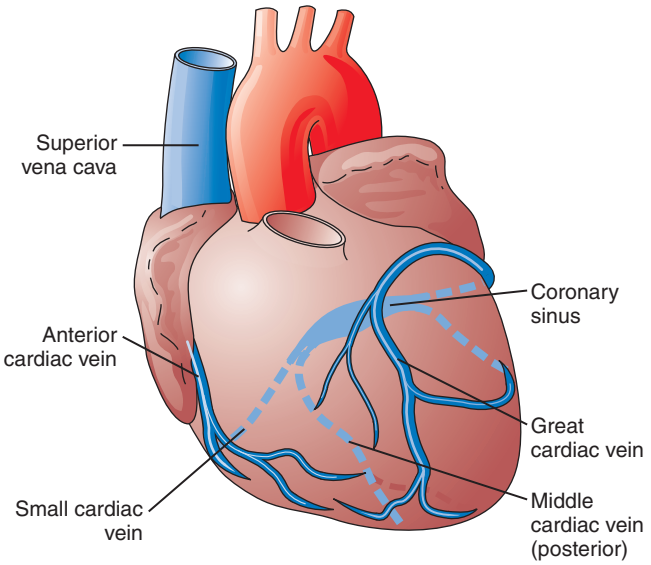


FIG 17.10 Venous drainage of the heart. Coronary veins drain blood from the myocardial capillary bed and deliver it into the right atrium.

Several vasodilating substances have been proposed, including potassium ions, hydrogen ions, carbon dioxide, nitric oxide, prostaglandins, and adenosine. The endothelial cells that line vessels are known to secrete a variety of relaxing and constricting factors, which may contribute to autoregulation. Vasodilatory substances are washed away as blood flow increases in response to increased vessel diameter. A declining level of vasodilatory chemicals results in vasoconstriction. Thus vessel diameter is continuously adjusted according to concentrations of vasodilatory chemicals, which are directly related to the tissue's metabolic activity.

One mechanism for autoregulation of coronary blood flow involves an adenosine triphosphate (ATP)–sensitive potassium channel in vascular smooth muscle. When ATP levels rise in response to increased coronary flow, the channel closes, making it easier to depolarize the cell and contract vascular smooth muscle. Contraction of vascular smooth muscle reduces the diameter of the coronary arteries and reduces blood flow. The opposite also occurs: a reduction in ATP level, due to low flow or increased metabolism, opens the K^+ channels. Potassium then leaks out of the vascular smooth muscle and short-circuits the depolarizing influences. This inhibits vascular contraction, leading to vasodilation and increased coronary blood flow. Adenosine also contributes to regulation of the ATP-sensitive K^+ channels, causing vasodilation when adenosine levels are elevated.

Nitric oxide (NO) produced by endothelial cells lining the coronary arteries is an important regulator of coronary blood flow. NO is a diffusible gas produced by the enzyme inducible nitric oxide synthase in response to numerous stimuli, including hypoxemia and platelet factors. NO is a potent vasodilator, and inhibition of its production is associated with reduced coronary blood flow. Many known risk factors for coronary heart disease have been shown to impair NO-dependent vasodilation of coronary arteries.

Vessel diameter also is regulated by the autonomic nervous system. The coronary arteries are primarily innervated by sympathetic nerves, but they also receive a small amount of parasympathetic innervation. The sympathetic neurotransmitter norepinephrine (NE) binds to both α_1 and β_2 receptors in coronary arteries; α_1 stimulation results in vasoconstriction, whereas β_2 stimulation dilates. Under normal conditions the vasodilator response predominates, but in pathologic states, excessive α_1-mediated constriction can occur. The increased metabolic activity associated with sympathetic nervous system stimulation generally causes autoregulatory vasodilation and overrides the direct effect of NE on the vessels. Parasympathetic activity contributes to vasodilation by promoting the production of NO by coronary endothelial cells.

In addition to vessel diameter, coronary resistance is affected by myocardial contraction. During systole, cardiac muscle compression creates a marked rise in coronary resistance that reduces coronary blood flow (perfusion). Blood flow to the left ventricle is greatly decreased during systole because of the pressures generated by the thick muscular layer. Blood vessels that penetrate the myocardium to supply the innermost endocardial areas are more compressed during contraction than are outer epicardial vessels. Even though coronary artery driving pressure is greatest during ventricular systole, little blood flow reaches the left ventricle because of the high external pressure applied to the coronary vessels as the myocardium contracts. Therefore most myocardial blood flow occurs during the diastolic interval between ventricular contractions. The time the heart spends in diastole is directly related to heart rate. Faster heart rates reduce diastolic time and decrease coronary artery blood flow.

Cardiac muscle needs a continuous supply of oxygen and nutrients to perform its pumping functions. A disruption in cardiac blood flow *(ischemia)* generally results in some degree of pump failure and damage to cardiac tissues. Myocardial ischemia may be caused by conditions that reduce coronary blood flow or increase myocardial demands for oxygen. These include (1) reduced driving pressure (e.g., low aortic blood pressure or high right atrial pressure), (2) reduced vessel diameter (e.g., myocardial hypertrophy, arteriosclerosis, thrombosis, vasoconstricting chemicals), (3) reduced perfusion time (e.g., high heart rates, some dysrhythmias), and (4) increased metabolic demands (e.g., fever, sepsis, anemia).

KEY POINTS

- The right and left coronary arteries originate from the aortic root within the sinuses of Valsalva. In most people the right coronary artery perfuses the right ventricle, atrioventricular (AV) node, sinoatrial (SA) node, and right atrium.
- The left coronary artery divides into the left circumflex artery and left anterior descending artery, which perfuse the left atrium and ventricle.
- Coronary blood flow is regulated centrally by the autonomic nervous system and locally by autoregulation. The amount of coronary flow depends on

driving pressure and coronary resistance. Coronary resistance is dependent on vessel diameter.

- Adenosine and nitric oxide (NO) are two important vasodilating chemicals that are produced in response to inadequate oxygen delivery to the heart and help increase blood flow to meet metabolic demands.
- Although driving pressure is highest during systole, there is little coronary flow to the left ventricle because of vessel compression by the contracting myocardium. Most coronary blood flow to endocardial areas of the ventricles occurs during diastole.

CARDIAC MYOCYTES

Cardiac muscle cells are divided into two general types: working cells, which have primarily mechanical pumping functions; and electrical cells, which primarily transmit electrical impulses. Both types are *excitable:* they are able to produce and transmit action potentials. Working myocardial cells are packed with contractile filaments and compose the bulk of the atrial and ventricular muscle. Electrical cells function to initiate and coordinate contraction of the working cells. Differentiated cardiac myocytes are unable to enter the cell cycle to proliferate; however, they can increase in size and synthesize more contractile proteins (hypertrophy). New myocardial cells can be formed from stem cells that have the potential to divide. Stem cells may be recruited from the circulation and stimulated to divide and mature into myocytes within the myocardium. Conditions that increase myocardial cell death are thought to stimulate recruitment of stem cells into the myocardium. A high turnover rate of cardiac myocytes occurs and increases with age, suggesting that the entire population of cells within the heart is completely replaced 11 to 15 times over a lifetime. When the rate of myocardial cell loss exceeds replacement by stem cells, the condition of heart failure may ensue (see Chapter 19).

Myocyte Structure

Typical myocardial cells (myocytes) are illustrated in Fig. 17.11. Cardiac myocytes are described as muscle "fibers" because of their long, narrow shape. The plasma membrane (sarcolemma) of one cardiac cell is joined end to end with its neighbors by intercalated disks, which contain gap junctions that allow the rapid passage of electrical impulses from one cell to the next. The intercalated disks permit the many separate cells of the myocardium to function together in a coordinated manner. This arrangement is called a *functional syncytium.* The sarcolemma also forms membrane-lined channels that penetrate the cell and become the transverse tubules (T tubules) (Fig. 17.12). The T tubules permit extracellular fluid and ions to diffuse near intracellular structures. Movement of ions across the sarcolemma is an essential part of cellular excitation and the subsequent contraction of intracellular elements. Cellular contractile elements are simultaneously activated because signals at the cell surface are rapidly transmitted internally by the T tubules.

The sarcoplasmic reticulum (SR) is an extensive labyrinth of hollow membrane that stores significant amounts of intracellular calcium. It contains Ca^{2+}-sensitive channels that open briefly during depolarization and allow calcium ions to flow into the cytoplasm. An action potential traveling along the T tubule opens voltage-sensitive calcium ion channels (L type) in the plasma membrane. The Ca^{2+} ions that enter the cell through these channels interact with receptors on the SR membrane called *ryanodine receptors* (see Fig. 17.12B). Activation of these receptors opens calcium gates on the SR, and Ca^{2+} rushes into the cytoplasm to initiate contraction. The SR also contains powerful sarcoplasmic endoplasmic reticulum calcium ATPase (SERCA) pumps that recover calcium ions from the cytoplasm and return them to the SR. Inside the SR, calcium is bound to specialized proteins, including calsequestrin. This helps keep the free calcium concentration in the SR lower such that the calcium transporters have a lower gradient to pump against.

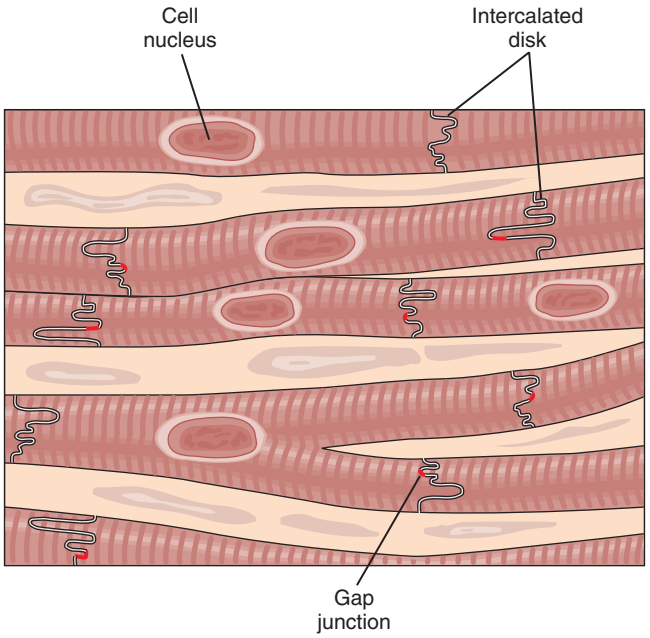


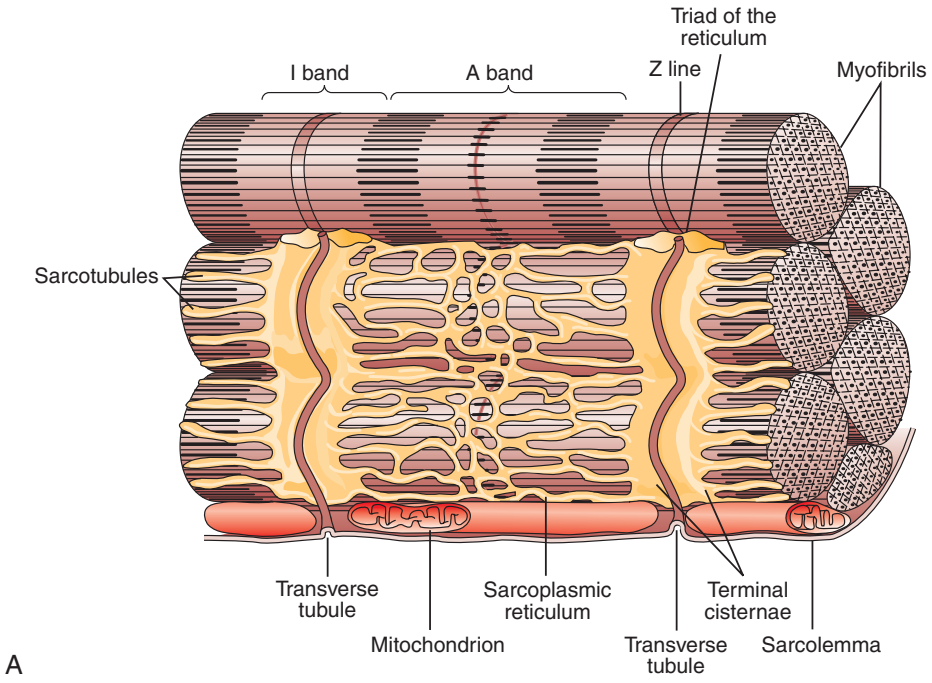
FIG 17.11 Myocardial cells, showing long narrow shape and interconnecting junctions, forming a functional syncytium. The end of one muscle cell is fused to the next by intercalated disks. Within these connections are specialized proteins that form a fluid-filled pore (gap junction) between the fused cells. Ions can travel through the gap junctions to transport changes in membrane potential from one cell to the next.

Cardiac muscle cells are packed with numerous mitochondria that are strategically positioned along the contractile fibers of the cell. The heart is also endowed with an extensive capillary network, approximately one capillary per muscle cell. The large number of mitochondria and abundant oxygen supply are necessary to keep pace with the high ATP requirements of the contractile elements and ion pumps.

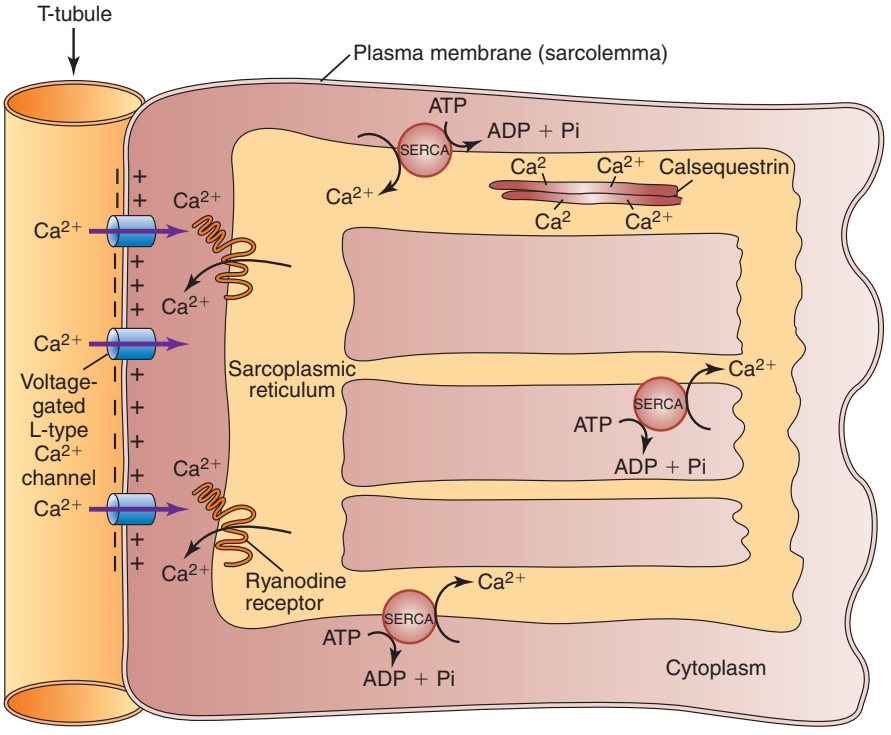
Structure of the Contractile Apparatus

Microscopic inspection of the cardiac myocyte reveals a typical pattern of banding called *striation.* This striated appearance is due to an organized structure of the proteins (myofibrils) of the contractile apparatus (Fig. 17.13). The contractile proteins, actin and myosin, are called *filaments* because they are long and narrow. Myosin filaments are larger and are referred to as *thick filaments. Thin filaments* are actually composed of several different types of protein bundled together. Actin is the primary constituent of thin filaments, with smaller amounts of the proteins tropomyosin and troponin bound to it.

The thick and thin filaments are specifically arranged in contractile units called *sarcomeres* (Fig. 17.14). Sarcomeres are defined by dark bands called *Z disks* (also called Z lines), which lie perpendicular to actin and myosin filaments. A sarcomere extends from one Z disk to the next. Thin actin filaments are attached to Z disks and extend from them. The I bands (isotropic) are light in color and correspond to the position of thin actin filaments extending in both directions from the Z disk. Thick myosin filaments lie parallel to and between the thin filaments. They are held in place by a very large and elastic protein called *titin* that extends from the Z disk to the center of the sarcomere. Each myosin filament is surrounded by six thin filaments (see Fig. 17.14). An efficient, synchronized contraction is enhanced by this precise arrangement of contractile elements.



A



B

FIG 17.12 A, Schematic diagram of a portion of a cardiac myocyte showing the transverse tubules (T tubules), which extend at right angles from the plasma membrane (sarcolemma) into the cell interior. The T tubules are extensions of the plasma membrane that bring the extracellular fluid into juxtaposition with the terminal ends of the sarcoplasmic reticulum (SR). The T tubule with the SR on either side of it is called the *triad of the reticulum*. **B,** Calcium ions that enter the cytoplasm through voltage-gated L-type channels on the T-tubule membrane interact with the ryanodine receptors on the sarcoplasmic reticulum. The activated ryanodine receptors allow calcium ions to flow into the cell cytoplasm where they initiate contraction. As soon as they are released, calcium ions are rapidly captured by the sarcoplasmic endoplasmic reticulum calcium ATPase (SERCA) pumps on the SR membrane.

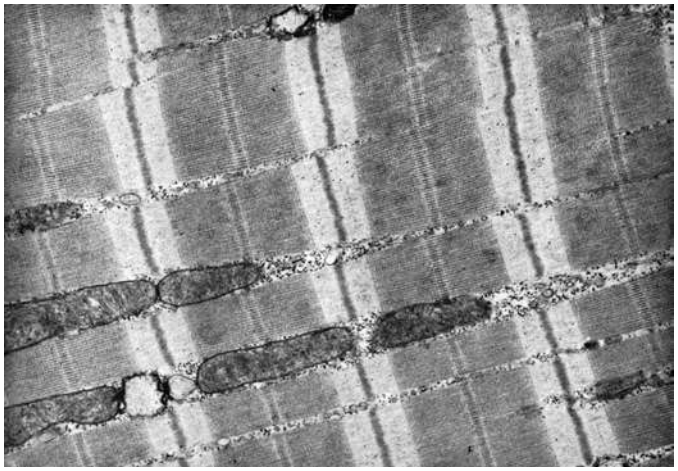


FIG 17.13 Electron micrograph of muscle fibrils showing characteristic banding pattern. The dark vertical lines are the Z disks. A sarcomere extends from one Z disk to the next. Compare with the schematic drawing in Fig. 17.14. (From Fawcett DW: *The cell,* Philadelphia, 1981, Saunders.)

Characteristics of Contractile Filaments

Myosin molecules are composed of six polypeptide chains: two heavy (H) chains and four light (L) chains. These light and heavy chains are organized into a tail region and two globular "head" areas (Fig. 17.15). The myosin heads interact with actin filaments to produce muscle contraction. Thick filaments consist of many myosin molecules with tail regions bundled together and heads protruding at intervals along the bundle. The head regions are flexible and can bend and pull on actin filaments to accomplish muscle contraction. Myosin heads are oriented in opposite directions on either side of the center tail region (see Fig. 17.15). Myosin heads have enzymatic properties and can cleave ATP to release the energy necessary for muscle contraction. Different forms of myosin have varying rates of ATP hydrolysis, which affects how quickly the muscle contracts. The serum level of thyroid hormone is known to affect the type of myosin produced in heart cells. Hyperthyroidism is associated with a fast-cycling type of myosin and hypothyroidism with a slow type of myosin. The rate of myosin cycling can also be regulated at the light chain of the myosin protein. Cellular enzymes that attach a phosphate to the light chain accelerate the rate

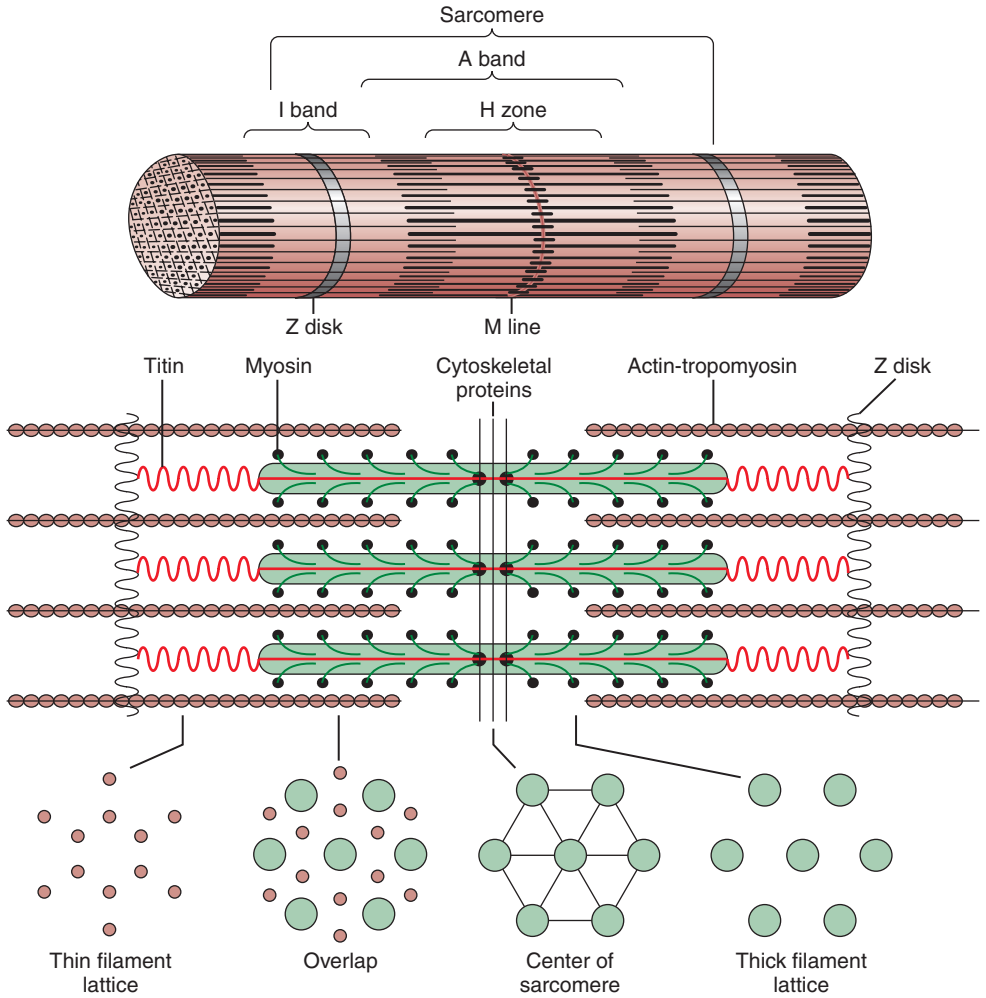


FIG 17.14 Thick and thin filaments are organized into contractile units called *sarcomeres*. A sarcomere extends from one Z disk to the next and represents the fundamental unit of muscle contraction. See text for description of bands, zones, and lines. Overlap of thick and thin filaments in each area is shown in cross-section at the bottom. Each thick filament interacts with six thin filaments that surround it.

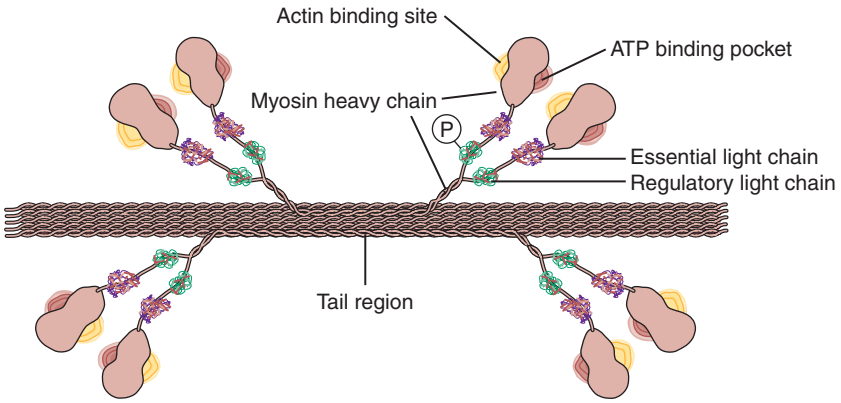


FIG 17.15 Thick filament of the sarcomere is composed of myosin proteins. Myosin head groups are oriented in opposite directions on either side of the center tail region. Phosphorylation (P) of the regulatory light chain increases myosin activity and rate of cross-bridge cycling.

of cycling. Phosphorylation is increased by activation of myocardial β_1 receptors and enhances contractility.

Thin filaments are composed of several different proteins, including actin, nebulin, tropomyosin, and troponin. Actin filaments are actually polymers of many globular actin proteins that are attached end to end, like two strings of beads, and then twisted together to form a helix (Fig. 17.16). Each of the actin beads has a site that can bind with myosin heads. Nebulin is a long protein that extends the entire length of the thin filament and is thought to regulate the length of the actin polymer such that all of the thin filaments are the same size. Tropomyosins are long, slender proteins that bind to a string of six or seven actin beads. When myocardial muscle is relaxed, tropomyosin molecules inhibit the myosin-binding sites on the actin beads. A third protein complex, troponin, is attached to the thin filament and regulates the availability of binding sites on the actin filament by controlling the position of tropomyosin. Each troponin is composed of three subunits, called troponins T, I, and C. Troponin T binds to tropomyosin, troponin I participates in the inhibitory actions of tropomyosin, and troponin C binds up to four molecules of Ca^{2+}. As described in the following section, tropomyosin and troponin are important regulatory proteins that control the activities of actin and myosin filaments. The specific isoforms (amino acid sequences) of troponins T and I present in heart tissue differ from those in other types of cells, and their presence in the serum can be used to detect myocardial infarction (see Chapter 18).

KEY POINTS

- The myocardial cells of the heart behave as a syncytium because they are joined by gap junctions within the intercalated disks that permit the flow of ions from one cell to the next.
- Myocytes are packed with actin and myosin proteins that form the contractile apparatus. The thick filaments are composed of myosin proteins. Myosin has enzymatic activity and splits ATP, releasing energy needed for movement of the filaments. Titin is a flexible protein that attaches the thick filament to the Z disk.
- The thin filament is composed of actin and two regulatory proteins: troponin and tropomyosin. At rest, tropomyosin inhibits myosin-binding sites on actin. The position of tropomyosin is regulated by the calcium-binding protein troponin. Nebulin is another actin-associated protein that helps regulate thin filament length.

MOLECULAR BASIS OF CONTRACTION

Overview of Contraction

The heart's pumping action is accomplished by the additive contractions of the many myocytes that form the cardiac chambers. Because each myocyte contributes only a small amount to overall muscle shortening, all cells of the chamber must shorten simultaneously to produce a forceful contraction. The specialized cells of the conduction system function to stimulate myocardial contraction in a coordinated way. An action potential traveling down the conduction system is the usual trigger for contraction. Cardiac myocyte depolarization causes ion channels in the plasma membrane and T tubules to open, permitting sodium and calcium entry and release of calcium from the SR. The presence of free calcium in the sarcoplasm (muscle cytoplasm) results in contraction. These events describe the process of *excitation-contraction coupling*.

Sliding Filament/Cross-Bridge Theory of Muscle Contraction

The sliding filament, or cross-bridge, theory of muscle contraction is suggested by the anatomic configuration of the sarcomere described earlier. Muscle shortening is accomplished by increasing the amount of overlap of actin and myosin filaments. The Z disks at the ends of the sarcomere move closer together as overlapping actin and myosin filaments pull past one another. Myosin heads grip binding sites on the actin beads and pull the thin filaments toward the sarcomere's center. Each time a myosin head binds an actin bead, it forms a so-called cross-bridge**.** Flexible myosin heads move in a ratchetlike manner to tug on the actin filaments (Fig. 17.17). Each ratcheting motion moves actin filaments only minutely, and many sequential cross-bridge formations are required to shorten the entire sarcomere. Thus myosin heads bend back and forth, binding and pulling on the actin filaments in a steplike fashion. Actin filaments are prevented from slipping back to their original position because some myosin–actin bonds are forming while others are disengaging. The making and the subsequent breaking of each actin–myosin cross-bridge requires one molecule of ATP. Consequently, tremendous quantities of ATP are hydrolyzed with each cardiac contraction.

ATP hydrolysis, which occurs at the myosin head region, provides the energy for contraction and affects the capability of myosin to bind actin. Myosin has two functional states or conformations: (1) a low-affinity state in which it binds weakly and (2) a high-affinity state in which it avidly binds actin. The affinity of the myosin head for actin depends on whether ATP is bound (low affinity) or ADP and inorganic

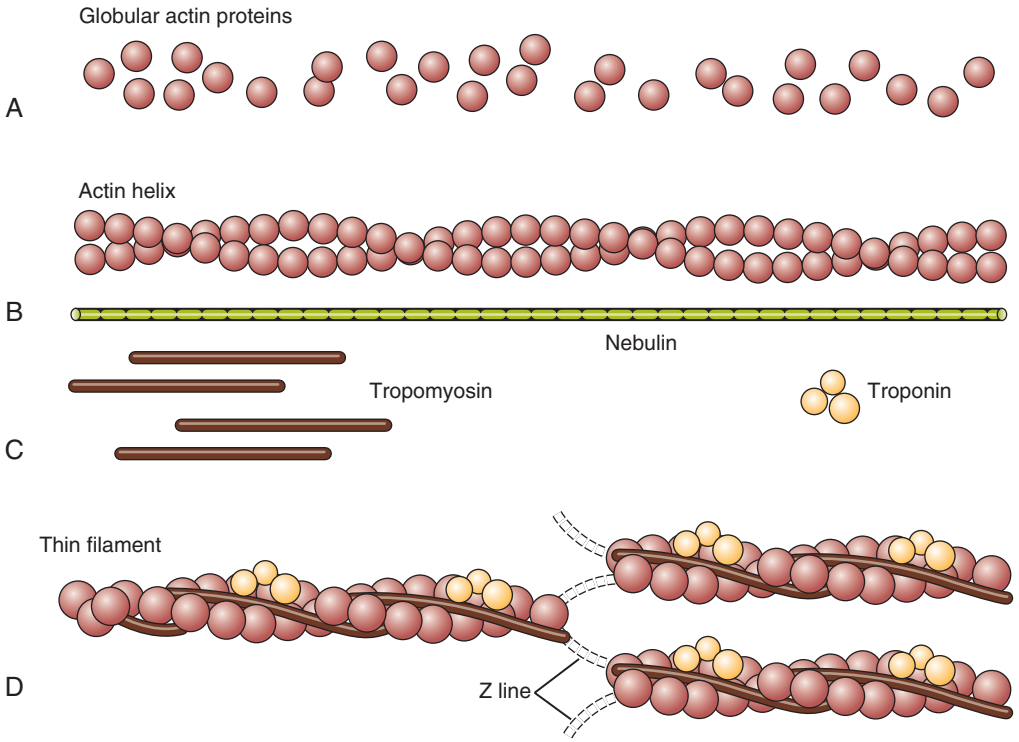


FIG 17.16 Schematic drawing of the proteins that comprise the thin filament. **A,** Globular actin proteins combine to form long double-helix filaments. **B,** Nebulin (nebulette) is a long cytoskeletal protein that extends the length of the thin filament and is thought to regulate filament length. **C** and **D,** The proteins troponin and tropomyosin combine with the actin helix to form the thin filament.

phosphate (P_i) are bound (high affinity). A proposed sequence of cross-bridge cycling is as follows (see Fig. 17.17):

1. Free myosin heads bind ATP and hydrolyze it to ADP and P_i, which remain on the myosin. Myosin heads now have a high affinity for actin and are in a high-energy conformation.
2. If binding sites on actin are accessible, myosin binds to the actin.
3. Binding results in release of ADP and P_i and a ratchet movement of the myosin as it assumes its low-energy conformation, which shortens the sarcomere (power stroke).
4. With loss of ADP and P_i, myosin can bind another molecule of ATP. The myosin heads with ATP bound now have a low affinity for actin and are released from the binding site. ATP is again hydrolyzed to ADP and P_i, and another cross-bridge cycle is initiated.

Continued cross-bridge cycling is dependent on the availability of ATP and calcium ions. A lack of ATP results in fewer cross-bridge cycles and inability of the muscle to shorten normally.

Role of Calcium in Muscle Contraction

Muscle contraction is dependent on the presence of an adequate amount of calcium ions in the cytoplasm. In the absence of free intracellular calcium, muscle contraction will not take place, even though myosin head groups have high affinity for actin-binding sites. This phenomenon can be explained in the following way. At rest, myosin heads are prevented from binding to actin by tropomyosin proteins, which inhibit actin-binding sites. The position of tropomyosin protein is controlled by troponin. When calcium is absent, troponin induces tropomyosin to inhibit the actin-binding sites. When calcium binds to troponin C, the troponin complex induces tropomyosin to move and expose the binding sites (see Fig. 17.17A and B). Cross-bridge formation immediately ensues because myosin heads have high affinity for these sites in the relaxed state. The concentration of free calcium ions in the myocardial cell determines how many actin sites are exposed and for how long and therefore determines the number of cross-bridges and extent of contraction. The release of Ca^{2+} into the cytoplasm is regulated by numerous neurotransmitters and hormones that affect contractility as described in later sections of this chapter.

Energy of Muscle Relaxation

Although muscle relaxation is generally viewed as a passive phenomenon, it actually requires significant energy to pump calcium ions out of the cytoplasm. As calcium levels fall, calcium diffuses away from the troponin molecules and tropomyosin is induced to cover the actin-binding sites. With actin-binding sites covered, myosin heads are unable to initiate cross-bridge formation, and thick and thin filaments slide back to their resting positions. Removal of calcium ions is an energy-requiring process. Membrane pumps located in the sarcolemma and SR actively move calcium out of the sarcoplasm against a concentration gradient (Fig. 17.18). The sarcolemma contains two different calcium pumps: one that requires ATP and one that uses the potential energy of the sodium gradient to remove calcium from the cell (3 Na^+ for 1 Ca^{2+}). Calcium pumps on the SR (SERCAs) require ATP. Thus energy deficiency attributable to myocardial ischemia can impair diastolic relaxation as well as systolic contraction of the heart muscle.

KEY POINTS

- Cardiac myocytes are terminally differentiated cells incapable of proliferation. New myocytes are formed from stem cells that are recruited from the circulation.
- Contraction of cardiac muscle is accomplished by shortening of individual sarcomeres. This is due to increased overlap of actin and myosin filaments. Myosin heads bind to specific sites on actin and pull the thin filaments toward the center of the sarcomere.

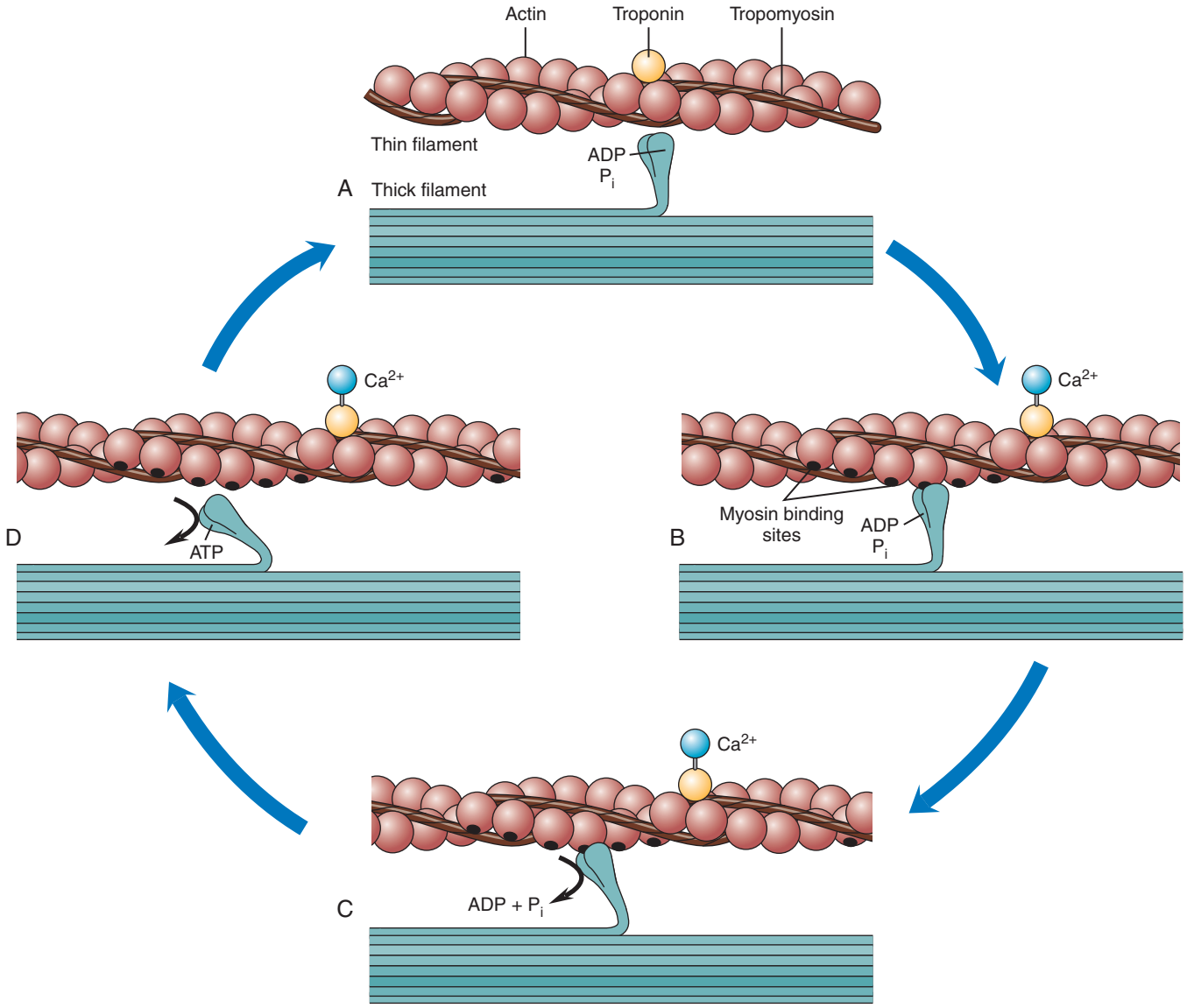


FIG 17.17 Cross-bridge cycle of muscle contraction. **A,** The myosin head has hydrolyzed its bound adenosine triphosphate (ATP) to adenosine diphosphate (ADP) and inorganic phosphate (Pi), which remain on the myosin. In this state the myosin has high affinity for actin but cannot bind because the actin-binding sites are not accessible. **B,** When calcium ions enter the cell and bind to troponin, the tropomyosin-blocking protein moves to allow myosin to bind actin, forming a cross-bridge. **C,** The act of binding changes the shape of myosin so that ADP and Pi are released. The "power stroke" is accomplished by movement of the myosin neck region. **D,** When a new molecule of ATP binds to the myosin, it changes to a low-affinity state and releases from the actin. ATP is again hydrolyzed to ADP and Pi to restart the cycle. Each cross-bridge cycle uses one ATP molecule.

- Adenosine triphosphate (ATP) hydrolysis provides the energy for cross-bridging and affects the affinity of myosin for actin. Myosin has high affinity for actin when ADP and P_i are bound and has low affinity when ATP is bound. Myosin cycles between high- and low-affinity states, making and breaking cross-bridges with the actin filament.
- The presence of intracellular free calcium ion (Ca^{2+}) is necessary for muscle contraction to occur. When Ca^{2+} is absent, actin-binding sites are inhibited and inaccessible for cross-bridging. Binding of Ca^{2+} to troponin induces the movement of tropomyosin to expose actin-binding sites and allows cross-bridge formation.
- Muscle relaxation (lusitropy) is due to removal of Ca^{2+} from the cytoplasm. This is an energy-requiring process.

CARDIAC ENERGY METABOLISM

The heart, like other tissues in the body, utilizes energy from ATP hydrolysis to drive its energy-requiring functions. Synthesis of ATP in cardiac muscle cells is accomplished by the same glycolytic and oxidative reactions described in detail in Chapter 3.

Oxygen Utilization

Because the heart is continuously active, its energy requirements are considerable. Very little ATP is stored in myocardial cells, so a continuous supply of oxygen and nutrients is necessary to support ongoing ATP synthesis. Even under normal resting conditions, the heart extracts a large portion of oxygen from the blood perfusing it. Therefore conditions

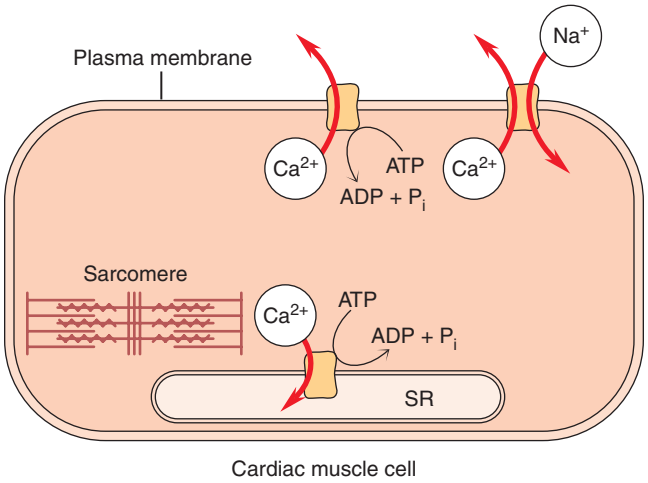


FIG 17.18 Calcium ions (Ca^{2+}) are removed from the cardiac muscle cell cytoplasm by energy-dependent protein transporters in the plasma membrane and sarcoplasmic reticulum (SR) membrane. Thus cardiac relaxation is an energy-requiring process. *ADP,* Adenosine diphosphate; *ATP,* adenosine triphosphate; *Pi,* inorganic phosphate.

of increased oxygen demand must be met by increasing the rate of coronary blood flow. When oxygen delivery is insufficient to meet requirements for oxidative phosphorylation, the cell must rely on ATP produced by glycolysis. Unfortunately, glycolysis results in production of only enough ATP to maintain the cell for seconds to minutes. In addition, anaerobic glycolysis results in local buildup of lactic acid, which may further impair cardiac performance.

Under conditions of relative ATP excess, myocardial cells are able to transfer energy to a storage form called *creatine phosphate* (CP). This transfer is accomplished by the enzyme creatine kinase (CK or CPK), in the following reaction:

$$\text{ATP} + \text{creatine} \leftrightarrow \text{ADP} + \text{CP}$$

Although amounts of cellular CP are limited, they provide an immediate source of energy when cellular ATP levels drop acutely. Under conditions of ischemia, the enzymatic reaction would proceed in reverse, utilizing CP and adenosine diphosphate (ADP) to produce ATP for immediate use by the cell. This phosphate transfer reaction is important during fluctuations in ATP supply because it does not require the presence of oxygen.

The enzyme CK is also useful in the diagnosis of myocardial cell damage. Myocardial cells that lose membrane integrity (necrosis) leak their enzymes into extracellular fluid and eventually into the bloodstream. Elevated levels of blood CK are indicative of the degree of acute myocardial cell death. Different types of tissue contain different forms of CK (isoenzymes). The myocardial band (MB) form of CK is found in cardiac muscle, and elevated serum levels of this enzyme are indicative of myocardial infarction. Other intracellular proteins, including troponin and myoglobin, are released during myocardial cell death and can be used as markers of myocardial infarction (see Chapter 18).

Substrate Utilization

The primary foodstuffs that provide fuel for energy-producing enzymatic processes in cardiac muscle are glucose and fatty acids. Amino acids are less important metabolic substrates for cardiac muscle except during states of starvation. The amount of fatty acids and glucose utilized by heart muscle cells depends on their relative concentrations in the blood. Fatty acids are the preferred fuel, particularly in a fasting state, when glucose levels are lower. Under fasting conditions, fatty acids account for approximately 85% of myocardial fuel, and glucose contributes only 15%. After eating, when blood glucose levels rise, glucose utilization may increase to about 50%. Fatty acid metabolism requires oxygen and is therefore not useful under conditions of ischemia. The heart is also able to use lactate and ketones as sources of energy when they accumulate in the circulation. Lactate is derived from pyruvate under conditions of anaerobic metabolism, whereas ketones are formed from lipid metabolism when carbohydrate supplies are low. Thus the heart is equipped to use a variety of substrates to produce ATP under varying metabolic conditions.

KEY POINTS

- Creatine phosphate (CP) is an immediately available storage form of energy. Under conditions of low adenosine triphosphate (ATP) availability, CP is converted to ATP by the enzyme creatine kinase (CK).
- The primary energy substrates for the heart are fatty acids and glucose, but the heart can utilize a variety of sources to produce energy depending on the nutrients available in the circulation.

CARDIAC ELECTROPHYSIOLOGY

The plasma membranes of cardiac cells are endowed with special ion channels that make the cells excitable. Excitable tissues are capable of generating and conducting action potentials. The heart is rhythmically activated by action potentials, which are generated and transmitted by a specialized conduction system. Spread of an action potential over cardiac muscle cell surfaces results in myocardial contraction. An understanding of the electrophysiologic properties of the heart is important because many cardiac disorders result in disturbances in electrical function that produce abnormal conduction pathways, dysrhythmias, and conduction blocks.

Cardiac Resting Potential

Like other cells, resting cardiac cells are negatively charged on the inside with respect to the outside (see Chapter 3). A difference in potassium ion concentration across the cell membrane is the primary determinant of the resting membrane potential. Atrial and ventricular muscle cells generally have a resting membrane potential of –85 to –95 mV. Pacemaker cells in the SA node are less polarized, having a resting membrane potential of about –60 mV. An increase in the concentration of extracellular potassium ion tends to hypopolarize the cell (make it less negative), and a lower-than-normal extracellular potassium concentration tends to hyperpolarize the cell (make it more negative). The degree of polarization is an important determinant of the ease with which an action potential can be initiated. Abnormalities in serum potassium level are a common source of cardiac dysrhythmias.

Cardiac Action Potential

Depolarization of cardiac cells to a threshold point results in activation of voltage-sensitive ion channels in the membrane. A myocardial action potential (Fig. 17.19) results from movement of ions through these open voltage-gated channels. The action potential in atrial and ventricular cells has five characteristic phases. Atrial action potentials are shorter in duration because they have a reduced phase 2 compared with ventricular cells.

Phase 0

Phase 0 begins when the membrane potential approaches threshold and voltage-gated "fast" sodium channels open momentarily. As a result of a steep electrochemical gradient for sodium entry, rapid influx of sodium ions occurs. Sodium entry depolarizes the cell by neutralizing

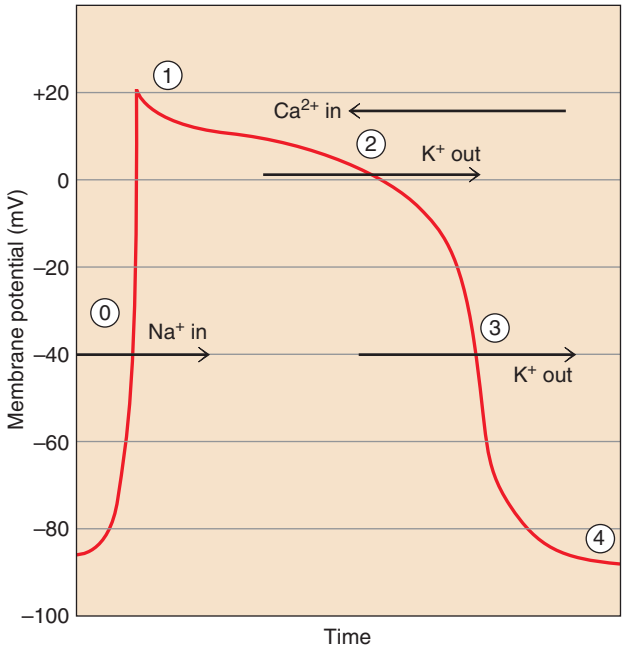


FIG 17.19 The ventricular myocardial action potential has five characteristic phases, representing changes in ion movement through the plasma membrane. Phase 0: Rapid upstroke attributable to sodium influx. Phase 1: Slight repolarization attributable to closure of sodium channels and initiation of potassium efflux. Phase 2: Plateau attributable to offsetting influx of calcium and efflux of potassium. Phase 3: Rapid repolarization attributable to closure of calcium channels and increased potassium efflux. Phase 4: Resting membrane potential reestablished attributable to closure of all voltage-sensitive channels.

the difference in charge (polarity) across the membrane. A steep depolarizing deflection (upstroke) is recorded. Class I antidysrhythmic agents such as quinidine and lidocaine block voltage-gated sodium channels and interfere with phase 0 depolarization (see Chapter 19).

Phase 1

Phase 1 is identified as a small repolarizing deflection that corresponds to closure of the fast sodium channels and transient efflux of potassium from the cell through K^+ channels. The interior of the cell is now more positively charged than at rest, which induces potassium ions to leave the cell.

Phase 2

Phase 2 is also called the *plateau phase* because little change in membrane potential occurs during this time, even though ions continue to move across the membrane. Phase 2 is primarily associated with an influx of calcium ions, which is offset by an efflux of potassium ions. The ability of K^+ to leave the cell down its electrochemical gradient is inhibited during the plateau phase by a relative inhibition of a subset of potassium channels. This is sometimes called potassium *rectification.* Chloride also may leave the cell during this phase. The voltage-gated calcium channels open and close slowly in comparison to fast sodium channels and are thus referred to as *slow channels* or *L-type channels* (long-lasting).

The calcium that enters the cell during phase 2 is linked to muscle contraction as previously described. The L-type calcium channels can be modified by agonists that prolong the open phase, such as catecholamines, and by antagonists that shorten the open phase, such as acetylcholine. Calcium channel–blocking agents (class IV antidysrhythmic agents) are used commonly in patients with cardiovascular diseases to inhibit calcium influx. β-Blockers (class II antidysrhythmics) also reduce calcium ion influx during phase 2 by indirectly inhibiting calcium channels.

Phase 3

Phase 3 is characterized by a rapid return to the resting membrane potential. This is accomplished by closure of the slow calcium channels and continued and even more rapid efflux of potassium ions from the cell through a variety of potassium channels. Sodium channels remain absolutely refractory during phases 1, 2, and early 3. The latter part of phase 3 represents a *relative refractory period,* when sodium channels may be induced to open, but a larger-than-normal depolarizing stimulus is required. If an abnormally early (premature) depolarization occurs during the relative refractory period, it will be conducted more slowly than usual because few fast Na^+ channels are ready to be activated. Slow conduction through the myocardium predisposes to cardiac dysrhythmias, such as ventricular fibrillation (see Chapter 19). Class III antidysrhythmic agents, such as amiodarone, increase the refractory period by inhibiting opening of potassium channels during phase 3.

Phase 4

Phase 4 of the ventricular myocyte action potential corresponds to the period of time between action potentials when no changes in membrane voltage are evident and the resting membrane potential is present. The resting membrane potential in ventricular myocardial cells is flat, and they do not spontaneously depolarize. In contrast, cells in the pacemaker and conduction system automatically depolarize and have a sloping phase 4. The Na^+–K^+ pumps and Ca^{2+} pumps work continuously throughout all phases to reestablish the internal and external concentrations of sodium, potassium, and calcium ions.

Rhythmicity of Myocardial Cells

Rhythmicity and automaticity refer to regular, spontaneous generation of action potentials. Rhythmic pacemaker cells have a recognizable action potential that is characterized by a sloping phase 4 (Fig. 17.20), in contrast to the flat phase 4 of ventricular muscle cells. A requirement for rhythmicity is that the cell membrane has channels that automatically open during phase 4. These channels begin to open as the membrane potential becomes more negative during the repolarization phase. Progressive channel opening makes the pacemaker cells leaky to Na^+, Ca^{2+}, and K^+. Gradually the flow of positive ions into a cell offsets the repolarizing currents and depolarizes the membrane, resulting in generation of an action potential. One contributor to the automatic depolarization during phase 4 is the I_f channel. The I_f channels originally were named for a "funny" current and later discovered to be sodium channels that are activated by membrane repolarization. Channels that allow calcium and potassium leakage are also operative during phase 4 in pacemaker cells. Late in phase 4, an increase in calcium ion influx occurs through voltage-gated calcium channels called *T type,* for "transient." These channels open and close more quickly than the L-type calcium channels that open during the action potential. Spontaneous release of Ca^{2+} from the SR also contributes to depolarization by activating the $3Na^+/Ca^{2+}$ exchanger and promoting Na^+ influx. Many of these channels can be regulated by various means, including autonomic neurotransmitters, in order to change heart rate. An action potential is initiated when phase 4 depolarization reaches the threshold for the opening of voltage-gated, L-type, slow calcium channels. Repolarization is achieved in large part by an exodus of potassium ions from the cell.

The rate of rhythmic discharge is determined by the relative influx of Na^+ and Ca^{2+} versus the efflux of K^+. In a normal heart, a cell with the fastest rate of spontaneous depolarization becomes the pacemaker

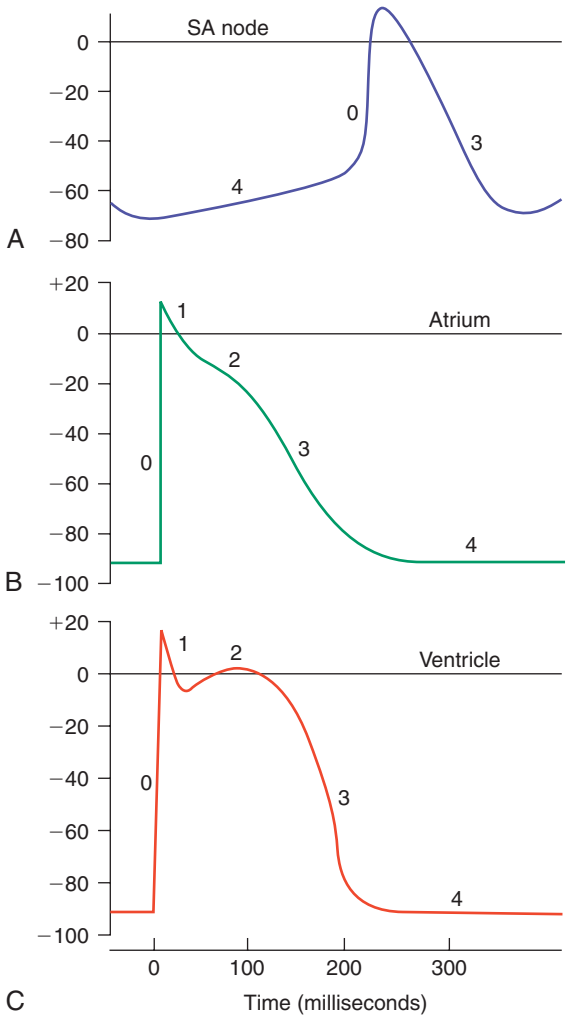


FIG 17.20 Rhythmic cells **(A)** have a sloping phase 4, in contrast to the flat phase 4 of the atrial **(B)** and ventricular **(C)** muscle cells. Spontaneous depolarization during phase 4 allows pacemaker cells to develop action potentials automatically. *SA,* Sinoatrial. (Adapted from Hoffman BF, Cranefield PF: *Electrophysiology of the heart,* New York, 1960, McGraw-Hill.)

for the rest of the heart. Cells in the SA node, located in the right atrium, generally function as the heart's pacemaker because they have the fastest rate of spontaneous depolarization. However, other cells in the conduction system are also capable of spontaneous depolarization and may initiate an action potential in certain circumstances.

The steepness of the slope of phase 4 depolarization determines the rate of action potential generation and therefore heart rate. Several factors determine the steepness of the slope, including membrane permeability to sodium, calcium, and potassium. For example, an increase in potassium ions leaving the cell would slow depolarization and result in a slower rate. Rhythmicity may be influenced by the autonomic nervous system, drugs, and electrolyte balance. These conditions are discussed in the following sections.

Specialized Conduction System of the Heart

Some myocardial cells are specialized to conduct action potentials throughout the heart in an organized and rapid manner. These cells constitute the conduction system of the heart, as shown in Fig. 17.21. Normal excitation of the heart follows a pathway beginning with the SA node, atrial internodal pathways, AV node, bundle of His, ventricular bundle branches, and, finally, Purkinje fibers.

The SA node is located in the right atrium near the superior vena cava inlet. It receives innervation from sympathetic and parasympathetic branches of the autonomic nervous system. The SA node generally serves as a pacemaker for the heart, generating about 70 (range, 60 to 100) action potentials per minute in a resting adult. SA action potentials are spread contiguously through gap junctions to adjacent atrial cells at a fast rate. A fibrous skeleton separates atria from ventricles and prevents the spread of impulses from atrial cells to ventricular cells. Several small bundles of atrial muscle cells conduct impulses slightly faster than the usual atrial cell. One such bundle, the anterior interatrial band (Bachmann bundle), conducts impulses from the SA node to the left atrium. Atrial depolarization results in atrial contraction, which increases the volume of blood delivered to the still-relaxed ventricular chambers.

After traversing the atria, the impulse initiated at the SA node arrives at the AV node (AV junction) located in the posterior septal wall of the right atrium just behind the tricuspid valve. There is a characteristic slowing of impulse conduction through the AV node, which allows for completion of atrial contraction before initiation of ventricular systole. The AV node is actually composed of different types of fibers that have somewhat different action potential conduction times. Overall, it normally takes about 0.13 second for an impulse to pass through the AV node. The AV node is richly innervated by the autonomic nervous system. The AV node spontaneously depolarizes at a rate of 40 to 60 times per minute and usually becomes the heart's pacemaker if the SA node fails.

Purkinje cells (fibers), which lead from the AV node to ventricular myocardium, are vastly different from AV nodal cells. They are large and well structured to conduct impulses very rapidly. After penetrating the AV fibrous barrier, the bundle of Purkinje fibers travels 5 to 15 mm down the intraventricular septum toward the apex. The main bundle then divides into left and right bundle branches, which travel down the left and right sides of the intraventricular septum. Successive branches of Purkinje fibers penetrate the ventricular muscle mass from the endocardial side. Intraventricular septal areas are depolarized first, followed by apical muscle, and finally the lateral walls. Early septal depolarization allows the septum to contract first and provide a stable wall against which the left and right ventricles can contract. The total time elapsed between main bundle branch and terminal Purkinje fiber depolarization is only 0.03 second. Therefore the entire ventricular endocardium is activated almost simultaneously. Purkinje fibers are capable of spontaneous depolarization at a rate of 15 to 40 times per minute and may become pacemakers for the heart if impulses from the SA and AV nodes are interrupted.

Action potentials are rapidly transmitted from the terminal Purkinje fibers to cardiac muscle fibers and then spread contiguously from cell to cell through gap junctions in the ventricular muscle. Approximately 0.03 second is required for the impulse to be transmitted through the ventricular myocardium. Impulses normally travel from the terminations of Purkinje fibers at endocardial surfaces toward the epicardial surfaces. Depolarization of the right ventricle is accomplished slightly sooner than the left because of differences in muscle mass. Depolarization of the ventricular myocardium is followed by contraction and ejection of blood from the ventricles.

The capability of faster pacemakers to suppress the automatic discharge of slower pacemakers is called *overdrive suppression.* A slower pacemaker may be revealed if the normal pacemaker is suddenly interrupted. Sometimes it takes time for the slower pacemaker to "kick in" and begin pacing at its intrinsic rate. A previously rapid rate of depolarization apparently enhances the activity of membrane Na^+–K^+ pumps, resulting

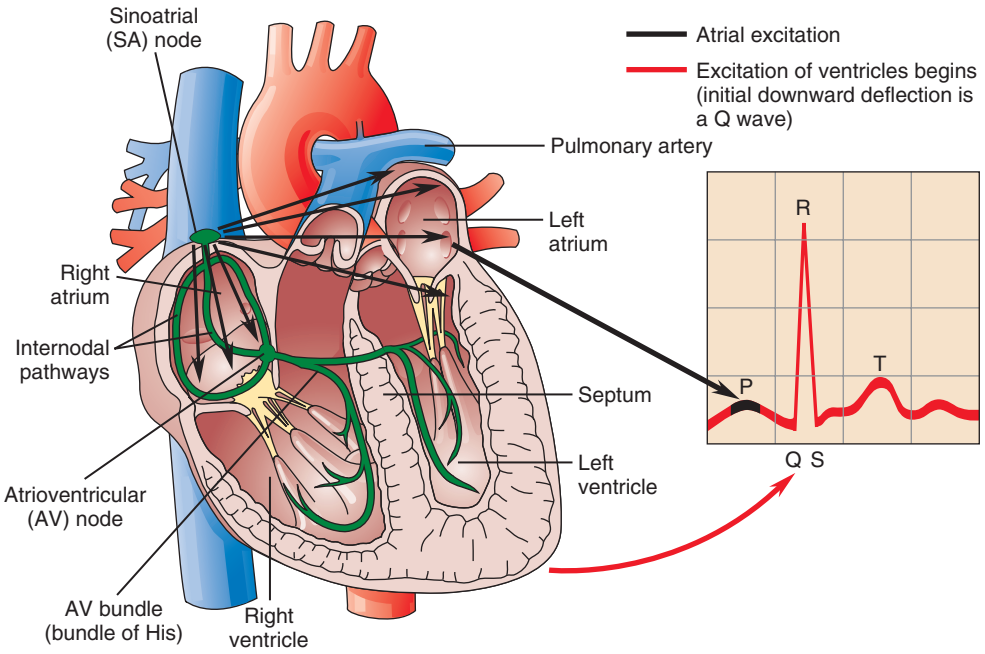


FIG 17.21 Schematic drawing of the conduction system of the heart. An impulse normally is generated in the sinus node and travels through the atria to the AV node, down the bundle of His and Purkinje fibers, and to the ventricular myocardium. Recording of the depolarizing and repolarizing currents in the heart with electrodes on the surface of the body produces characteristic waveforms.

in a period of hyperpolarization (more negative resting potential) when the faster pacemaker suddenly stops. Thus it takes slightly longer to reach threshold and initiate the first action potential.

Autonomic Regulation of Rhythmicity

Both sympathetic and parasympathetic nerves supply the heart. Sympathetic innervation is widespread to all areas, including the ventricular myocardium. Parasympathetic innervation, by way of the vagus nerves, is localized primarily in SA and AV nodal areas. The right vagus nerve supplies the SA node, whereas the left vagus nerve supplies the AV node. The autonomic nervous system exerts control over heart rate and velocity of impulse conduction. In general, sympathetic activation increases heart rate *(chronotropic effect)* and increases speed of conduction *(dromotropic effect)*, as well as inducing heart muscle to contract more forcefully *(inotropic effect)* and relax more quickly *(lusitropic effect)*. These effects are achieved by release of NE from sympathetic nerve endings. Binding of NE to β receptors on heart muscle cell membranes increases production of cAMP, which regulates several membrane channels and pumps and increases depolarizing ion currents.

Parasympathetic stimulation primarily results in a reduction in heart rate and speed of action potential conduction. Acetylcholine is the neurotransmitter released by parasympathetic nerve endings. Acetylcholine binding to muscarinic receptors on heart cells inhibits cAMP production and increases membrane permeability to potassium ions, allowing them to leak from the cell. The resulting hyperpolarization makes it more difficult to reach threshold and initiate an action potential. The resting heart is normally under a predominant parasympathetic influence, which results in an SA discharge rate of about 70 beats/min. If parasympathetic activity is blocked, the spontaneous discharge rate of SA nodal cells increases to about 100 beats/min. An increase in vagal activity can reduce heart rate significantly. Breath holding, bearing down during defecation, and pressing on the carotid arteries may increase vagal tone and reduce heart rate. This is sometimes called a *vasovagal response* and may lead to dizziness and fainting.

KEY POINTS

- The cardiac resting membrane potential is about –90 mV. The resting membrane potential is determined by the ratio of intracellular to extracellular K^+ concentration.
- The five phases of the cardiac action potential are due to changes in ion conductance through the plasma membrane. The main changes in ion conductance result from opening of the fast Na^+ channels (phase 0), slow Ca^{2+} channels (plateau), and K^+ channels (repolarization) in the plasma membrane.
- Spontaneous generation of action potentials in automatic cells is due to a progressive leak of Na^+ and Ca^{2+} into the cell via channels that automatically open during repolarization. The rate of cation leak determines the rate of pacemaker discharge. Parasympathetic influence increases K^+ efflux and slows the rate. Sympathetic influence increases influx of Na^+ and Ca^{2+} and increases the rate.
- The usual conduction pathway for depolarization of the heart begins at the SA node, progresses through the atria, enters the AV node, and activates the bundle of His and Purkinje fibers. Purkinje fibers leave the AV node and divide into left and right branches that innervate the endocardial surface of the ventricular myocardium.

ELECTROCARDIOGRAPHY

As action potentials spread from cell to cell throughout the myocardium, an electrical current is transmitted to the body surface and can be detected by electrodes placed on the skin. A recording of these electrical currents is called an *electrocardiogram* (ECG). The ECG is a useful indicator of abnormalities of the heart's conduction system. Irregularities in initiation of impulses, conduction rates, and conduction pathways can be identified. The ECG has a different appearance from the cardiac action potential described previously because it registers depolarizing and repolarizing currents in the whole heart rather than the activity of individual myocytes (Fig. 17.22).

Electrical currents traveling through the heart have both direction and magnitude and are often described as *vectors*. At any instant, electrical currents are moving in various directions through different regions of the heart. Waveforms recorded at the ECG electrodes are algebraic sums of all of these vectors. Patterns of electrical activity shown on the ECG vary according to the placement of electrodes on the body. In general, a wave of depolarization moving toward a positive recording electrode will register as an upward deflection on the ECG. A wave of repolarization moving away from a positive electrode also will register as an upward deflection on the ECG. A downward deflection results from a wave of depolarization moving away from a positive electrode (Fig. 17.23). Placement of a recording electrode on the lower-left extremity (lead II) results in the typical ECG pattern shown in Fig. 17.24. A description of the usual electrode placements is included in the section "Tests of Cardiac Function" at the end of this chapter.

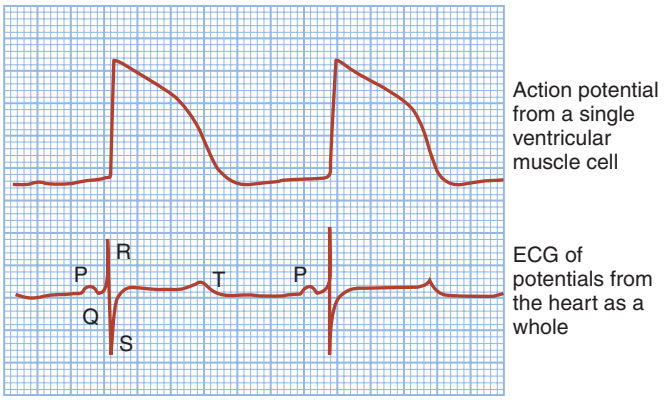


FIG 17.22 Comparison of the action potential from a single ventricular muscle cell, showing rapid depolarization and prolonged repolarization phases, against an electrocardiogram of potentials from the heart as a whole. Ventricular myocytes remain depolarized and refractory throughout the entire QT interval.

Each deflection on the ECG has a normal characteristic shape and time interval (see Fig. 17.24). The three major wave complexes are the *P wave*, which corresponds to atrial depolarization; the *QRS complex*, which represents ventricular depolarization; and the *T wave*, which reflects ventricular repolarization. The PR interval, between the beginning of the P wave and the beginning of the QRS complex, includes atrial, AV node, and His–Purkinje fiber depolarization. The normal sequence

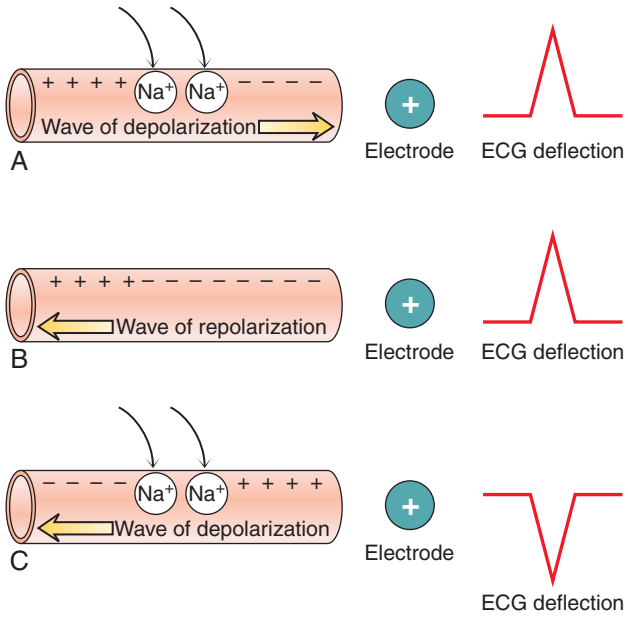


FIG 17.23 Electrocardiographic waveforms may be positive (upward) or negative (downward), depending on the location of electrodes on the chest. **A,** A wave of depolarization moving toward a positive electrode results in a positive deflection. **B,** A wave of repolarization moving away from a positive electrode results in a positive deflection. **C,** A wave of depolarization moving away from a positive electrode results in a negative deflection.

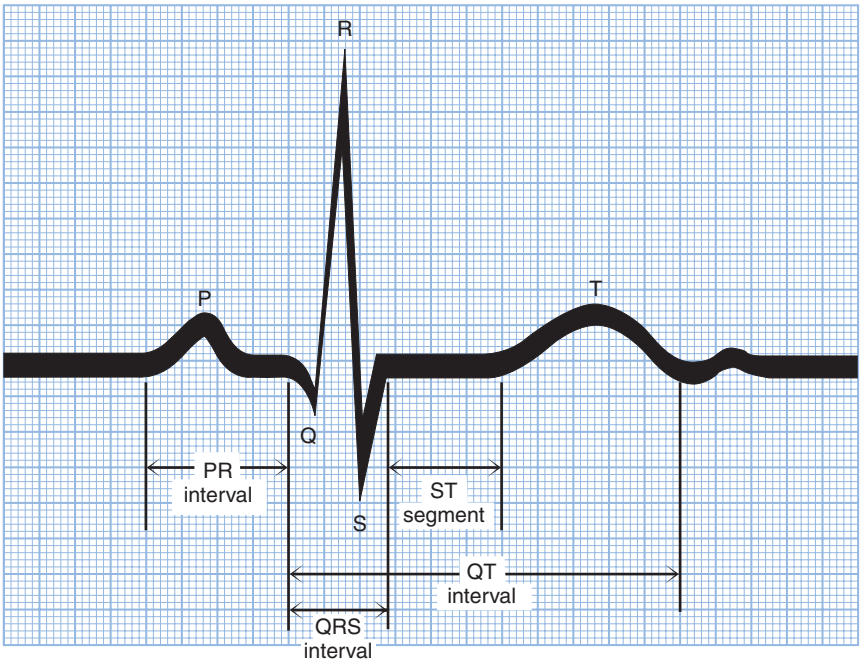


FIG 17.24 Usual electrocardiographic pattern recorded from lead II, showing characteristic waves and intervals.

of ventricular depolarization begins with the septum, followed by the apex, and finally the base of the ventricular walls. Septal depolarization begins on the left septal surface and then travels toward the right, resulting in a small negative deflection, the *Q wave* in lead II (Fig. 17.25). A large upright *R wave* corresponds to a wave of depolarization traveling down the ventricles toward the apex. Depolarization of the ventricular base, because it moves in a direction away from the lower limb electrode, is recorded as a negative *S wave*. The ST interval, between the S wave and the beginning of the T wave, is isoelectric (flat) because the entire ventricle is depolarized and no detectable current is flowing. The QT interval, from the beginning of the QRS complex to the end of the T wave, is commonly measured as an indicator of ventricular systole. The T wave is normally upright in lead II, representing a wave of repolarization moving away from a positive electrode. In some patients, particularly those with slow heart rates, the T wave is followed by a small positive deflection, called a *U wave*. Prominent U waves also are a sign of a low potassium level. Abnormalities in any time intervals may indicate abnormal conduction pathways and enhanced or slowed conduction times. Rhythm disturbances are discussed in detail in Chapter 19.

KEY POINTS

- The Electrocardiogram (ECG) represents an algebraic sum of all depolarizing and repolarizing currents occurring in the heart. ECGs are useful for detecting conduction and rhythm disturbances.
- The major deflections of the ECG are:
  P wave: atrial depolarization
  PR interval: atrial, AV node, and Purkinje depolarization
  Q wave: septal depolarization
  R wave: apical depolarization
  S wave: depolarization of lateral walls (base)
  T wave: ventricular repolarization

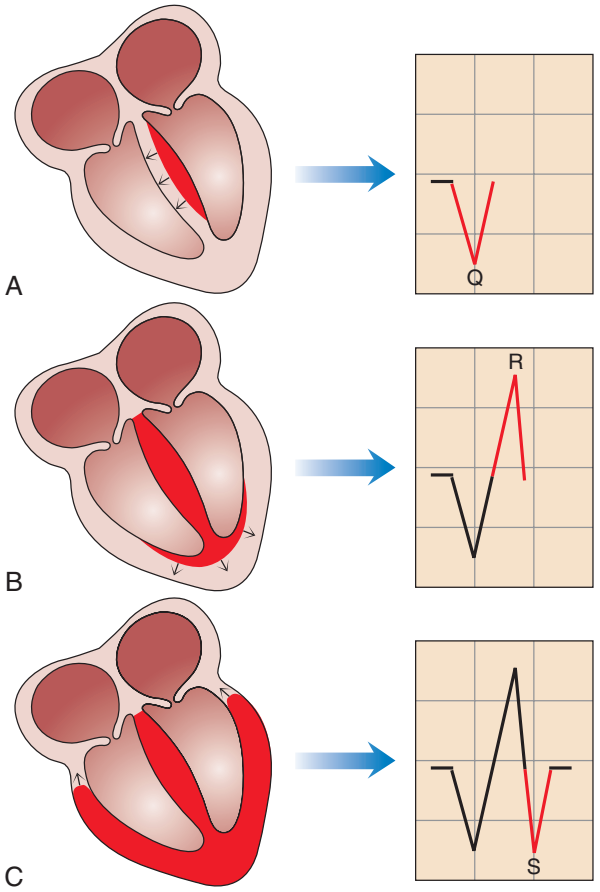


FIG 17.25 QRS complex results from the sequence of ventricular depolarization. **A,** In lead II, septal depolarization is in a direction away from the positive electrode, resulting in a negative Q wave. **B,** Depolarization of the apex of the heart is in a direction toward the positive electrode, resulting in a large positive R wave. **C,** Depolarization of the lateral walls and base of the ventricles is in a direction away from the positive electrode, resulting in a negative S wave.

DETERMINANTS OF CARDIAC OUTPUT

Cardiac output is a measure of the amount of blood pumped out of the heart each minute. Because the heart's primary function is to pump enough blood to circulate oxygen and nutrients to tissues, the cardiac output is an extremely important indicator of cardiovascular health. Normal resting cardiac output is approximately 5 to 6 L/min, but it varies with body size and age. Cardiac output is often indexed to body surface area in an attempt to adjust for these differences (cardiac index = cardiac output/body surface area). A normal cardiac index ranges from 2.8 to 3.3 L/min/m^2. Regardless of the actual number of liters of blood pumped per minute, the adequacy of tissue perfusion is ultimately important.

Cardiac output is a product of heart rate and stroke volume (CO = HR × SV). Stroke volume refers to the amount of blood ejected from the ventricle with each contraction. An increase in heart rate (to a point) and/or an increase in stroke volume will result in a greater cardiac output. Conversely, a low heart rate and/or a decreased stroke volume will cause cardiac output to fall. To a certain extent, a change in one factor can be compensated for by a change in the other, thus maintaining cardiac output at a constant level. For example, it is common for an individual with limited stroke volume attributable to cardiac disease to have a high resting heart rate. Any physiologic, pharmacologic, or pathologic process that alters heart rate or stroke volume may affect cardiac output and therefore tissue perfusion.

Determinants of Heart Rate

Heart rate is primarily influenced by the autonomic nervous system. Release of NE by sympathetic nerve endings results in an increased heart rate. A similar effect results from circulating NE and epinephrine released from the adrenal gland during sympathetic stimulation. Sympathetic activation of the heart is regulated by several reflex pathways that constantly monitor blood pressure and metabolic activity in the body. In general, detection of inadequate blood pressure, a lack of oxygen, or a buildup of metabolic end products results in activation of the sympathetic nervous system. Specialized sensory nerve endings, called *baroreceptors*, located in the aortic arch and carotid arteries respond to changes in blood pressure and transmit this information to the central nervous system (CNS) by way of cranial nerves IX and X. A decline in blood pressure causes parasympathetic system inhibition and cardiac sympathetic nerve activation, resulting in a rise in heart rate. Conversely, a rise in blood pressure causes the heart rate to fall because of parasympathetic activation and sympathetic inhibition. Under normal resting conditions, the heart rate is under parasympathetic influence, with a usual rate of approximately 70 beats/min.

In addition to baroreceptors, other sensory fibers that detect pressure are located in the cardiac chambers. These sensory receptors respond to changes in intrachamber pressure, which reflect the volume of blood in the chamber. Atrial or ventricular overdistention suppresses parasympathetic influence and increases heart rate (Bainbridge reflex). Heart rate may also be influenced by higher CNS activities that do not involve reflex pathways. Anxiety, fear, stress, excitement, trauma, and fever may

activate the sympathetic system, for example. A variety of drugs can mimic or block the effects of both sympathetic and parasympathetic systems and therefore influence heart rate (see Chapter 18).

In general, an increase in heart rate results in an increase in cardiac output; however, at very high heart rates, cardiac output may actually fall. At high heart rates (e.g., more than 200 beats/min in the young, even lower in the adult), the time for diastolic ventricular filling can be significantly reduced, resulting in a low stroke volume. The benefit of increased heart rate is therefore undermined by impaired pumping efficiency.

Determinants of Stroke Volume

Three major factors influence stroke volume: (1) the volume of blood in the heart (preload), (2) the contractile capabilities of heart muscle (contractility), and (3) the impedance opposing ejection of blood from the ventricle (afterload). Each of these factors is in turn influenced by many other physiologic, pharmacologic, and sometimes pathologic variables.

Volume of Blood in the Heart (Preload)

The heart can only pump as much blood as is delivered to it by the circulatory system. Blood returning to the heart from the circulation is often called *venous return*. Normally, venous return is equal to cardiac output because the circulatory system is just that—a circuit. However, there may be inequalities over several heartbeats when changes in blood volume or blood distribution occur. The heart is well suited to adjust to these beat-to-beat changes in venous return such that the healthy heart pumps essentially whatever amount is delivered to it.

The amount of blood present in the ventricles just before contraction (end-diastolic volume) is an important determinant of stroke volume. The relationship between diastolic volume and the force of myocardial contraction is known as the Frank–Starling law of the heart. In essence, this law states that an increase in resting muscle fiber length results in a greater development of muscle tension. Ventricular muscle fiber length is determined by the volume of blood it contains, commonly called the *preload*. An increase in preload results in a greater force of contraction and a larger stroke volume. In this way, the ventricle is able to adjust its stroke volume, beat by beat, according to the amount of blood to be pumped.

The Frank–Starling law of the heart (also called the *length–tension relationship*) may be understood by recalling the molecular structure of contractile units of heart muscle. For contraction to occur, the actin and myosin filaments that make up the sarcomere must form cross-bridges and slide together. Stretching the muscle before contraction is believed to optimize the space between the actin and myosin filaments, bringing them closer together and resulting in more cross-bridge formation (Fig. 17.26). Stretching the muscle before contraction also makes the contractile apparatus more sensitive to calcium ions such that a greater contractile force occurs for a given calcium concentration.

The cardiac function curve describes the effects of preload on ventricular stroke volume (Fig. 17.27). In practice, stroke volume and ventricular end-diastolic volume are difficult to measure, and other indicators, such as ventricular pressure and cardiac output, may be used. Cardiac function curves can be measured in persons with poorly functioning hearts to determine the best filling volume (preload) for optimizing cardiac output. Often, the failing heart requires a higher-than-normal preload to maintain a normal cardiac output. However, there are limits to the improvement in stroke volume with increased diastolic filling, and beyond that point the curve will flatten. On the flat part of the curve, an increase in preload increases the workload of the heart, but does not provide an improvement in output. The workload imposed on the heart chambers by preload is sometimes called the *volume work of the heart*. An increase in preload increases the volume work of the heart, which must be met by increased oxygen uptake to maintain adequate ATP production.

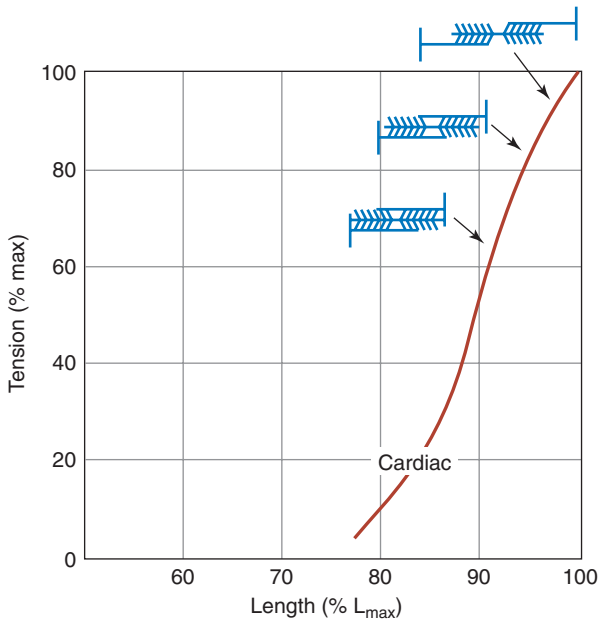


FIG 17.26 Force of muscle contraction depends in part on its resting length before activation. At optimal lengths, the greatest tension is developed, and cross-bridge formation is enhanced.

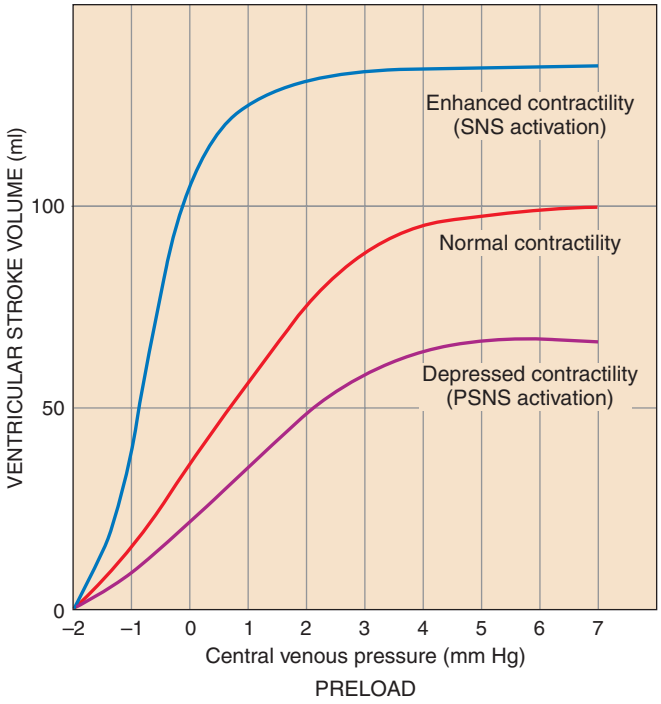


FIG 17.27 Cardiac function curves showing the dependence of ventricular stroke volume on preload. Different hearts have different cardiac function curves and may respond differently to the same degree of preload. *PSNS,* Parasympathetic nervous system; *SNS,* sympathetic nervous system.

Contractile Capabilities of the Heart (Contractility)

Heart muscle contractility depends on several factors, including (1) the amount of contractile proteins in the muscle cells, (2) the availability

of ATP, and (3) the availability of free calcium ions in the cytoplasm. Contractility is, by definition, independent of fiber end-diastolic length and is therefore not affected by preload. Given an adequate ATP supply, the contractile state of the normal myocardium is primarily determined by factors that increase the availability of free calcium ions within the myocardial cell. In general, an increased intracellular free calcium level can be accomplished by enhanced release from internal stores, enhanced entry from extracellular fluid, and reduced rates of extrusion across the plasma membrane.

A variety of agents that increase contractility, called *positive inotropes,* are associated with increased intracellular calcium levels in the heart. These include the sympathetic neurotransmitters NE and epinephrine, thyroid hormone, caffeine, digitalis, and many others. Agents that depress contractility, called *negative inotropes,* achieve their effects by reducing intracellular calcium levels. These agents include L-type calcium channel blockers, parasympathomimetics, and sympathetic blocking drugs. The baroreceptor reflex, described previously in relation to heart rate, is also an important regulator of stroke volume through its effects on contractility. Positive inotropic agents increase ATP utilization by the myocardium, whereas negative inotropes decrease myocardial workload and reduce ATP requirements.

Cardiac disease may adversely affect contractility because of an inadequate oxygen supply or because of loss of myocardial pumping cells. These disorders are discussed in Chapter 18.

Impedance to Ejection From the Ventricle (Afterload)

The third major determinant of stroke volume is afterload, which refers to the impedance or resistance that must be overcome to eject blood from the chamber. Left ventricular afterload is determined primarily by aortic blood pressure. Because high blood pressure increases left ventricular afterload, vasodilating agents that reduce blood pressure can significantly decrease afterload. Normally the aortic valve offers little impedance to flow; however, aortic valve narrowing may significantly increase afterload. An increase in afterload will result in a decrease in stroke volume unless contractility or preload (or both) is adjusted to compensate. Conversely, a decrease in afterload will allow a larger-than-normal volume of blood to be ejected from the heart, requiring less myocardial work. The work done by the heart to overcome afterload is often called the *pressure work of the heart.* An increase in afterload increases pressure work and requires greater tension development within the walls of the chamber (wall stress). Increased wall stress not only increases myocardial workload and oxygen consumption but also, if prolonged, may trigger structural changes leading to hypertrophy of myocytes.

The ventricles normally eject about 60% to 70% of their end-diastolic volume during contraction; the remaining 30% to 40% remains in the ventricle. Ejection fraction is influenced by afterload as well as preload and contractile state. A reduced ejection fraction is a common finding in persons suffering from myocardial infarction. Ejection fractions less than 40% indicate significant myocardial impairment and may be associated with systolic heart failure (see Chapter 19).

Cardiac Workload

The oxygen requirements of the heart are related to the amount of energy (ATP) exerted to perform its pumping function. The four determinants of cardiac output described in the previous section—heart rate, preload, contractility, and afterload—are also the major determinants of cardiac energy requirements. An increase in any of these four factors will increase ATP requirements and therefore cardiac cell oxygen requirements. High afterload is most detrimental because it greatly increases cardiac work without producing a higher cardiac output. When oxygen supply to the heart is impaired, as in coronary atherosclerosis, it may be beneficial to reduce myocardial oxygen demand by reducing cardiac workload. This may be accomplished by reductions in heart rate, preload, afterload, and contractility.

KEY POINTS

- Cardiac output is the product of the heart rate times the stroke volume (CO = HR × SV). An increase in heart rate or stroke volume will increase cardiac output. An increase in heart rate can compensate for a decrease in stroke volume.
- Heart rate is controlled primarily by the autonomic nervous system. Factors that increase heart rate include low blood pressure (baroreceptors), acidemia (chemoreceptors), atrial and ventricular overdistention (Bainbridge reflex), and emotions. Activation of the vagus nerve will decrease heart rate.
- Stroke volume is influenced by preload. According to the Frank–Starling law, increased preload stretches the sarcomere, resulting in more forceful contraction.
- Increased contractility increases stroke volume by causing a greater percentage of the ventricular volume to be ejected. Any factor that enhances the availability of cytoplasmic free Ca^{2+} will increase contractility.
- Increased afterload will decrease stroke volume. Afterload is determined primarily by the resistance of the arterial system. Vasoconstriction and high aortic pressure increase afterload.
- Any factor that increases heart rate, preload, contractility, or afterload will increase the workload of the heart.

ENDOCRINE FUNCTION OF THE HEART

In addition to its pumping function, the heart has an endocrine function: secretion of natriuretic peptides. Atrial natriuretic peptide (ANP) is synthesized by myocytes in the atria and released in response to atrial stretch. Increased atrial stretch occurs when blood volume becomes excessive. The ventricles produce a related peptide called *B-type natriuretic peptide (BNP)* when they are chronically overdistended. An elevated BNP value is a marker for congestive heart failure. ANP and BNP cause enhanced excretion of sodium and water by the kidney. In general, the effects of the natriuretic peptides are antagonistic to those of the renin–angiotensin–aldosterone system (see Chapter 26).

TESTS OF CARDIAC FUNCTION

In addition to patient history, laboratory results, and physical assessment, a number of diagnostic tests may be employed to evaluate cardiac function. The ECG is routinely obtained and provides information about the heart's conduction patterns. Echocardiography and nuclear cardiography are tests that use various modes to image the heart. A more direct assessment of cardiac function can be obtained by cardiac catheterization. In addition, a number of methods have been developed to quantify myocardial blood flow. Each of these studies is briefly described in this section.

Electrocardiography

The ECG graphically indicates electrical currents generated by cardiac cells. The current is registered by skin electrodes placed in particular positions on the body. The standard ECG has 12 different leads that are obtained through 10 skin electrodes: 3 standard bipolar limb leads, 3 augmented unipolar limb leads, and 6 unipolar chest leads. Bipolar leads represent a difference in electrical potential between two electrodes, one positive and one negative. Augmented unipolar limb leads represent a difference in potential between one electrode and the average of the other two limb electrodes. Unipolar chest leads represent a difference

in potential between the chest electrode and a location at the center of the heart. Each lead provides a different ECG because of its particular "view" of current flow through the heart.

The three standard bipolar limb leads are lead I, lead II, and lead III (Fig. 17.28): lead I measures the current between the right arm and left arm, lead II measures the current between the right arm and left leg, and lead III measures the current between the left arm and left leg. A normal ECG from leads I, II, and III is illustrated in Fig. 17.29.

Electrode placement for the augmented unipolar limb leads is illustrated in Fig. 17.30. Unipolar limb lead electrodes provide the positive pole: lead aVR is recorded from the right arm, lead aVL is recorded from the left arm, and lead aVF is recorded from the left leg. In these leads, *a* stands for augmented; *V* stands for voltage; and *R*, *L*, and *F* indicate the location of the unipolar lead (*r*ight arm, *l*eft arm, and *f*oot [left]). A normal ECG from these leads is illustrated in Fig. 17.31.

Precordial unipolar chest leads are recorded from electrodes placed in six positions over the heart on the anterior chest (Fig. 17.32). Chest leads are designated as V_1, V_2, V_3, V_4, V_5, and V_6. The normal ECG from the chest leads is shown in Fig. 17.33. The chest leads provide a horizontal view of the heart, whereas the limb leads provide a view of the frontal plane.

Twelve-lead ECGs are usually recorded for a short period when the patient is resting (Fig. 17.34). Sequential ECGs are useful for determining changes over time. In some cases it is necessary to monitor the ECG recording for an extended period to capture rhythm problems that

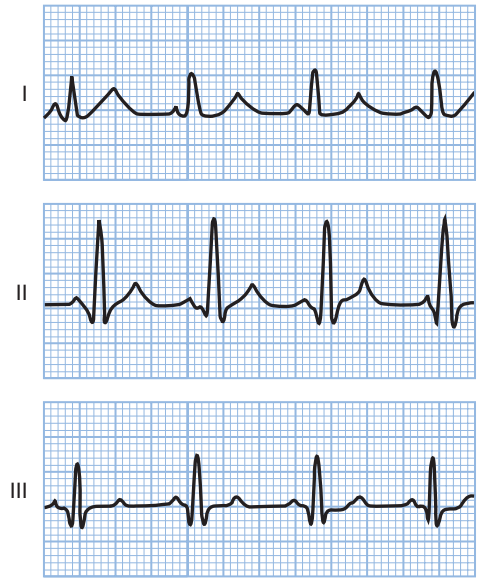


FIG 17.29 Normal electrocardiogram recorded from the three standard bipolar limb leads. The R wave is normally upright in leads I, II, and III.

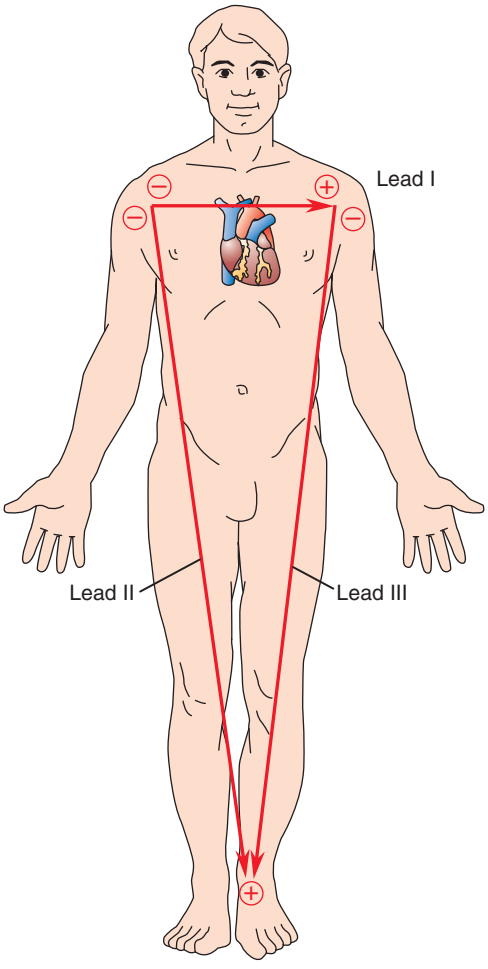


FIG 17.28 Positions of standard bipolar limb leads I, II, and III. The positive (+) lead is the recording lead.

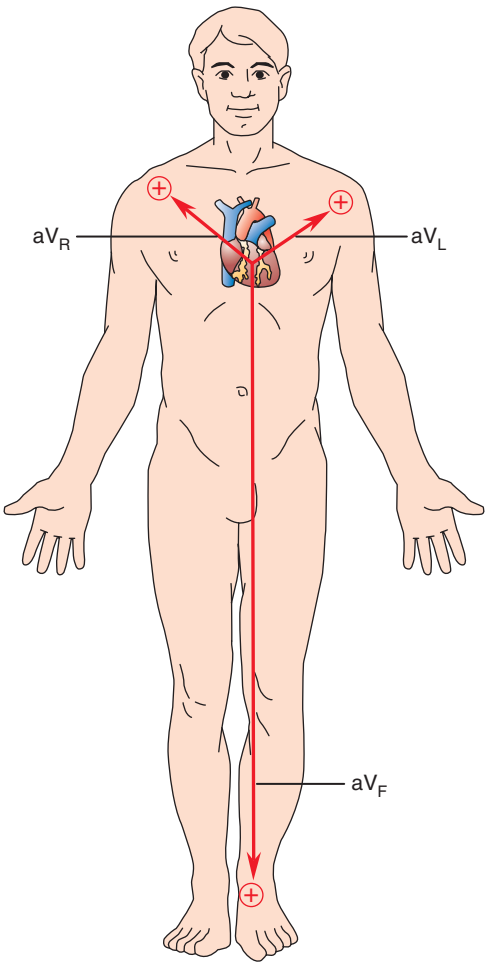


FIG 17.30 Unipolar augmented leads aVR, aVL, and aVF.

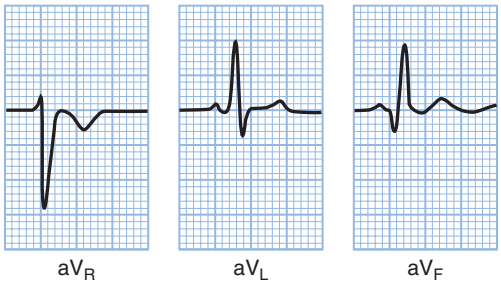


FIG 17.31 Normal electrocardiogram recorded from the three unipolar augmented leads. The aVR lead is characterized by a large S wave and an inverted T wave. The aVL and aVF leads have an upright R wave and T wave.

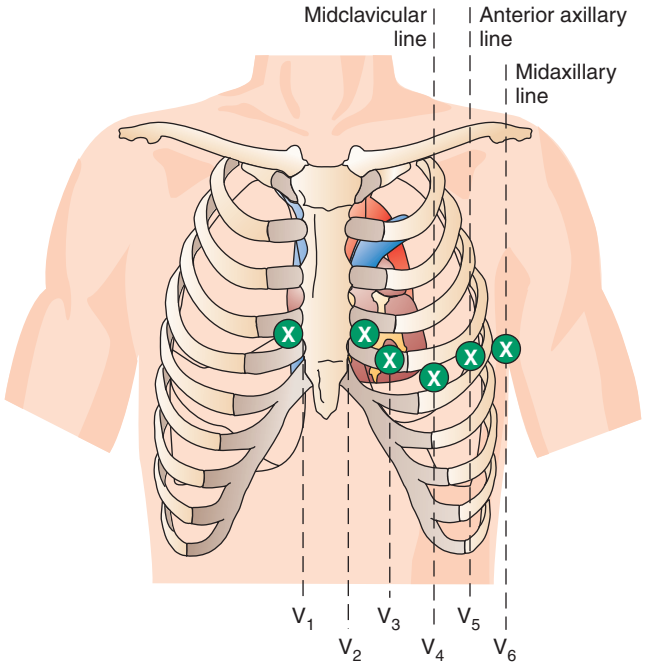


FIG 17.32 Unipolar chest (precordial) leads V1 through V6.

occur infrequently or with particular activities. This is accomplished by continuous ambulatory monitoring (e.g., Holter monitoring) over a 24- to 48-hour period. An ECG can also be recorded during exercise to monitor the effects of exercise stress on cardiovascular function. An exercise test (stress test) is usually performed while the subject progressively increases his or her effort on a treadmill or stationary bicycle. The exercise ECG is particularly useful for assessing the adequacy of coronary circulation when the myocardial workload is increased. Impaired myocardial oxygen delivery may be evident on the ECG as ST segment elevation or depression and abnormal T waves.

Magnetic Resonance Imaging and Computed Tomography

Magnetic resonance imaging (MRI) and computed tomography (CT) are useful for imaging cardiac structures. Myocardial thickening, pericardial sac disease, valvular structures, and congenital malformations may be visualized by MRI. Contrast-enhanced imaging identifies acute and chronic myocardial infarcts with high specificity and sensitivity. MRI and CT are used to detect coronary plaque burden and assess vulnerable plaque morphology in the arterial walls. Excessive plaque burden and unstable plaques are correlated with a greater degree of coronary atherosclerosis and may be used to predict coronary artery disease risk or progression.

Echocardiography

Echocardiography uses reflected sound waves (ultrasound) to provide an image of cardiac structure and motion within the chest. The cardiac echo is obtained by placing a blunt probe on the chest surface, which transmits and receives high-frequency sound waves. Sound waves traveling through chest and heart structures are reflected back to the receiving probe. The time between sound wave emission and detection of reflected waves is used to calculate distances between the probe and reflecting tissue. The sound waves are not heard or felt by the subject and have no known detrimental effects on tissues. The probe is moved across the chest to assess cardiac structures of interest, and recordings are videotaped for later viewing.

Echocardiograms are particularly useful for diagnosis of heart enlargement, valvular disorders, collections of fluid in the pericardial space, cardiac tumors, and abnormalities in left ventricular motion. Estimations of ejection fraction and assessments of ventricular systolic and diastolic function can be made noninvasively by echocardiogram. An echocardiogram is shown in Fig. 17.35.

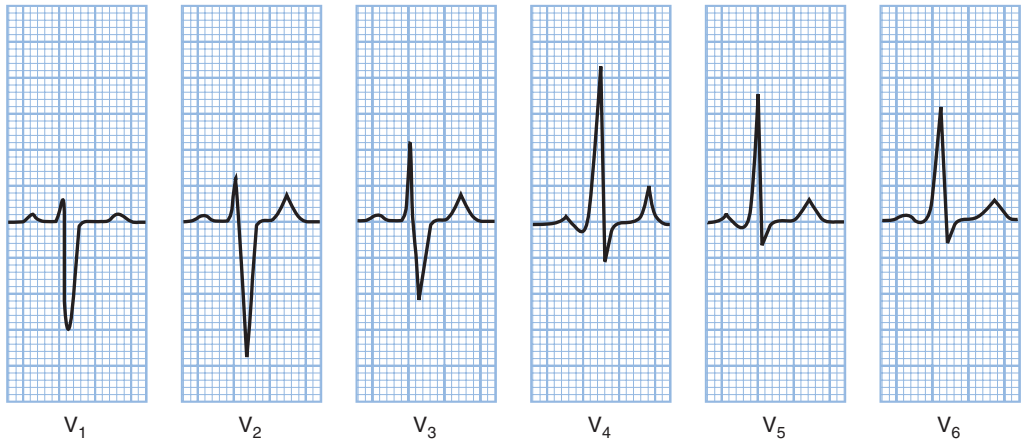


FIG 17.33 Normal electrocardiogram recorded from the six unipolar chest leads, V1 through V6. Note the progression across the precordial leads as the R waves become increasingly positive.

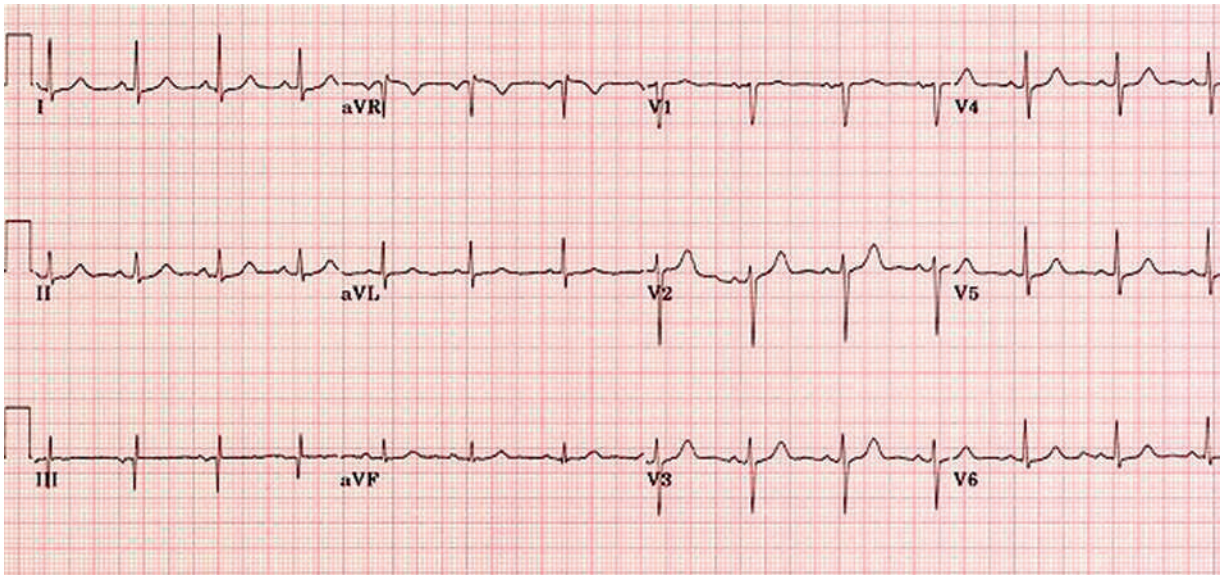


FIG 17.34 Example of a normal 12-lead ECG showing R wave progression in V leads.

Nuclear Cardiography

Radioactive substances injected into the bloodstream can be used to trace the patterns of blood flow in the heart. Radiation exposure is minimal because very small amounts of radioactive substances are used. Radioactive tracers can be linked to substances that accumulate in normal myocardial cells while the tracer is delivered by coronary blood flow. Areas with impaired perfusion will absorb less radioactivity and appear as "cold spots" on the scan. Scanning usually is done using single-photon emission computed tomography (SPECT), which images numerous slices through the heart, or by planar imaging, which gives an overall picture in one plane only.

Thallium-201 (^{201}Tl) and technetium-99 labeled compounds (^{99m}Tc sestamibi) are used to assess the adequacy of blood flow to cardiac tissues. After injection of the radioactive compound, the heart is scanned to visualize the amount of radioactivity absorbed by cardiac tissues. Healthy cardiac tissues that receive adequate blood supply actively accumulate these isotopes. Areas of inadequate blood flow or infarcted tissue do not accumulate isotopes and appear as cold spots on the scan. Resting and exercise scanning are done to assess for exercise-induced perfusion defects.

Gated pool scanning (radionuclide ventriculogram) is used primarily to assess left ventricular motion and ejection fraction. Before it is injected intravenously, radioactive technetium is attached to albumin or red blood cells, and therefore it remains in the bloodstream and is not absorbed by cells. Computer imaging is used to analyze blood flow through the chambers of the heart over many cardiac cycles. The dynamics of ventricular motion, such as hypercontractility or hypocontractility, may be visualized. Separate radionuclide ventriculogram evaluation has largely been replaced by SPECT scanning that allows simultaneous evaluation of perfusion and left ventricular function.

Positron emission tomography scans may also be used to evaluate cardiac perfusion and metabolism. Radiotracers can be incorporated into substances normally used in cellular metabolic processes, such as glucose. Metabolic activity in different areas of the heart can then be tracked over time under different conditions.

Cardiac Catheterization/Coronary Angiography

Cardiac catheterization/coronary angiography may be used to determine important structural and hemodynamic characteristics because it affords direct measurement of pressures within cardiac chambers; visualization of chamber size, shape, and movement; sampling for blood oxygen content in various heart regions; measurement of cardiac output and ejection fraction; and visualization and management of coronary artery obstructions.

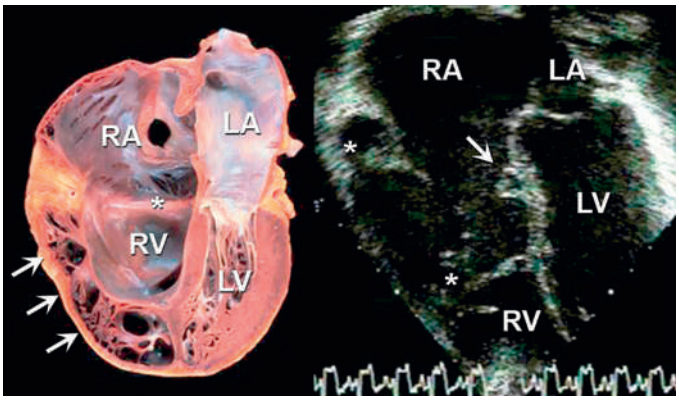


FIG 17.35 Pathology specimen *(left)* and echocardiography image *(right)* of heart with enlarged right atrium and right ventricle. *LA,* Left atrium; *LV,* left ventricle; *RA,* right atrium; *RV,* right ventricle. (From Connolly H, Oh J: Echocardiography. In Bonow R et al, editors: *Braunwald's heart disease: a textbook of cardiovascular medicine,* ed 9, vol 1, Philadelphia, 2012, Saunders, p 264.)

Cardiac catheterization angiography is associated with several serious risks, including bleeding, dysrhythmias, heart perforation, and coronary ischemia. The advent of noninvasive high-resolution MRI, echocardiograms, and SPECT scanning has replaced the need for cardiac catheterization in many instances; however, the value of information supplied is generally believed to outweigh the risks in certain cases. Catheterization is frequently used for interventions to rapidly improve coronary blood supply and to evaluate suspected or confirmed coronary artery disease, valvular dysfunction, congenital defects, left ventricular dysfunction, and coronary bypass graft patency.

Assessment of the left side of the heart, including the coronary arteries, is achieved by passing a catheter through a femoral or brachial artery into the aorta. The catheter is then manipulated into the left

ventricle or left atrium to assess chamber pressures, and a ventriculogram is obtained. Contrast dye injected into the ventricular chamber is monitored fluoroscopically to assess ventricular function. The catheter is usually pulled back into the aorta and then advanced into one or more of the coronary arteries. The patency of the coronary arteries can be visualized by injecting contrast dye into them and monitoring by fluoroscopy (Fig. 17.36). When contrast dye is in the coronary artery, a period of cardiac ischemia is produced during which the patient may experience angina, dysrhythmias, and coronary spasms. Coronary catheterization may also be done to insert a probe for obtaining intracoronary ultrasounds. Ultrasounds are useful for assessing plaque morphologic characteristics (Fig. 17.37).

Right-sided heart catheterization is done to evaluate right-sided heart structures. The catheter is introduced into a vein, usually femoral, and then threaded through the inferior vena cava and into the heart. Pressures and blood samples are obtained as the catheter is advanced into the right atrium, ventricle, and pulmonary artery. Right heart catheterization is useful in assessing tricuspid and pulmonary valve disorders, pulmonary hypertension, septal defects, and right ventricular function.

Coronary angiography is commonly followed by interventions to treat detected abnormalities. The coronary catheter can be used to direct thrombolytic agents to the site of coronary thrombosis and rapidly restore blood flow to ischemic areas. Laser therapy, coronary balloon angioplasty, and stent placement can also be performed during coronary angiography. These methods clear the coronary obstruction through thermal and mechanical means. The success of these approaches to management of coronary obstruction depends largely on how soon after an ischemic event they are performed (see Chapter 18).

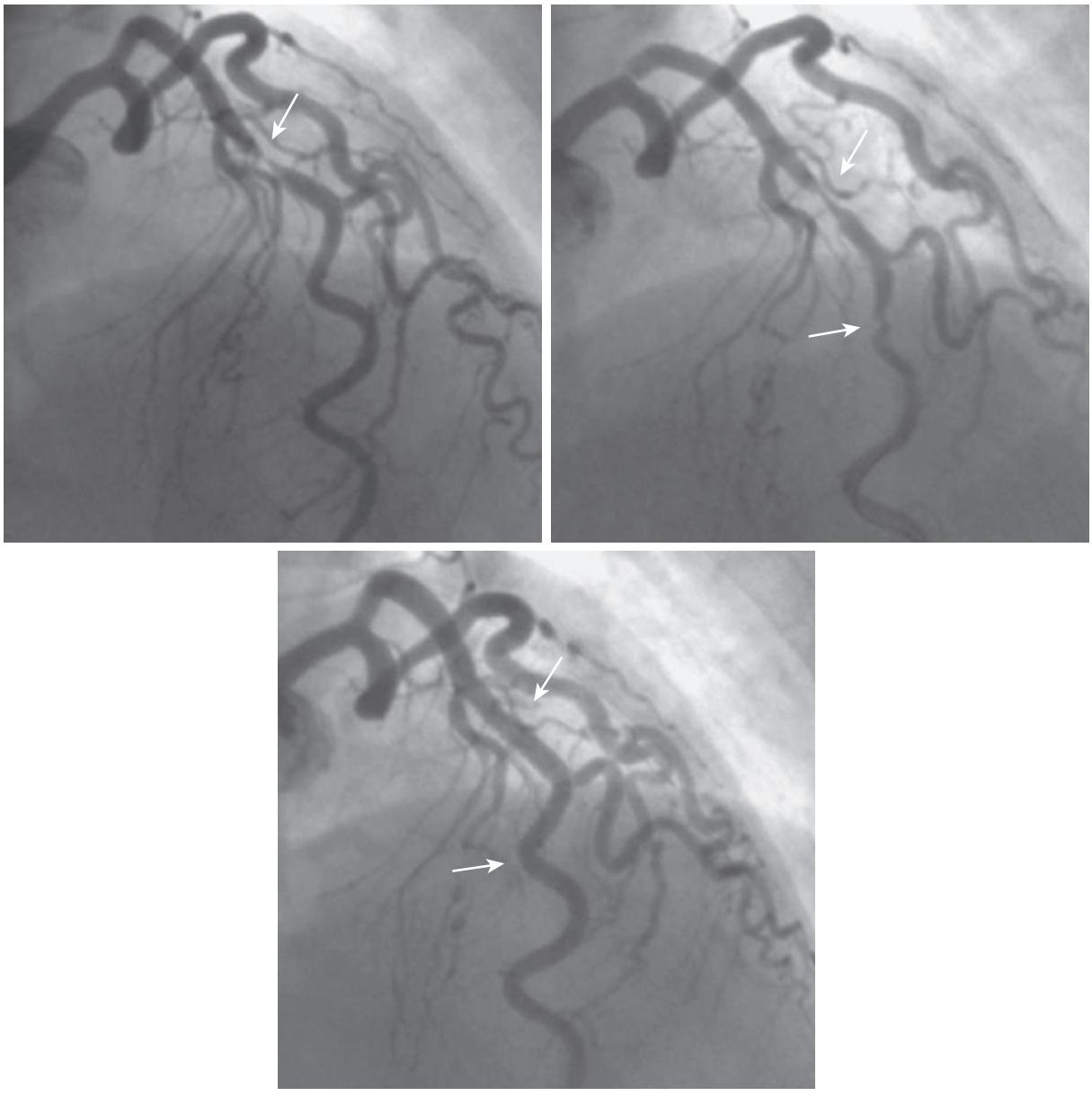


FIG 17.36 Coronary artery angiography. The arrows show an area of obstruction of the coronary artery. (From Popma J et al: Coronary arteriography and intracoronary imaging. In Mann D et al, editors: *Braunwald's heart disease: a textbook of cardiovascular medicine,* ed 10, Philadelphia, 2015, Saunders, p 414.)

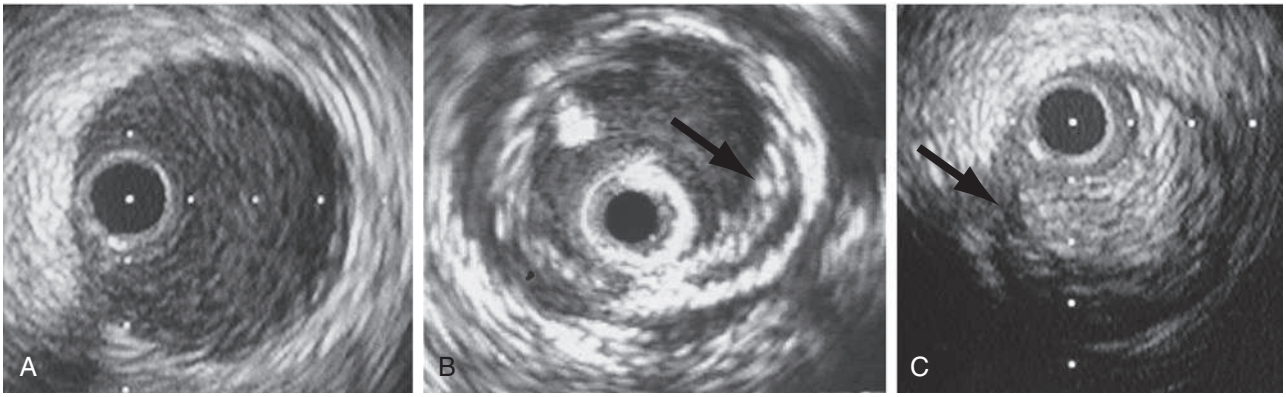


FIG 17.37 Intracoronary ultrasonographic examples of plaque morphology. **A,** A normal vessel wall. **B,** Fibrous cap on coronary plaque. **C,** A soft plaque with rupture of the fibrous cap. (From Braunwald E, Zipes D, Libby P, editors: *Braunwald's heart disease: a textbook of cardiovascular medicine,* ed 6, Philadelphia, 2001, Saunders, p 416.)

SUMMARY

The heart's primary function is to pump sufficient blood to deliver oxygen and nutrients to the body. The heart may be viewed as two separate pumps: a right-sided pump that perfuses the lungs and a left-sided pump that perfuses the systemic circulation. The left ventricle must generate higher pressures and therefore has a thicker myocardial mass and higher energy requirements. Because little ATP storage in cardiac cells is possible, the coronary arteries must deliver a steady supply of oxygen and nutrients. Cardiac contraction can be described by the sliding filament/cross-bridge theory and occurs only in the presence of ATP and free calcium ions. Factors that enhance intracellular calcium ion concentration will result in generation of a greater contractile force.

A coordinated cardiac contraction is possible because the heart's conduction system activates the chambers in a sequential manner. The sinoatrial node is the usual pacemaker because it has the highest intrinsic rate of diastolic depolarization. The diastolic depolarization rate is strongly influenced by the autonomic nervous system. The ECG shows the electrical activity of the heart and is a useful indicator of cardiac conduction abnormalities.

The ultimate indicator of cardiac function is the cardiac output, which is the product of heart rate and stroke volume. The autonomic nervous system is the main regulator of heart rate, whereas stroke volume is influenced by preload, afterload, and contractility. These factors are also the primary determinants of myocardial work and energy expenditure.

RESOURCES

Canty JM, Duncker DJ: Coronary blood flow and myocardial ischemia. In Mann D, Zipes D, Libby P, Bonow R, editors: *Braunwald's heart disease: a textbook of cardiovascular medicine*, ed 10, Philadelphia, 2015, Saunders, pp 1029–1056.

Clerico A, Giannoni A, Vittorini S, Passino C: Thirty years of the heart as an endocrine organ: physiological role and clinical utility of cardiac natriuretic hormones. *Am J Physiol Heart Circ Physiol* 301(1):H12–H20, 2011.

Fang JC, O'Gara PT: The history and physical examination: an evidence-based approach. In Mann D, Zipes D, Libby P, Bonow R, editors: *Braunwald's heart disease: a textbook of cardiovascular medicine*, ed 10, Philadelphia, 2015, Saunders, pp 95–113.

Favero G, Paganelli C, Buffoli B, et al: Endothelium and its alterations in cardiovascular diseases: life style intervention. *Biomed Res Int* 2015. http://dx.doi.org/10.1155/2014/801896.

Leri A, Rota M, Hosoda T, et al: Cardiac stem cell niches. *Stem Cell Research* 13(3ptB):631–646, 2014.

Libby P, Bonow R, editors: *Braunwald's heart disease: a textbook of cardiovascular medicine*, ed 10, Philadelphia, 2015, Saunders, pp 429–453.

Mirvis DM, Goldberger AL: Electrocardiography. In Mann D, Zipes D, Libby P, Bonow R, editors: *Braunwald's heart disease: a textbook of cardiovascular medicine*, ed 10, Philadelphia, 2015, Saunders, pp 114–152.

Opie LH, Bers DM: *Mechanisms of cardiac contraction and relaxation*. In Mann D, Zipes D, Libby P, Bonow R, editors: *Braunwald's heart disease: a textbook of cardiovascular medicine*, ed 10, Philadelphia, 2015, Saunders, pp 429–453.

Panteghini M, et al: Use of biochemical biomarkers in acute coronary syndromes. IFCC Scientific Division, Committee on Standardization of Markers of Cardiac Damage, International Federation of Clinical Chemistry. *Clin Chem Lab Med* 37:683–693, 1999.

Pettersen KH, Bugenhagen SM, Nauman J, et al: *Arterial stiffening provides sufficient explanation for primary hypertension.* arXiv:1305.0727 [q-bio.TO], 2013.

Popma J, Kinley S, Bhatt DL: Coronary arteriography and intracoronary imaging. In Mann D, Zipes D, Libby P, Bonow R, editors: *Braunwald's heart disease: a textbook of cardiovascular medicine*, ed 10, Philadelphia, 2015, Saunders, pp 392–423.

Starling EH: *The Linacre lecture on the law of the heart*, London, 1918, Longmans Green.

Strandring S: *Gray's anatomy*, ed 41, London, 2016, Churchill Livingstone.

18

Alterations in Cardiac Function

Jacquelyn L. Banasik

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- What is the role of injury, inflammation, and lipid oxidation in coronary plaque initiation and progression?
- What factors alter the balance between myocardial oxygen supply and demand?
- How do the clinical features of the coronary heart disease syndromes differ?
- How do valvular disorders alter cardiac pressure dynamics and workload?
- What are the similarities and differences among the cardiomyopathies and myocarditis?
- How do pericarditis and pericardial effusions differ in regard to cause and significance?
- What factors determine whether a congenital heart defect will produce cyanosis?

CHAPTER OUTLINE

Coronary Heart Disease, 383
Etiology of Coronary Heart Disease, 383
Risk Factors and Mechanisms of Coronary Atherosclerosis, 383
Pathophysiology of Ischemia, 385
Clinical Features and Management of Coronary Syndromes, 388
Angina Pectoris, 388
Acute Coronary Syndrome, 388
Sudden Cardiac Arrest, 393
Chronic Ischemic Cardiomyopathy, 393
Endocardial and Valvular Diseases, 394
Disorders of the Mitral Valve, 394
Mitral Stenosis, 395
Mitral Regurgitation, 396
Mitral Valve Prolapse, 396
Disorders of the Aortic Valve, 397
Aortic Stenosis, 397
Aortic Regurgitation, 397
Diseases of the Endocardium, 398
Rheumatic Heart Disease, 398
Infective Endocarditis, 398
Myocardial Diseases, 399
Myocarditis, 399
Cardiomyopathy, 400
Dilated Cardiomyopathy, 400
Hypertrophic Cardiomyopathy, 400
Restrictive Cardiomyopathy, 401
Pericardial Diseases, 402
Pericardial Effusion, 402
Cardiac Tamponade, 402
Pericarditis, 402
Acute Pericarditis, 402
Chronic Pericarditis, 403
Congenital Heart Diseases, 403
Embryologic Development, 403
Etiology and Incidence of Congenital Heart Disease, 404
Pathophysiology of Congenital Heart Disease, 405
Acyanotic Congenital Defects, 406
Atrial Septal Defect, 406
Ventricular Septal Defect, 407
Patent Ductus Arteriosus, 407
Coarctation of the Aorta, 408
Pulmonary Stenosis or Atresia, 408
Aortic Stenosis or Atresia, 408
Cyanotic Congenital Defects, 408
Tetralogy of Fallot, 408
Transposition of the Great Arteries, 408
Truncus Arteriosus, 409
Tricuspid Atresia, 409

The incidence of cardiovascular disease (CVD) increased rapidly in the United States during the last century, but the death rates from CVD declined. CVD accounts for approximately 31% of all deaths. Approximately 45% of these deaths are due to coronary heart disease (CHD), whereas stroke, high blood pressure, heart failure, and others claim the remainder. The decline in cardiac mortality has been achieved in the United States because of improvements in treatment and prevention. More than 16 million people living today have a history of CHD. Men and women are equally represented, although women tend to be older when their heart disease becomes apparent. The direct and indirect economic cost of cardiovascular diseases, including stroke, is well over $315 billion annually. CHD is the most important cardiovascular disorder in terms of numbers affected and economic impact.

CORONARY HEART DISEASE

CHD is also called *ischemic heart disease* and *coronary artery disease (CAD)* in some sources. These terms are related because CHD is characterized by insufficient delivery of oxygenated blood to the myocardium (ischemia) because of atherosclerotic coronary arteries (CAD). The American Heart Association compiles statistics under the heading of CHD, which includes the diagnoses of angina pectoris and myocardial infarction (MI). CHD causes about one in six deaths in the United States. Other sequelae of CHD include dysrhythmias, sudden cardiac arrest, and heart failure. When metabolic demand for oxygen exceeds supply, the myocardium becomes ischemic, which leads to a dysfunction in cardiac pumping and predisposes to abnormal heart rhythms. If the ischemic episode is severe or prolonged, irreversible damage to myocardial cells may result in MI.

Etiology of Coronary Heart Disease

Atherosclerosis of coronary arteries is the source of nearly all CHD. Atherosclerosis causes progressive narrowing of the arterial lumen and predisposes to a number of processes that can precipitate myocardial ischemia, including thrombus formation, coronary vasospasm, and endothelial cell dysfunction. Uncommon causes of cardiac ischemia include abnormalities of blood oxygen content (e.g., respiratory failure) and poor perfusion pressure through the coronary arteries (e.g., hypotension, hypovolemia). Occasionally, patients experience the signs and symptoms of cardiac ischemia but show no evidence of significant coronary artery atherosclerosis when evaluated by angiography. These patients are thought to have abnormalities of the microcirculation. Abnormal vascular regulation by endothelial cells in small vessels of the heart has been suggested as a probable mechanism. Endothelial cells secrete variable quantities of vascular relaxing and contracting factors and play a key role in controlling myocardial blood flow. Abnormalities of the microcirculation are more difficult to detect than coronary artery plaque, which is evident on coronary angiography (see Chapter 17).

Risk Factors and Mechanisms of Coronary Atherosclerosis

Knowledge about mechanisms of plaque formation in the coronary arteries has rapidly accumulated in recent decades. Epidemiologic studies reported in the 1960s suggested associations among certain traits and habits and the development of CHD. More recent studies have confirmed and expanded on these *risk factors,* which include several major risks (e.g., age, family history, abnormal lipid levels, cigarette smoking, hypertension, diabetes, and obesity) and numerous probable risks (Box 18.1). Although males and females succumb to heart disease in equal numbers, male gender is a risk factor for earlier development of heart disease—on average about 10 years earlier. The risk factors for CHD are the same as those for atherosclerosis in other arteries and are discussed in Chapter 15.

The observation that atherosclerotic plaque is composed primarily of lipid prompted the idea that abnormal lipid metabolism was a probable culprit in the development of CHD, and a great deal of attention has been focused on therapies to reduce levels of serum cholesterol in individuals with dyslipidemia. Lipids are transported through the bloodstream in combination with specific proteins (apoproteins). Certain lipid–protein molecules (lipoproteins) are associated with a greater risk of atherosclerosis. The five major kinds of lipoproteins are shown in Fig. 18.1. High levels of low-density lipoproteins (LDLs), which are high in cholesterol, have been associated with the highest risk. Very-low-density lipoproteins, which have large amounts of triglycerides, also appear to increase the risk. High-density lipoproteins, on the other hand, have been correlated with a decreased risk of atherosclerosis.

High-density lipoproteins are thought to transport cholesterol from the peripheral tissues back to the liver, thus removing atheromatous plaque. The role of low-density and, indirectly, very-low-density lipoproteins is to transport cholesterol to the peripheral tissues (Fig. 18.2). Cholesterol uptake by peripheral cells is mediated by LDL receptors on cell surfaces that bind and promote endocytosis of cholesterol. The liver normally binds and internalizes about 75% of the circulating LDL cholesterol.

BOX 18.1 Risk Factors for Coronary Heart Disease

Nonmodifiable Risks

Age: ≥45 years for men; ≥55 years for women
Gender: male (earlier onset)
Family history of premature coronary heart disease

- Myocardial infarction or sudden cardiac death in male first-degree relative at age less than 55 years or female first-degree relative at age less than 65 years
- Genetic markers identified

Lipid Risk Factors

Total cholesterol >200 mg/dL
LDL cholesterol >130 mg/dL
Triglycerides >150 mg/dL
HDL cholesterol <40 mg/dL

Nonlipid Risk Factors

Hypertension >140/90 mm Hg
Cigarette smoking
Thrombogenic state
Diabetes
Obesity
Physical inactivity
Poor diet (atherogenic)

Probable Risk Factors (Emerging)

Lipoprotein(a)
Small LDL particles (pattern B)
HDL subtypes
Apolipoprotein B
Homocysteine
Fibrinogen
Chronic inflammation
Impaired fasting glucose (100–125 mg/dL)

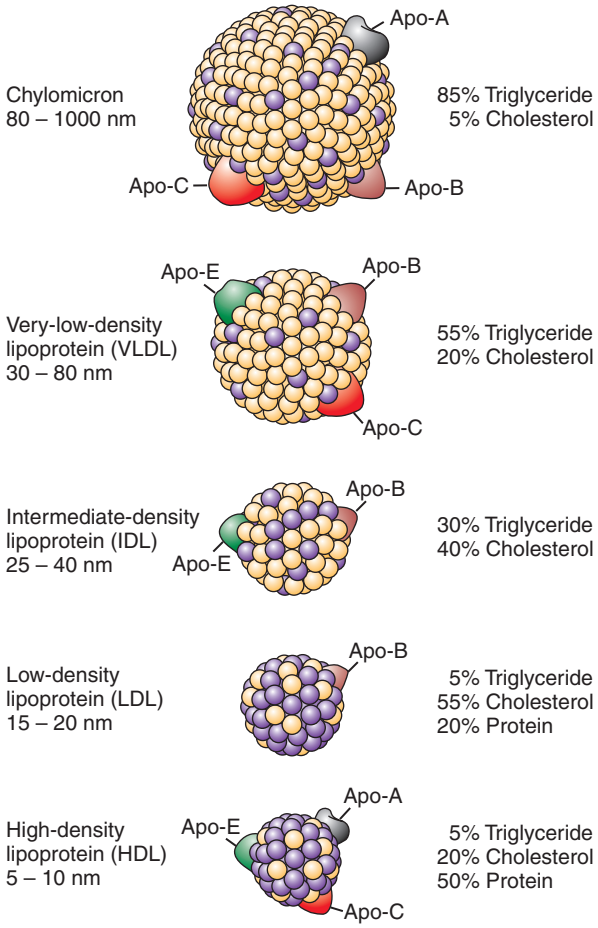


FIG 18.1 Serum lipoprotein fractions showing lipid composition and apoprotein components. Binding of lipoproteins to receptors is mediated through apoproteins.

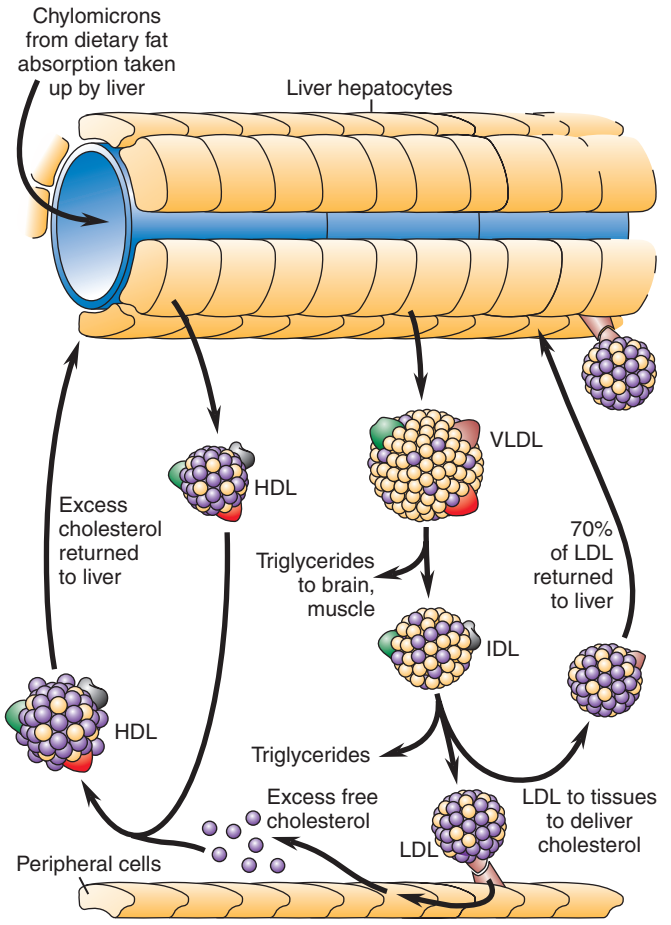


FIG 18.2 Schematic of lipoprotein metabolism in the body. Chylomicrons from dietary fat absorption are taken up by the liver and resynthesized into high-density lipoprotein (HDL) and very-low-density lipoprotein (VLDL). HDL circulates to the peripheral tissues and takes up excess cholesterol for transport back to the liver. Triglycerides are removed for tissue use from VLDL, which becomes intermediate-density lipoprotein (IDL). More triglyceride removal leads to the formation of low-density lipoprotein (LDL). LDL is absorbed by peripheral tissues to obtain cholesterol. About 70% of the circulating LDL returns to the liver.

Extreme cases of hyperlipidemia occur in individuals who have genetic derangements in lipid metabolism. These disorders run in families, and some are associated with the development of severe coronary atherosclerosis at a young age unless aggressively managed. The most common form of genetic hyperlipidemia (familial hypercholesterolemia) is associated with a defect in the LDL receptor on liver cells. Inability of the liver to efficiently remove cholesterol from the bloodstream results in hyperlipidemia. Genetic disorders of lipid metabolism are described in Table 18.1. Even when lipid metabolism is normal, a high-fat diet can overwhelm the liver's ability to clear LDL cholesterol from the circulation and results in hyperlipidemia. Dietary fat restriction may be beneficial in reducing cholesterol level in this case.

Atherosclerotic plaque formation is initiated by injury to the coronary artery endothelium. The specific cause of endothelial dysfunction in the early stage of atherosclerosis is uncertain; however, several potential mechanisms have been described. These include chronic hemodynamic wall stress, which may explain the typical localization of plaques at arterial branch points and the role of hypertension as a risk factor; toxins from cigarette smoke; circulating inflammatory cytokines; and hyperlipidemia. Once the injury occurs, the endothelium may become more permeable and recruit leukocytes (Fig. 18.3). LDLs leak through the endothelium and into the vessel wall (insudation) where they are oxidized by endothelial cells and macrophages. Oxidized lipids are damaging to the endothelial and smooth muscle cells and stimulate the recruitment of macrophages into the vessel wall, where they engulf the lipids. Lipid-filled macrophages are called *foam cells.* The macrophages and foam cells release inflammatory mediators and growth factors that attract more leukocytes and stimulate smooth muscle proliferation. Excess lipid and debris begin to accumulate within the vessel wall and to coalesce into a pool called the *lipid core* (Fig. 18.4). Atherosclerotic plaques with large lipid cores are fragile and prone to rupture. Rupture of a plaque exposes subendothelial proteins and initiates platelet aggregation and thrombus formation. Thrombi may be asymptomatic if they are small and do not occlude the artery. Components of the thrombus may be incorporated into the plaque, causing it to enlarge. Older plaques have significant collagen and fibrin, which form a cap and tend to make the plaque more stable. Numerous therapies aimed at stabilizing vulnerable plaques and preventing thrombus formation have been studied in clinical trials.

In addition to lifestyle modification to improve risks, statins are recommended as the drug class of choice for patients with significant risks for cardiovascular events. The groups that benefit most from statin therapy are shown in Box 18.2. The use of nonstatin agents, such as fibrates and niacin, is no longer recommended. In addition, the use of serum lipid measurements as medication dosing targets is no longer recommended because lipid lowering is only one of the mechanisms

TABLE 18.1 Genetic Lipoprotein Disorders

Disorder	Gene
LDL Particles	
Familial hypercholesterolemia	*LDL-R*
Familial defective ApoB-100	*ApoB*
Gain of function PCSK9 mutations	*PCSK9*
Autosomal recessive hypercholesterolemia	*ARH*
Abetalipoproteinemia	*MTP*
Hypobetalipoproteinemia	*ApoB*
Familial sitosterolemia	*ABCG5/ABCG8*
Familial LP(a) hyperlipoproteinemia	*Apo(a)*
Remnant Lipoproteins	
Dysbetalipoproteinemia type III	*ApoE*
Hepatic lipase deficiency	*HL*
Triglyceride-Rich Lipoproteins	
Lipoprotein lipase deficiency	*LPL*
ApoC-II deficiency	*ApoC-II*
Apo-AV	*ApoA-V*
Familial hypertriglyceridemia	Polygenic
Familial combined hyperlipidemia	Polygenic
Hdl Particles	
Apo-A1 deficiency	*Apo-A1*
Tangier disease, familial HDL deficiency	*ABCA1*
Familial LCAT deficiency syndromes	*LCAT*
CETP deficiency	*CETP*
Niemann-Pick disease types A and B	*SMPD1*
Niemann-Pick disease type C	*NPC1*

Adapted from Genest J, Libby P: Lipoprotein disorders and cardiovascular disease. In Mann D, Zipes D, Libby P, Bonow R: *Braunwald's heart disease: a textbook of cardiovascular medicine,* ed 10, Philadelphia, 2015, Elsevier, p 988.
CETP, Cholesteryl ester transfer protein; *HDL,* high-density lipoprotein; *LCAT,* lecithin cholesterol acyltransferase; *LDL,* low-density lipoprotein; *Lp(a),* lipoprotein a.

BOX 18.2 Major Statin Benefit Groups

For secondary prevention for those with clinical atherosclerotic cardiovascular disease
For those with LDL cholesterol >190 mg/dL
For those with diabetes aged 40 to 75 years with LDL cholesterol between 70 and 189 mg/dL
For those without diabetes aged 40 to 75 years with LDL cholesterol between 70 and 189 mg/dL who have an estimated risk of ≥7.5% using the pooled cohort algorithm

of statin benefit. Instead the doses used are based on risk of cardiovascular events, with higher doses prescribed for those with higher risk (Table 18.2). Nearly 50% of MIs occur in patients who do not have hyperlipidemia. Statin medications do carry significant risk of liver and muscle damage and should be used in patients only when benefits outweigh these risks. Atherosclerotic lesions generally increase in size over many years and progressively occlude the lumen of vessels. A significant reduction in blood flow can result when plaque occupies 75% or more of the arterial lumen. Clinically significant atherosclerotic plaque may be located anywhere within the three major coronary arteries or major secondary branches. Often all three coronary arteries are simultaneously

TABLE 18.2 Statin Therapy for High-, Moderate-, and Low-Intensity Therapy

High-Intensity Therapy	Moderate-Intensity Therapy	Low-Intensity Therapy
Atorvastatin 40–80 mg daily	Atorvastatin 20–40 mg daily	Simvastatin 10 mg daily
Rosuvastatin 20–40 mg daily	Rosuvastatin 5–10 mg daily	Pravastatin 10–20 mg daily
	Simvastatin 20–40 mg daily	Lovastatin 10 mg daily
	Pravastatin 40–80 mg daily	Fluvastatin 20–40 mg daily
	Lovastatin 40 mg daily	Pitavastatin 1 mg daily
	Fluvastatin 40 mg twice daily	
	Or Fluvastatin XL 80 mg daily	
	Pitavastatin 2–4 mg daily	

From Ridker PM, Libby P, Buring JE: Risk markers and the primary prevention of cardiovascular disease. In Mann D, Zipes D, Libby P, Bonow R: *Braunwald's heart disease: a textbook of cardiovascular medicine,* ed 10, Philadelphia, 2015, Elsevier, p 931.

affected, although some individuals have only one or two diseased vessels. Surprisingly, the extent and severity of atherosclerotic lesions are not good predictors of the severity of ischemia.

Atherosclerotic coronary lesions have been characterized and attempts have been made to correlate the anatomic descriptions with plaque development and behavior. Typically atherosclerotic lesions begin as fatty streaks and progress to small regions of medial wall thickening with scattered macrophages at a young age. As the plaques acquire more free lipid within the arterial wall, they are more vulnerable to rupture, thrombus formation, and progressive plaque growth. These advanced lesions carry a significant risk of producing disruptions in coronary blood flow. Critical narrowing of the coronary lumen over time or sudden rupture of a plaque followed by thrombus formation causes the clinical syndromes of CHD, including angina, infarction, ischemic cardiomyopathy, and sudden cardiac arrest.

Stable plaques usually are asymptomatic or may be associated with exercise-induced angina pain (stable angina pectoris). However, plaques are vulnerable to rupture or erosion, which can initiate thrombus formation and acute coronary occlusion. A variety of factors have been identified as markers of increased plaque vulnerability. These factors include (1) active inflammation within the plaque, (2) a large lipid core with a thin cap, (3) endothelial denudation (erosion) with superficial platelet adherence, (4) fissured or ruptured cap, and (5) severe stenosis predisposing to high shear stress. Acute coronary syndrome (ACS), or unstable angina or MI, as well as sudden cardiac arrest, is nearly always associated with acute disruption of a vulnerable plaque. Because the types of plaques that are most vulnerable often do not significantly obstruct the lumen before they rupture, ACS frequently occurs in individuals whose disease had been asymptomatic. Patients with a high risk for or known presence of vulnerable plaques benefit from therapies such as statin agents (to stabilize plaques) and antiplatelet agents such as aspirin (to prevent thrombosis).

Pathophysiology of Ischemia

Ischemia of cardiac cells occurs when the oxygen supply is insufficient to meet metabolic demands. Myocardial cells are unable to store much

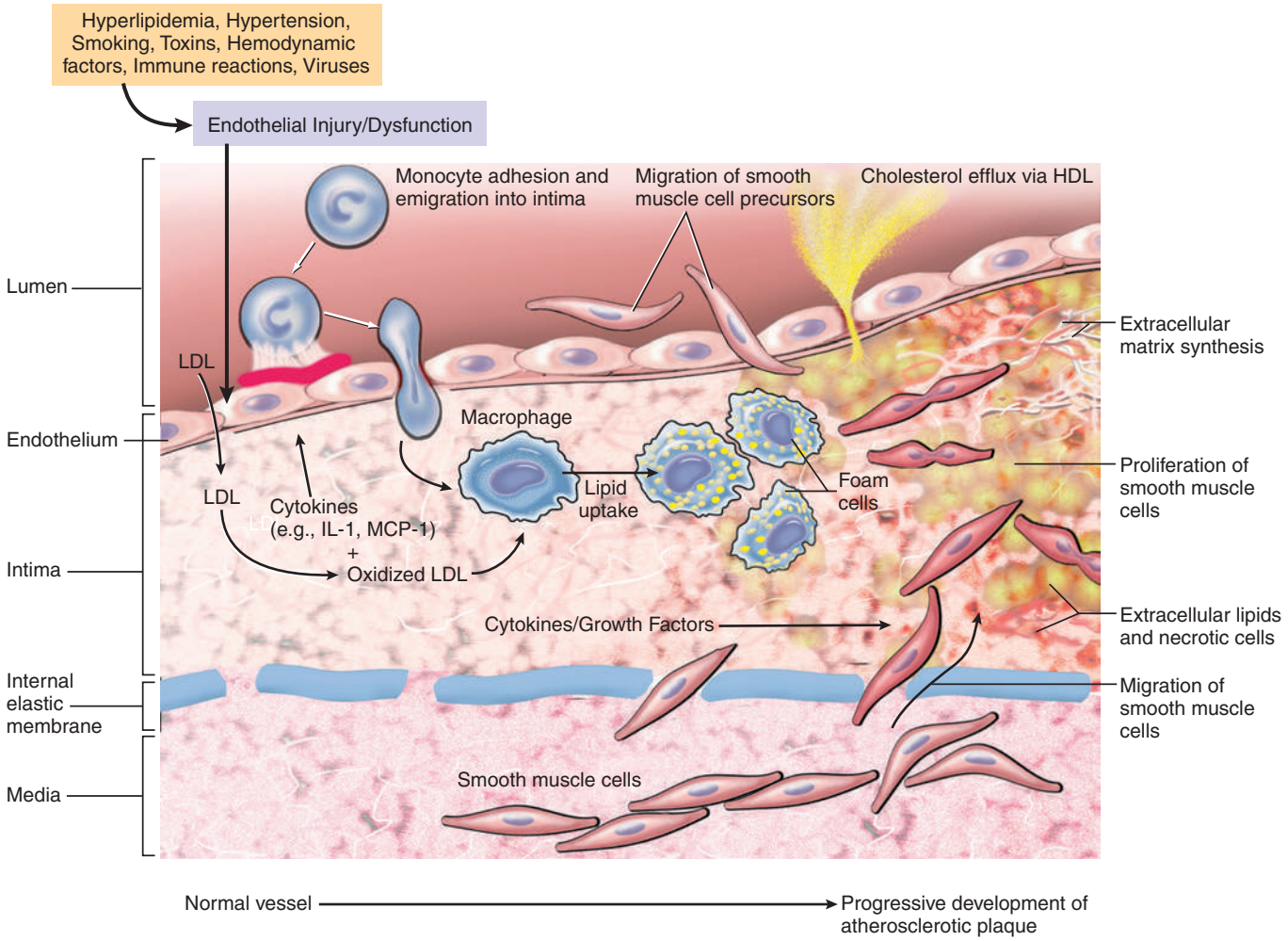


FIG 18.3 Sequence of events in the arteriolar wall associated with development of atherosclerosis. Note that smooth muscle cells migrate from the intima through the internal elastic membrane and into the intimal layer where they proliferate in response to growth factors. Macrophages in the intima release signals that alter the endothelial cell layer and induce expression of cell adhesion molecules that recruit monocytes into the tissue. (From Hovland A et al: The complement system and toll-like receptors as integrated players in the pathophysiology of atherosclerosis. *Atherosclerosis* 2015;241(2):480–494.)

energy in the form of adenosine triphosphate (ATP) and must therefore continuously receive a supply of oxygen for aerobic synthesis of ATP. ATP is essential for powering myocardial contraction as well as for cell maintenance. Because the heart is unable to slow its activity when ATP supplies dwindle, a steady flow of oxygen is essential.

Factors that decrease myocardial oxygen supply or increase myocardial oxygen demand can upset the balance and result in cellular ischemia. Thus the critical factors in meeting cellular demands for oxygen are (1) the rate of coronary perfusion and (2) the myocardial workload. Coronary perfusion can be impaired in several ways, including (1) large, stable atherosclerotic plaque, (2) acute platelet aggregation and thrombosis, (3) vasospasm, (4) failure of autoregulation by the microcirculation, and (5) poor perfusion pressure.

Myocardial workload depends on heart rate, preload, afterload, and contractility (see Chapter 17). An increase in any of these variables increases myocardial oxygen requirements and may precipitate ischemia. However, even conditions resulting in very high myocardial oxygen consumption will seldom lead to ischemia unless some underlying impairment in coronary perfusion is present.

One or more of the aforementioned mechanisms are operative in producing clinically significant myocardial ischemia resulting in the acute or chronic coronary syndromes. Advanced fibrous plaque is thought to produce intermittent ischemia when 65% to 75% or more of the arterial lumen is occluded. Because fibrous plaque progresses slowly over many years, the heart can develop alternative pathways for myocardial blood flow. This collateral circulation can preserve blood flow despite almost total occlusion of the coronary artery. Thus stable fibrous plaque may produce no symptoms of ischemia unless the demand of the heart for oxygen is suddenly elevated, as occurs in exercise or stress. When the onset of ischemia is predictable with certain activities and subsides with rest, the patient is said to have a chronic coronary syndrome, called *classic* or *stable angina pectoris*.

ACS occurs when sudden obstruction of coronary blood flow results in acute myocardial ischemia. Acute obstruction is usually associated with the formation of a clot in the coronary artery at the site of a vulnerable plaque. Rupture of the plaque exposes a rough area composed of collagen and other molecules that are thrombogenic. A high fibrinogen level, as occurs in smokers, and enhanced platelet adhesiveness, as occurs in hyperlipidemia, may enhance the risk of thrombus formation. Clot formation begins with adherence of platelets to the ruptured plaque. The platelets that initially attach release chemicals that attract more platelets, which aggregate and form a plug. The coagulation cascade

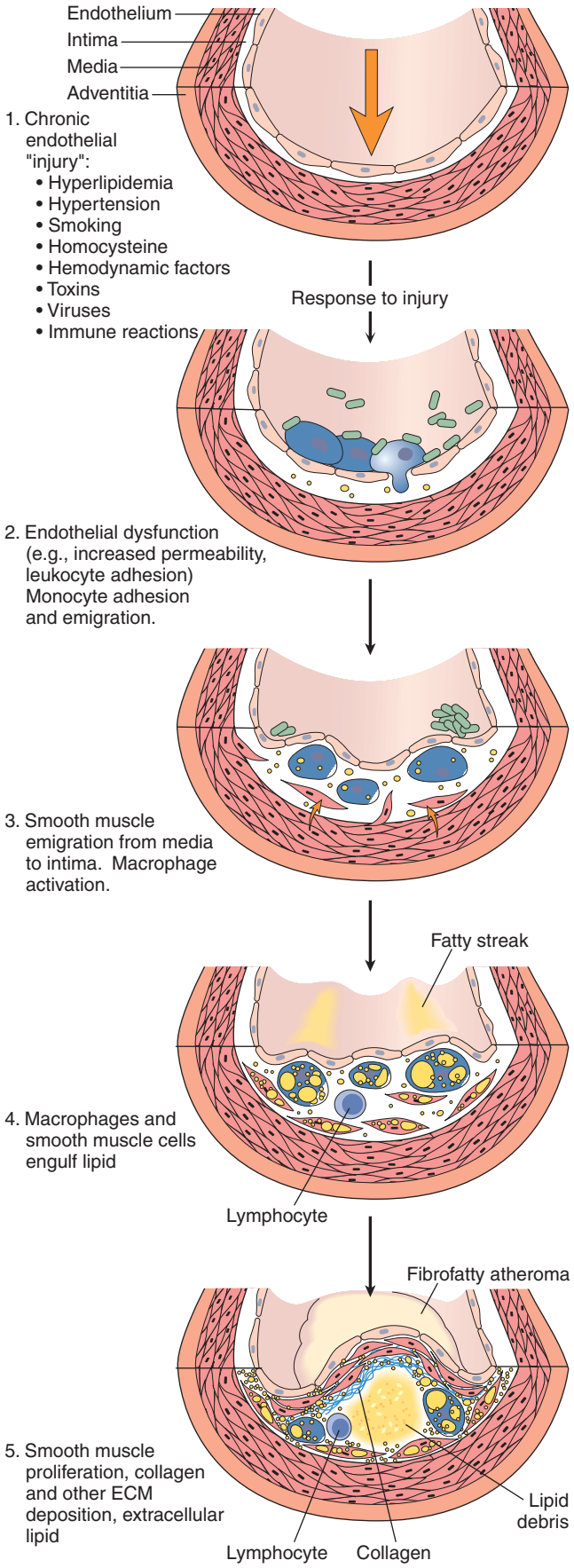


FIG 18.4 Pathogenesis of atherosclerosis. *1,* Chronic endothelial injury leads to 2. *2,* Endothelial dysfunction, permeability, and inflammation. *3,* Activated monocytes infiltrate the arterial wall and smooth muscle proliferates. *4,* Macrophages engulf lipid to become foam cells. *5,* A lipid core forms in the arterial wall and a fibrous cap evolves. (From Garden JO, Bradbury AW, Forsythe JLR, Parks RW: *Principles and practice of surgery*, ed 6, London, 2012, Churchill Livingstone.)

may also be initiated and result in the formation of a platelet–fibrin clot that may occlude the vessel or break loose and travel farther along the vessel. Chemicals released by activated platelets include several vasoactive products (e.g., serotonin, thromboxane) that may contribute to spasm of the coronary vessel, further reducing blood flow.

Thrombosis occurs suddenly and may partially or completely obstruct the artery and cause acute ischemia. The ACS may present as unstable angina, MI, or sudden cardiac arrest. Appreciation of the role thrombus formation plays in coronary obstruction has resulted in the prophylactic use of antithrombotics, such as aspirin. Research indicates that the long-term use of small doses of aspirin reduces mortality from ischemic heart disease.

Vasospasm usually occurs in areas of atherosclerotic plaque, but is also proposed as a mechanism of ischemia in patients who have anginal signs and symptoms but no significant amount of fibrous plaque in the coronary arteries. *Variant,* or *Prinzmetal,* angina is the term applied to vasospasm-initiated anginal symptoms. The etiology of spasm in vessels with no significant atherosclerotic plaques is unclear, but usually responds promptly to vasodilating agents. Intense vasospasm can occur in response to certain drugs, such as cocaine.

As previously mentioned, endothelial cells are important regulators of vascular tone. They secrete variable amounts of constricting and relaxing factors to control tissue perfusion. This autoregulation of blood flow allows the microvasculature to dilate when the need for oxygen in a particular area is increased. Failure of endothelial cells to appropriately regulate flow is a potential mechanism of myocardial ischemia. Endothelial cells can be damaged by circulating toxins from cigarette smoke, immune cells, and infectious agents. Inflammatory disorders that may alter endothelial cell function include lupus erythematosus, Kawasaki syndrome, and polyarteritis nodosa. The importance of inflammatory processes in the pathogenesis of CHD has been recognized, resulting in efforts to find markers (e.g., serum high-sensitivity C-reactive protein) and methods to reduce inflammation in those at risk.

Even if the coronary arteries and microcirculation are functioning properly, coronary perfusion may still be inadequate if perfusion pressure is low. Recall from Chapter 17 that coronary blood flow occurs primarily during diastole and depends on the driving pressure in the aorta. A fall in aortic blood pressure can significantly reduce coronary perfusion, particularly in vessels with high resistance to flow. Conditions such as shock, hemorrhage, and anesthesia may be associated with a decline in blood pressure that decreases driving pressure and coronary perfusion and results in myocardial ischemia. However, the most common cause of cardiac ischemia is atherosclerotic coronary arteries.

KEY POINTS

- Cardiac ischemia occurs when the heart's demand for oxygenated blood exceeds its supply. In most cases, ischemia is a result of impaired blood flow through the coronary arteries.
- Coronary heart disease (CHD) is associated with coronary atherosclerosis. Risk factors for CHD are the same as those for atherosclerosis of other

arteries and include advancing age, male gender, family history, hyperlipidemia, diabetes, smoking, hypertension, and obesity. Endothelial injury and inflammation and lipid accumulation in the intima are thought to be the primary initiators of coronary atherosclerosis.

- Early atherosclerotic lesions are asymptomatic precursor lesions in which lipids begin to accumulate in the arterial wall. Advanced lesions may cause symptoms because of progressive arterial occlusion or acute plaque disruption and thrombus formation.
- Vulnerable plaques may rupture or become eroded, which stimulates clot formation on the plaque. Plaques with a large lipid core, thin cap, or high shear stress are vulnerable.
- Chronic occlusion of a coronary vessel is associated with the clinical syndrome of stable angina. Acute occlusion is associated with plaque disruption and thrombus formation and results in ACS (unstable angina or MI).
- Myocardial ischemia may uncommonly be caused by coronary vasospasm, hypoxemia, or low perfusion pressure from volume depletion or shock.

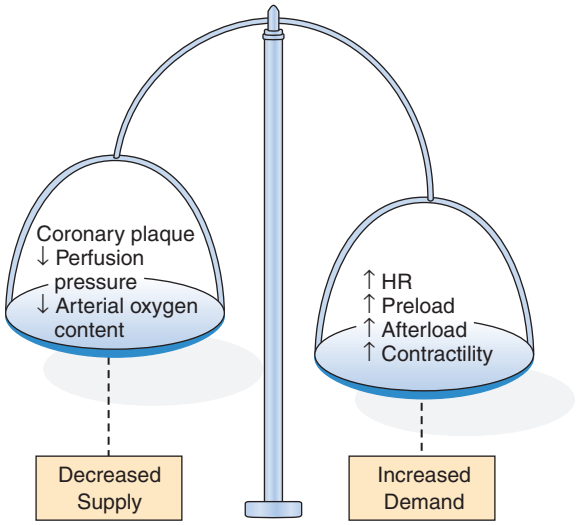


FIG 18.5 Factors that decrease coronary blood supply or increase myocardial oxygen demand can upset the balance and lead to ischemia and anginal pain.

Clinical Features and Management of Coronary Syndromes

Five syndromes can be differentiated according to the severity and onset of cardiac symptoms. Stable *angina pectoris* and *ischemic cardiomyopathy* are chronic syndromes that usually progress slowly and are a consequence of chronic obstruction from stable atherosclerotic plaques. ACS has an abrupt onset and life-threatening consequences and is associated with acute changes in plaque morphology and thrombosis. ACS includes *unstable angina* and *MI*. Unstable angina and MI are combined together because they are difficult to differentiate in the acute stage when therapeutic decisions must be made. Any of the coronary heart syndromes may precipitate *sudden cardiac arrest* and associated dysrhythmias.

Angina Pectoris

Angina pectoris literally means *chest pain* and is associated with intermittent myocardial ischemia. Bouts of chest pain and associated symptoms are generally recurrent and may be precipitated by conditions that increase myocardial oxygen demand, such as exercise; stress; sympathetic nervous system activation; and increased preload, afterload, heart rate, or muscle mass. Ischemic pain receptors from the myocardium travel to the central nervous system with the eighth cervical nerve and the first through fourth thoracic dorsal root ganglia. Sensory neurons from the jaw, neck, and arm also travel in these nerve trunks, so heart pain may be perceived as emanating from these body parts. This phenomenon is called *referred pain*. Anginal pain may be described as burning, crushing, squeezing, or choking. Pain is sometimes represented by expressions such as "an elephant is sitting on my chest" or by the patient placing a tight fist on the chest. Anginal pain may be mistakenly attributed to indigestion or dental pain. In some cases, patients have atypical symptoms of myocardial ischemia, such as back pain, fatigue, or weakness, rather than the classic symptom of chest pain.

Anginal ischemia, even when temporary, may result in inefficient cardiac pumping with resultant pulmonary congestion and shortness of breath. Three patterns of angina pectoris have been described: (1) stable or typical angina, (2) Prinzmetal or variant angina, and (3) unstable or crescendo angina. All these patterns are associated with underlying coronary vessel disease and may be exhibited in a particular individual at different times and under different conditions. Unstable angina may progress to acute ischemia and is discussed in the "Acute Coronary Syndrome" section along with MI.

Stable angina. *Stable angina* is the most common form and is therefore called *classic* or *typical angina*. Stable angina is characterized by stenotic atherosclerotic coronary vessels that reduce coronary blood flow to a critical level. The stenosed arteries dilate poorly in response to increased myocardial oxygen requirements. Under conditions of increased myocardial workload, such as during physical exertion or emotional strain, coronary perfusion is inadequate and ischemia results (Fig. 18.5). The onset of anginal pain is generally predictable and elicited by similar stimuli each time. Stable angina is generally relieved by rest and nitroglycerin, a drug that causes coronary and peripheral vasodilation, reduces preload, and, consequently, reduces myocardial workload.

Prinzmetal variant angina. *Prinzmetal variant angina* is characterized by unpredictable attacks of anginal pain. Although most individuals with Prinzmetal angina have some coronary atherosclerosis, the onset of ischemic symptoms is unrelated to physical or emotional exertion, heart rate, or other obvious causes of increased myocardial oxygen demand. Vasospasm has been identified as the probable mechanism leading to variant angina, although the cause of the vasospasm is unknown. Proposed mechanisms include atherosclerosis-induced hypercontractility, abnormal secretion of vasospastic chemicals by local mast cells, and abnormal calcium flux across vascular smooth muscle. Variant angina responds well to treatment with calcium channel–blocking drugs, which inhibit vascular smooth muscle contraction.

Patients with angina are at risk for developing ACS and need aggressive treatment for risk factor reduction and therapies to reduce the risk of plaque rupture, thrombosis, and dysrhythmia.

Acute Coronary Syndrome

Unstable angina and MI are difficult to distinguish on the basis of clinical manifestations and are lumped together as ACS. Both are characterized by chest pain that may be more severe and lasts longer than the patient's typical angina and may occur in individuals whose disease was previously asymptomatic. In both cases, plaque rupture with subsequent acute thrombus development is thought to occur. In unstable angina, the occlusion is partial or the clot is dissolved before the death of myocardial tissue. In MI, the occlusion is complete and the thrombus persists long enough for development of irreversible damage to myocardial cells, resulting in necrosis. In the past, differentiation of unstable angina and MI was based on laboratory evaluation of serum biomarker levels (e.g., MB band of creatine kinase [CK-MB], troponins I and T). If cardiac biomarkers were elevated, which is

indicative of necrosis, a diagnosis of MI was made; if not, a diagnosis of unstable angina was appropriate. In a time when monitoring and management of complications were the mainstay of treatment, this approach worked well. With the advent of reperfusion therapy, which is effective only if administered early in the course of infarction, the distinction between unstable angina and MI has become less clinically relevant. Because unstable angina and MI present a similar clinical picture in the acute phase, they have been combined in treatment protocols for ACS (Fig. 18.6). Patients with chest pain and evidence of acute ischemia on the electrocardiogram (ECG) (ST-segment elevation) are candidates for acute reperfusion therapy with fibrinolytic therapy. Patients presenting with symptoms of unstable angina and no ST elevation on the ECG do not benefit from thrombolytics and may be candidates for percutaneous coronary interventions (PCI). PCI is accomplished by inserting a catheter into the coronary arteries followed by interventions to open the occluded artery such as removal of the thrombus, disruption of plaque, or placement of stents in the artery to keep it open. Differentiation between unstable angina, NSTEMI (non–ST-elevation MI), and ST-elevation MI (STEMI) is made after obtaining cardiac necrosis markers; those patients with elevations in markers are diagnosed with NSTEMI or STEMI, and those without elevations are diagnosed with unstable angina (see Fig. 18.6).

Etiology and pathogenesis. MI results when prolonged or total disruption of blood flow to the myocardium causes cellular death by necrosis or apoptosis. Acute MI is an important form of CHD resulting in more than 125,000 deaths annually in the United States. An MI may occur at any age, but the frequency rises with advancing age. Females younger than 45 years have a sixfold lesser risk of MI than men of the same age. After menopause, the rate of MI in women approaches that of their male counterparts and becomes essentially equal by age 80.

As previously described, the initiating event in most MIs is believed to be development of a thrombus on top of an ulcerated or cracked atherosclerotic plaque. The initiating event is a sudden change in structure of the plaque. Platelets passing by the surface of the ruptured plaque adhere to it, initiate formation of a platelet plug, and activate the clotting cascade. The resultant thrombus grows until it occludes the vessel and triggers the MI because of prolonged ischemia.

The cellular consequences of an acute interruption in blood flow to the myocardium do not occur instantaneously or uniformly. Acute occlusion causes a range of cellular events, depending on the availability and adequacy of collateral blood flow, the relative workload, and the length of time that flow is interrupted. A typical myocardial infarct has several zones composed of cells in various stages of ischemia and death.

Experiments in animal models indicate that complete occlusion of a coronary vessel results in a predictable pattern of cellular dysfunction and death. Depletion of ATP in acutely ischemic cells begins immediately, followed within 1 to 2 minutes by an impaired ability to contract. Within 10 minutes, cellular concentrations of ATP fall to half of normal, and irreversible cell injury occurs after 30 to 40 minutes of complete occlusion (Table 18.3). Ischemic necrosis begins in the subendocardial zone and spreads across the ventricular wall toward epicardial surfaces. Epicardial areas are spared for longer periods because they have the greatest collateral network of arterial vessels. The ultimate size of the infarcted tissue depends on the extent, duration, and severity of ischemia. Areas of necrosis may be intermixed with or surrounded by zones of reversibly injured cells that are marginally perfused by collaterals. Injured cells die both from necrosis and from apoptosis. Restoring perfusion to potentially salvageable cells is an important focus of treatment. (See Chapter 4 for a discussion of necrosis and apoptosis.)

Nearly all infarcts are located in the left ventricular walls. Isolated right ventricular infarction occurs in only 1% to 3% of MIs. Occlusion of the left anterior descending artery causes 40% to 50% of acute MIs, the right coronary artery contributes another 30% to 40%, and the left circumflex contributes 15% to 20%. The locations of the resulting infarcts are shown in Table 18.4. It is common for individuals with CHD to suffer from more than one MI during their lifetime.

The area of necrosis resulting from MI undergoes a series of morphologic changes as the infarct ages. These morphologic changes generally cannot be detected on gross examination until 6 to 12 hours after infarct. After 18 to 24 hours, the area of infarction becomes paler than the surrounding tissues. Thereafter the area of infarction becomes obvious because it turns yellowish and soft with a rim of red vascular connective tissue (Fig. 18.7). At 1 to 2 weeks, the necrotic tissue is progressively degraded and cleared from the site. Infarcted myocardium is particularly weakened and susceptible to rupture at this time. By 6 weeks, the necrotic tissue has been replaced by tough fibrous scar tissue.

Diagnosis of MI. The diagnosis of MI is based on three primary indicators: signs and symptoms, electrocardiographic changes, and elevations in the levels of specific marker proteins in the blood. Other diagnostic examinations such as cardiac catheterization, echocardiography, and

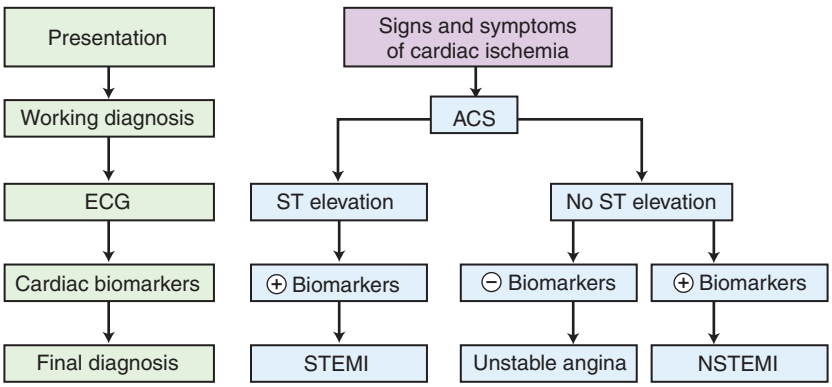


FIG 18.6 The etiologies of acute coronary syndrome (ACS) are difficult to differentiate by presenting symptoms because all involve some degree of myocardial ischemia. The electrocardiogram (ECG) is used to distinguish those patients with ST elevation from those with no ST elevation. Serum levels of cardiac biomarkers are then used to make a final diagnosis. Most patients with ACS characterized by ST elevation are diagnosed with MI (STEMI), and a proportion of patients with no ST elevation also will have elevated serum markers and are diagnosed with MI (NSTEMI). Patients who do not exhibit serum enzyme elevations are usually diagnosed with unstable angina.

TABLE 18.3 Evolution of Morphologic Changes in Myocardial Infarction

Time	Gross Features	Light Microscopic	Electron Microscopic
Reversible Injury			
0–½ hr	None	None	Relaxation of myofibrils; glycogen loss; mitochondrial swelling
Irreversible Injury			
½–4 hr	None	Usually none; variable waviness of fibers at border	Sarcolemmal disruption; mitochondrial amorphous densities
4–12 hr	Dark mottling (occasional)	Beginning coagulation necrosis; edema; hemorrhage	
12–24 hr	Dark mottling	Ongoing coagulation necrosis; pyknosis of nuclei; myocyte hypereosinophilia; marginal contraction band necrosis; beginning neutrophilic infiltrate	
1–3 days	Mottling with yellow-tan infarct center	Coagulation necrosis, with loss of nuclei and striations; interstitial infiltrate of neutrophils	
3–7 days	Hyperemic border; central yellow-tan softening	Beginning disintegration of dead myofibers, with dying neutrophils; early phagocytosis of dead cells by macrophages at infarct border	
7–10 days	Maximally yellow-tan and soft, with depressed red-tan margins	Well-developed phagocytosis of dead cells; early formation of fibrovascular granulation tissue at margins	
10–14 days	Red-gray depressed infarct borders	Well-established granulation tissue with new blood vessels and collagen deposition	
2–8 weeks	Gray-white scar, progressive from border toward core of infarct	Increased collagen deposition, with decreased cellularity	
>2 mo	Scarring complete	Dense collagenous scar	

From Kumar V et al: *Robbins and Cotran pathologic basis of disease,* ed 10, Philadelphia, 2015, Saunders, p 544.

TABLE 18.4 Location of Myocardial Infarction According to Coronary Artery Affected

Arterial Obstruction	Location of Infarct
Left anterior descending (40%–50% of infarcts)	Anterior wall of LV near apex Anterior portion of interventricular septum Apex
Right coronary (30%–40% of infarcts)	Posterior wall of LV Posterior interventricular septum
Left circumflex (15%–20% of infarcts)	Lateral wall of LV

Data from Kumar V et al: *Robbins and Cotran pathologic basis of disease,* ed 9, Philadelphia, 2015, Saunders, p 544.
LV, Left ventricle.

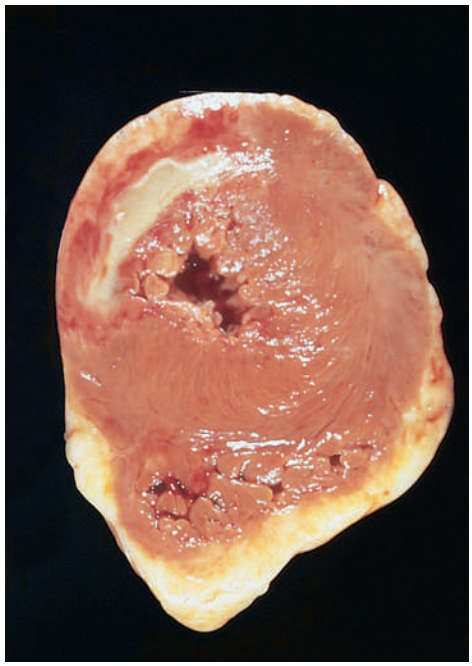


FIG 18.7 This photograph of a 5- to 7-day-old posterolateral infarction clearly shows a large, pale yellow lesion surrounded by a dark red zone of inflammation. (From Kumar V et al, editors: *Robbins basic pathology,* ed 7, Philadelphia, 2003, Saunders, p 369.)

radionuclide scintigraphy may also be performed to provide additional information (see Chapter 17).

Severe, crushing, excruciating chest pain that may radiate to the arm, shoulder, jaw, or back is the harbinger of MI. Pain is commonly accompanied by nausea, vomiting, diaphoresis (sweating), and shortness of breath. In contrast to anginal pain, infarction pain generally lasts more than 15 minutes and is not relieved by rest or nitroglycerin. In some instances, however, the MI is entirely asymptomatic and may elude detection. Asymptomatic MI has been called *silent MI* and may be detected only serendipitously at a later date. Pain may be difficult to assess in individuals with atypical presentations or a tendency to ignore or deny their symptoms. Thus although pain is an important indicator of acute ischemia, other clinical information is often needed to correctly distinguish between angina, infarction, and noncardiac sources of pain. Women, the elderly, and patients with diabetic neuropathies more commonly complain of atypical symptoms, including fatigue, nausea, back pain, and abdominal discomfort. Atypical complaints in patients with CHD risk factors should prompt a high suspicion of MI.

Electrocardiographic changes. Myocardial ischemia and infarction often result in characteristic changes on ECG waveforms. Injury and

ischemia are indicated on the ECG by ST-segment changes. ST-segment elevation is thought to represent acute cellular injury and ischemia. The presence of ST-segment elevation on the ECG indicates that the ischemic injury is ongoing and that immediate efforts to improve perfusion or reduce oxygen demand may be effective in preserving myocardial muscle mass. Patients with ST depression or T wave changes may benefit from timely PCI.

Infarcted muscle that is necrotic and no longer electrically active usually is indicated by the appearance of abnormally deep (>0.1 mV) or wide (>0.03 seconds) Q waves and inverted T waves (Fig. 18.8). These changes are specific for MI and, when present, are diagnostic. Q waves are usually persistent findings, whereas ST-segment and T wave changes may resolve over time. Q waves can take time to develop and may not be present in the acute phase of an MI. Q waves may never develop if the area of infarction is small.

Dysrhythmias and the characteristic ST-segment changes that accompany MI are attributed to injured and ischemic cells that have not yet become necrotic. Reversibly injured cells have limited ATP supplies to power membrane pumps and are predisposed to leakage of ions across their cell membranes. Abnormal ion flux may result in continuous current flow even when the heart is at rest.

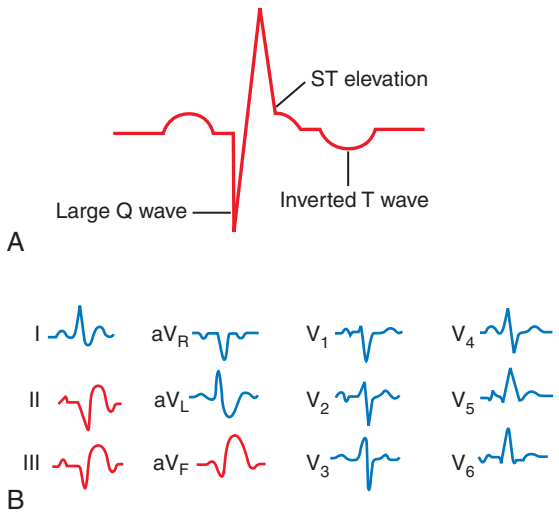


FIG 18.8 A, Typical ECG infarction pattern showing abnormally large Q wave, ST elevation, and inverted T wave. **B,** Typical ECG in acute inferior myocardial infarction. Note the Q waves and ST-segment elevation in leads II, III, and aVF.

The 12-lead ECG is used to localize the injured region of the left ventricle. Various leads of the 12-lead ECG "view" different regions of the heart. Abnormalities such as Q waves and ST-segment elevation in a particular lead or leads indicate that the damage is localized to the part of the left ventricle "seen" by that lead. MIs may thus be described as anterior, lateral, posterior, septal, inferior, or a combination of these sites. (The 12-lead ECG is described in Chapter 17.)

Serum biomarkers. The appearance of certain proteins in the blood after myocardial cell death is a sensitive and reliable indicator of MI. Myocardial cell death leads to elevated serum levels of myoglobin, troponin, lactate dehydrogenase, and creatine kinase. Cardiac cells contain particular forms of these proteins called *isoforms.* An increase in the concentration of these proteins suggests leakage from fatally damaged cells that have lost plasma membrane integrity. Cardiac biomarkers have a slightly different amino acid sequence than other cell types. In particular, myocytes contain the isoforms CK-MB, troponin I, and troponin T. An elevated level of serum CK-MB is a highly specific indicator of MI and is considered to be diagnostic. However, the level of CK-MB remains elevated for only 48 to 72 hours after MI. Two proteins that comprise part of the cardiac cell contractile apparatus, troponins I and T, have become the biomarkers of choice for detecting MI. Cardiac troponin levels become elevated in serum at about the same time as the CK-MB level, but they remain elevated for a longer period. Cardiac troponins I and T are highly sensitive and specific for cardiac cell death, but less helpful in detecting new infarction (reinfarction) because levels remain elevated for a prolonged period (Fig. 18.9). Cardiac myoglobin levels are elevated in serum very quickly after MI and may be helpful in early detection; however, cardiac myoglobin is less specific than the other biomarkers. All serum biomarkers are useful diagnostically only during the acute period of MI. Patients with ACS who do not develop elevations of these serum biomarkers are diagnosed with unstable angina, whereas those who do have elevations are classified as STEMI if they had ST elevation during the period of ACS, and NSTEMI if they did not have ST elevation during the acute phase.

Clinical course. In addition to chest pain, electrocardiographic abnormalities, and serum protein marker elevations, a person experiencing an MI may exhibit signs of cardiac inflammation, including fever,

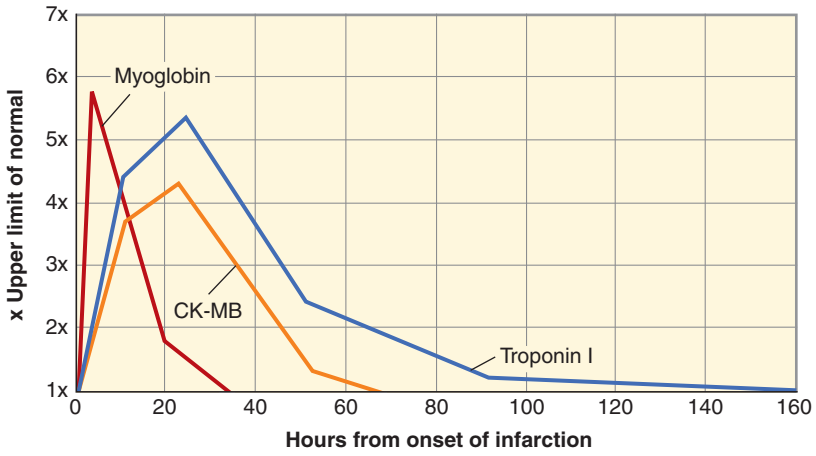


FIG 18.9 Time course of serum marker protein elevations after acute myocardial infarction. The MB band of creatine kinase (CK-MB) and troponin I are the most specific of the protein markers. Myoglobin is an early marker but is not very specific.

leukocytosis, and an elevated sedimentation rate. Symptoms of circulatory inadequacy, including fatigue, restlessness, anxiety, and weakness, may be present.

The events associated with MI are summarized in Fig. 18.10. Note that totally ischemic cells, which die and become electrically silent, are the source of the clinical findings of Q waves and release the indicative serum marker proteins (CK-MB, troponins). Partially ischemic cells are potentially salvageable but are unable to maintain normal ion flux across the cell membrane. Abnormal ion flux is responsible for the ST-segment changes in acute ischemia and predisposes to a variety of cardiac dysrhythmias, including ventricular ectopy and conduction blocks (see Chapter 19).

Neither partially ischemic nor totally ischemic cells are able to contract effectively, and poor stroke volume leads to a drop in cardiac output. Decreased stroke volume triggers a number of compensatory actions designed to improve cardiac output. In particular, activation of the sympathetic nervous system increases the heart rate, contractility, blood pressure, and fluid retention by the kidney. Unfortunately, these compensatory efforts impose a greater workload on the heart and may contribute to further ischemic damage. Compensatory mechanisms are shown in Fig. 18.11 and include sympathetic nervous system activation, enhanced preload, and hypertrophy of cardiac myocytes.

Prognosis and treatment. An overall prognosis for acute MI is difficult to determine because many variables affect the outcome, including the extent and location of the infarct, previous cardiovascular health, age, and the presence of other disease processes. Of particular importance is how quickly treatment is sought. Most deaths from MI occur before the victim reaches the hospital. The in-hospital mortality rate of MI is about 5%. In some cases, an MI is not accompanied by any complications (uncomplicated MI) and the patient recovers rapidly. However, the majority of MIs are followed by one or more complications. Potential complications include cardiac dysrhythmias, heart failure, cardiogenic shock, ventricular rupture, pericarditis, and thromboembolism.

Treatment for MI is directed at decreasing myocardial oxygen demand and increasing myocardial oxygen supply while monitoring and managing complications as they arise. Measures to reduce myocardial workload frequently include preload and afterload reduction, heart rate control, pain relief, and activity restriction. Sympathetic antagonists, nitrates, and morphine sulfate are the mainstays of drug therapy. Measures to increase oxygen delivery to ischemic areas include oxygen administration, antiplatelet therapy with aspirin and other antiplatelet agents, thrombolytic (fibrinolytic) drugs, anticoagulants, angioplasty, stent placement, and coronary artery bypass grafting (CABG) (Fig. 18.12). Therapies aimed at opening the blocked coronary artery are called *reperfusion therapies*.

Early detection and management of dysrhythmias and conduction disorders are an important part of the immediate care of a patient with MI. Many dysrhythmias are life threatening and, at the very least, lead to decreased cardiac output or increased myocardial workload. Continuous electrocardiographic monitoring is generally the standard of care

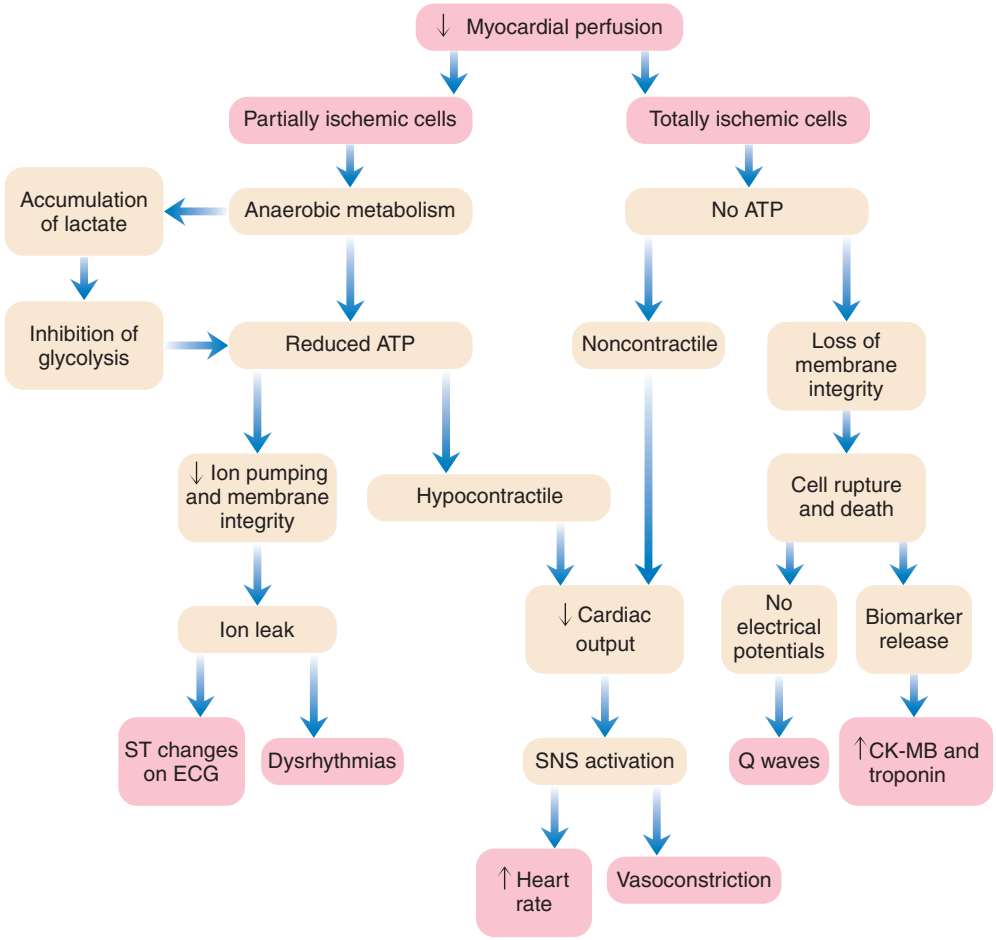


FIG 18.10 Summary of events after myocardial infarction. *ATP,* Adenosine triphosphate; *CK-MB,* MB band of creatine kinase; *ECG,* electrocardiography; *SNS,* sympathetic nervous system.

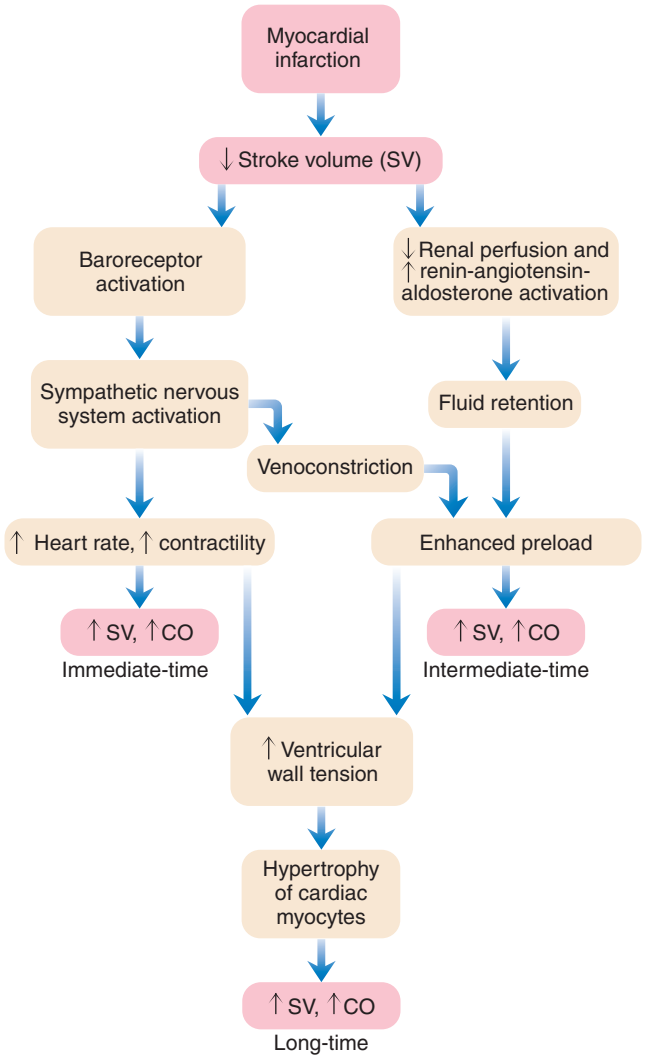


FIG 18.11 Compensatory responses to increase stroke volume (SV) and maintain cardiac output (CO) after myocardial infarction.

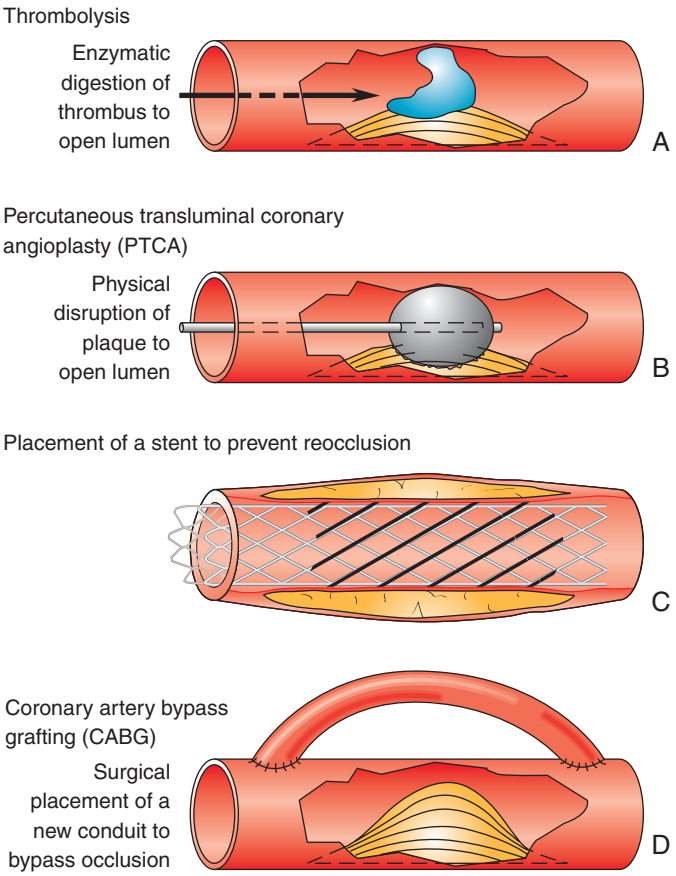


FIG 18.12 Management of blocked coronary arteries includes **(A)** thrombolysis with drugs such as streptokinase and tissue plasminogen activator; **(B)** plaque disruption with percutaneous transluminal coronary angioplasty, followed by anticoagulation or stent placement; **(C)** placement of a stent to prevent reocclusion; and **(D)** coronary artery bypass grafting—surgical placement of a new conduit to bypass the occluded area of the artery.

because of the high incidence of electrical disturbances after MI. Common dysrhythmias and conduction disorders are described in Chapter 19.

Sudden Cardiac Arrest

Also called *sudden cardiac death (SCD)*, this is usually defined as unexpected death from cardiac causes within 1 hour of the onset of symptoms. Successful resuscitation efforts by those trained in cardiopulmonary resuscitation (CPR) and use of the automated external defibrillators found in many public places have resulted in the increased survival of persons who experience witnessed SCD. Persons who survive SCD are at high risk for recurrence. CHD is the source of the vast majority of cases of sudden cardiac arrest. Sudden cardiac arrest may be a complication of hereditary or acquired structural or electrical abnormalities, such as long QT syndrome. An estimated 350,000 to 400,000 individuals experience SCD each year in the United States. Most cases are associated with coronary atherosclerosis and may be the initial manifestation of the disease. Acute MI occurs in a subset of cases of sudden cardiac arrest. A lethal dysrhythmia, such as ventricular fibrillation, is usually the primary cause (see Chapter 19). Ischemia from multivessel atherosclerosis, diffuse myocardial atrophy, scarring and fibrosis of old MI tissue, and electrolyte imbalances are factors that may predispose the heart to the electrical abnormalities that lead to sudden cardiac arrest.

Chronic Ischemic Cardiomyopathy

Chronic ischemic cardiomyopathy refers to a disorder in which heart failure develops insidiously as a consequence of progressive ischemic myocardial damage. In most cases, individuals affected have a history of angina or MI, often many years before the onset of heart failure. Heart failure appears to be a consequence of slow, progressive apoptotic death of myocytes from chronic ischemia. The disease is usually found in elderly individuals. Atrophic and dead cells are scattered throughout the myocardium rather than being localized, as occurs with MI. The prognosis for patients with chronic ischemic cardiomyopathy is quite poor, with death from heart failure the common outcome. Heart failure is further discussed in Chapter 19.

KEY POINTS

- The clinical syndromes of CHD include stable angina pectoris, ACS (unstable angina, NSTEMI, STEMI), chronic ischemic cardiomyopathy, and sudden cardiac arrest. These conditions are associated with coronary atherosclerosis.
- Stable angina is characterized by intermittent bouts of chest pain triggered by exertion and generally relieved by rest. No permanent myocardial damage occurs.

- Prolonged or severe ischemia results in MI that is characterized by severe, unrelieved chest pain, nausea and vomiting, diaphoresis, shortness of breath, and inflammation (e.g., fever, increased white blood cell [WBC] count, increased sedimentation rate).
- Serum protein biomarker elevations and electrocardiographic changes are diagnostic of MI. The most specific and sensitive serum biomarkers are increased levels of CK-MB and troponins I and T. ECG changes include ST-segment elevation (or depression), large Q waves, and inverted T waves.
- A drop in cardiac output as a result of MI triggers a number of compensatory responses, including sympathetic activation. The sympathetic nervous system increases the heart rate, contractility, and blood pressure, all of which increase myocardial workload.
- Treatment of acute ischemia usually includes efforts to decrease myocardial oxygen demand (e.g., sympathetic antagonists, rest, heart rate control, pain relief, afterload reduction) and increase oxygen delivery (e.g., thrombolysis, PCIs, CABG).

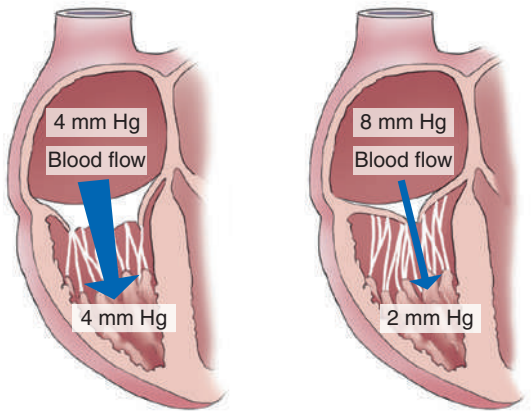


FIG 18.13 Development of a pressure gradient across a stenosed valve. The chamber behind the narrowed valve opening must work harder to force blood through the valve.

ENDOCARDIAL AND VALVULAR DISEASES

Endocardial and valvular structures may be damaged by inflammation and scarring, calcification, or congenital malformations. These processes interfere with the normal valvular property of unimpeded, unidirectional flow. Although congenital malformations may affect any valve, acquired valvular disorders generally involve the mitral or aortic valves. Abnormalities in valvular function cause altered hemodynamics in the heart and generally result in increased myocardial workload. Ultimately, heart failure may result from significant valvular dysfunction.

Normally, heart valves open completely, so blood flows through with little or no pressure difference across the valve. Failure of a valve to open completely is termed *stenosis*. Significant hemodynamic consequences generally begin to occur when the valve opening is reduced to half its normal diameter. The severity of stenosis can be estimated by the degree of pressure gradient across the valve (Fig. 18.13). Stenosis results in extra pressure work for the heart because blood must be forced through the high resistance of a narrow valve opening. Stenosis generally progresses slowly over years to decades, which allows time for affected heart chambers to compensate through myocardial cell hypertrophy.

Regurgitation (insufficiency) refers to the inability of a valve to close completely, thereby allowing blood to flow backward across the valve when no flow should be occurring. Regurgitation may develop suddenly from valvular infection or rupture of a supporting papillary muscle. Sudden regurgitation is poorly tolerated inasmuch as little compensation is possible. Regurgitation results in extra volume work for the heart because more blood must be pumped to maintain adequate forward flow.

Diseased valves may exhibit elements of both stenosis and regurgitation, although one problem usually predominates. Postinflammatory scarring from rheumatic heart disease and valvular calcification with aging are the primary causes of stenosis (Fig. 18.14). A wide variety of diseases of the endocardium may lead to valvular regurgitation, including rheumatic heart disease and infective endocarditis, which are discussed later in this chapter. Damaged valves are susceptible to infection, and antibiotic prophylaxis may be indicated for dental, surgical, and diagnostic procedures in some patients. The major causes of acquired mitral and aortic valvular diseases are listed in Table 18.5. Valvular disorders are often associated with abnormal turbulence of blood flow that produces heart sounds called *murmurs*. Careful assessment of the location and character of a murmur can help identify the underlying valvular abnormality. Defining characteristics of common valve disorders are described in Table 18.6.

TABLE 18.5 Major Etiologies of Acquired Heart Valve Disease

Mitral Valve Disease	Aortic Valve Disease
Mitral Stenosis	**Aortic Stenosis**
Postinflammatory scarring (rheumatic heart disease)	Postinflammatory scarring (rheumatic heart disease)
	Senile calcific aortic stenosis
	Calcification of a congenitally deformed valve
Mitral Regurgitation	**Aortic Regurgitation**
Abnormalities of leaflets and commissures:	
Postinflammatory scarring	Postinflammatory scarring (rheumatic heart disease)
Infective endocarditis	
Mitral valve prolapse	Degenerative aortic dilation
Drugs (e.g., Fen-Phen)	Syphilitic aortitis
Abnormalities of the tensor apparatus:	Ankylosing spondylitis
Rupture of papillary muscle	Rheumatoid arthritis
Papillary muscle dysfunction (fibrosis)	Marfan syndrome
Rupture of chordae tendineae	
Abnormalities of left ventricular cavity and/or annulus:	
LV enlargement (myocarditis, dilated cardiomyopathy)	
Calcification of mitral ring	

Modified from Schoen FJ: Surgical pathology of removed natural and prosthetic heart valves, *Hum Pathol* 1987;18(6):558–567; and from Kumar V et al: *Robbins and Cotran pathologic basis of disease,* ed 9, Philadelphia, 2015, Saunders, p 554.
LV, Left ventricle.

Disorders of the Mitral Valve

Three important disorders of the mitral valve are stenosis, regurgitation, and prolapse.

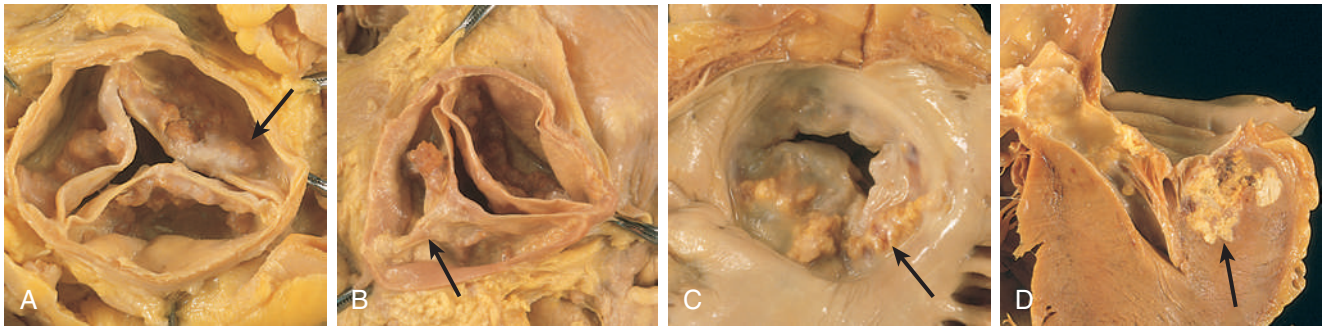


FIG 18.14 Valvular degeneration and calcification. **A,** Calcific aortic stenosis of previously normal (three-cusp) valve. **B,** Calcific aortic stenosis of congenital bicuspid valve. **C,** Mitral valve calcification. **D,** Cut section of valve from C. (From Kumar V et al, editors: *Robbins and Cotran pathologic basis of disease*, ed 9, Philadelphia, 2015, Saunders, p 555.)

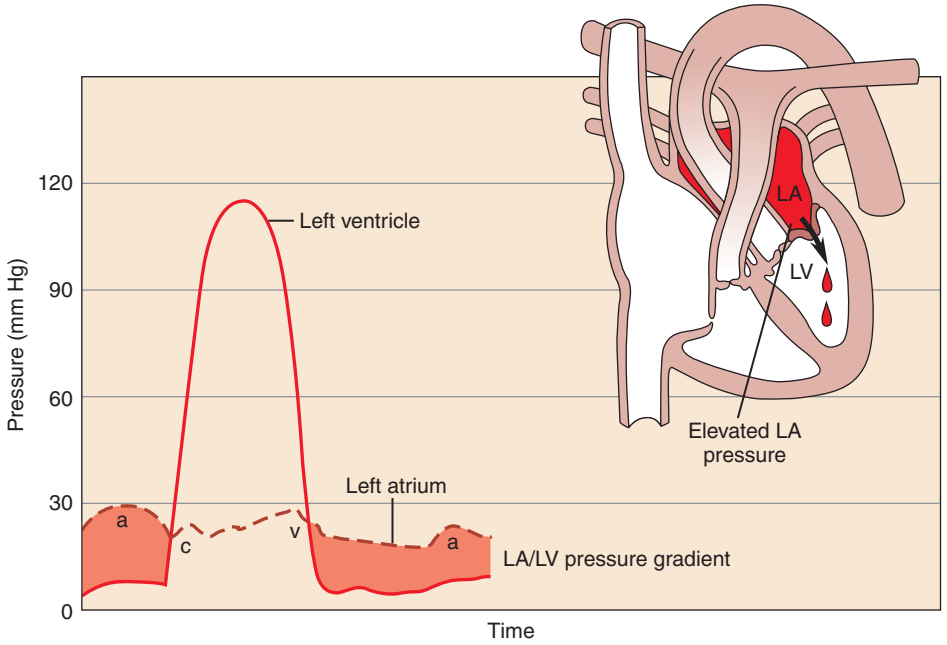


FIG 18.15 Mitral stenosis is characterized by an abnormal left atrial (LA)–to–left ventricular (LV) pressure gradient during ventricular diastole (shaded area).

TABLE 18.6 Defining Characteristics of Murmurs

Valve Disorder	Quality	Location, Radiation
Mitral stenosis	Low-pitched rumble, diastolic	At apex
Mitral regurgitation	Loud, pansystolic, high pitched, blowing	Loudest at apex, transmitted to left axilla
Aortic stenosis	Harsh, midsystolic, crescendo-decrescendo	Right second intercostal space, transmitted to neck
Aortic regurgitation	Faint, blowing, diastolic	Left sternal border, aortic area, apex

Mitral Stenosis

In *mitral stenosis* the flow of blood from the left atrium into the left ventricle is impaired. Mitral stenosis is therefore characterized by an abnormal left atrial–left ventricular pressure gradient during ventricular diastole (Fig. 18.15). Normally the pressures in the atrium and ventricle are nearly equal during ventricular diastole when the mitral valve is open. Fig. 18.15 shows that with mitral valve stenosis, atrial pressure remains higher than ventricular pressure throughout diastole. As the stenosis worsens, the pressure gradient often increases. In normal adults the area of the mitral valve orifice is 4 to 6 cm^2, and symptoms of stenosis do not appear until the orifice is narrowed to 2 cm^2. When the mitral valve orifice narrows to 1 cm^2, a critical stenosis is present and a pressure gradient of 20 mm Hg or more usually develops across the valve. Increased pressure work of the left atrium leads to atrial chamber enlargement and hypertrophy. Progressive narrowing of the mitral valve may lead to markedly elevated left atrial pressures and subsequent increased pulmonary vascular pressure. If uncorrected, mitral stenosis may result in chronic pulmonary hypertension, right ventricular hypertrophy, and right-sided heart failure.

The signs and symptoms of mitral stenosis are due to congestion of blood volume and increased pressure in the left atrium and pulmonary circulation, as well as decreased stroke volume of the left ventricle because of deficient filling. Symptoms are exacerbated by conditions that further decrease left ventricular filling such as an increased heart rate. Atrial dysrhythmias such as atrial fibrillation are common because of excessive atrial volume. Atrial enlargement and fibrillation also predispose to the

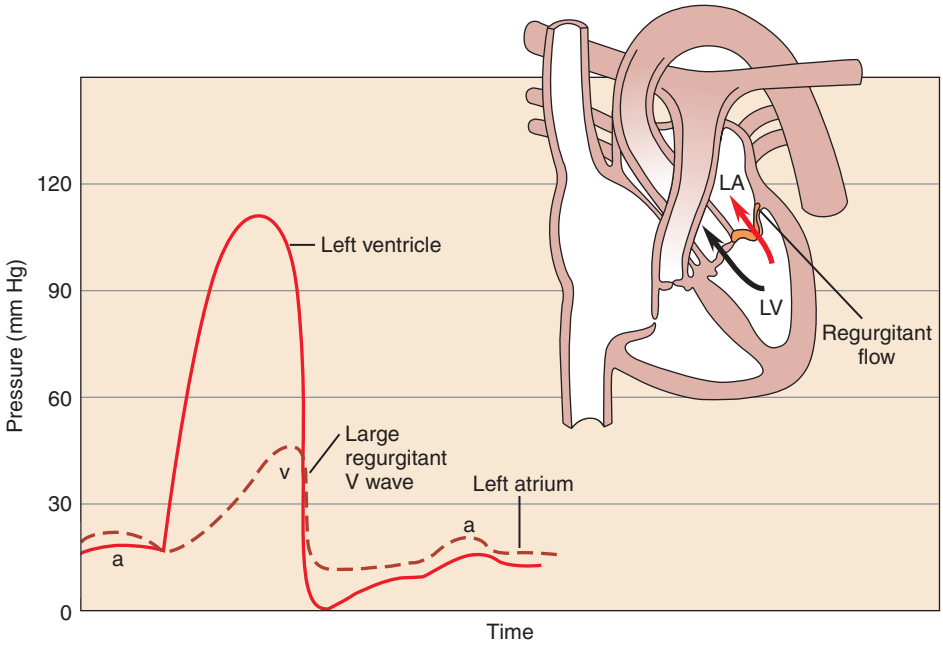


FIG 18.16 Mitral regurgitation causes characteristic giant V waves on the left atrial (LA) pressure monitor. *LV,* left ventricular.

development of atrial clots, which may dislodge and result in systemic embolization and stroke. Signs and symptoms of mitral stenosis secondary to pulmonary congestion may include orthopnea, cough, dyspnea on exertion, paroxysmal nocturnal dyspnea, abnormal breath sounds, and poor arterial oxygenation. Reduced left ventricular stroke volume may be apparent as fatigue, poor activity tolerance, and weakness. Exertional dyspnea is the most common complaint. Blood rushing through the narrowed mitral valve during ventricular diastole can sometimes be heard as a low-pitched, rumbling diastolic murmur at the heart's apex. In many cases, an opening snap may also be heard.

Mitral Regurgitation

Mitral regurgitation is characterized by backflow of blood from the left ventricle to the left atrium during ventricular systole. Elevation of left atrial volume and pressure by regurgitant flow leads to characteristic giant V waves on the atrial pressure monitor (Fig. 18.16). The severity of mitral insufficiency is related to the amount of left ventricular stroke volume that is regurgitant and depends, in part, on the aortic resistance to flow (afterload). A high afterload increases the amount of regurgitant flow. The left ventricle must pump a greater volume to compensate for the regurgitant flow and maintain an effective stroke volume. Both the left atrium and the left ventricle generally dilate and hypertrophy to compensate for the extra volume that they are required to pump. In most patients with mitral regurgitation, compensation is maintained for many years before symptoms occur. If severe and uncorrected, mitral regurgitation may eventually lead to left-sided heart failure. The signs and symptoms of mitral regurgitation are similar to those described for mitral stenosis and result from pulmonary congestion and poor cardiac output. Chronic weakness and fatigue are common complaints. The murmur of mitral regurgitation usually occurs throughout ventricular systole (pansystolic), radiates toward the left axilla, and has a high-pitched blowing character. The arterial pulse may be helpful in distinguishing the systolic murmur of mitral regurgitation from that of aortic stenosis. The upstroke of the pulse is sharp and full in mitral regurgitation, whereas it is weaker and delayed in aortic stenosis.

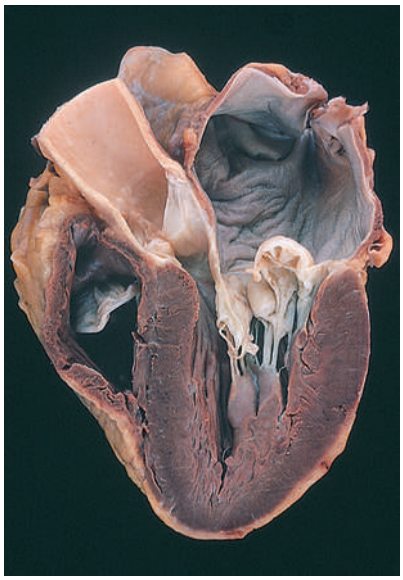


FIG 18.17 Appearance of mitral valve prolapse. Note how the valve balloons up into the left atrium. (From Kumar V et al, editors: *Robbins and Cotran pathologic basis of disease,* ed 9, Philadelphia, 2015, Saunders, p 557. Courtesy William D. Edwards, MD, Mayo Clinic, Rochester, MN.)

Mitral Valve Prolapse

Approximately 2% to 3% of the population have mitral valves that balloon into the left atrium during ventricular systole. This condition is called *mitral valve prolapse* (Fig. 18.17). Women are affected twice as often as men. In a great majority of cases, the disorder is asymptomatic and diagnosed only incidentally on routine physical examination. In some cases, the prolapse is sufficient to cause a degree of mitral regurgitation. The cause of this valvular abnormality is uncertain, although it is commonly associated with other connective tissue disorders such as Marfan syndrome or scoliosis.

The finding of 2 mm or more displacement of the mitral valve leaflets above the annulus on echocardiogram is an important diagnostic criterion. Mitral valve prolapse may be detected by a midsystolic click or systolic murmur. Individuals whose disease is symptomatic may experience palpitations, rhythm abnormalities, dizziness, fatigue, dyspnea, chest pain, or psychiatric manifestations such as depression and anxiety. The large majority of affected persons have no untoward effects, and most are unaware of their condition. Complications of mitral valve prolapse are relatively rare and include infective endocarditis, sudden cardiac arrest, cerebral embolic events, and progression to mitral regurgitation.

Disorders of the Aortic Valve

The primary disorders of the aortic valve are stenosis and regurgitation.

Aortic Stenosis

With the decline in incidence of rheumatic fever, the predominant cause of aortic stenosis is age-related calcification. The hallmark of this disorder is the formation of calcium deposits on the aortic cusps (see Fig. 18.14). Calcification is particularly common in patients with a congenital bicuspid aortic valve. Aortic calcifications accumulate over several decades and generally become clinically apparent in individuals 70 to 90 years old. Rheumatic heart disease, on the other hand, occurs primarily in children and young adults and now accounts for only a small percentage of cases of acquired aortic stenosis in the United States.

Aortic stenosis results in obstruction to aortic outflow from the left ventricle into the aorta during systole. This condition is characterized by a left ventricular–aortic pressure gradient during ventricular ejection (Fig. 18.18). The left ventricle produces high systolic pressure to overcome resistance of the stenotic aortic valve. The slow development of aortic stenosis allows the heart to maintain stroke volume by compensatory left ventricular hypertrophy. The combination of high left ventricular pressure and hypertrophy predisposes the heart to ischemia and attacks of anginal pain. Continued high left ventricular afterload from a stenotic aortic valve may lead to left-sided heart failure.

Critical obstruction is characterized by a mean systolic pressure gradient exceeding 40 mm Hg and an effective aortic valve orifice less than 25% of normal (≤1.0 cm^2 in an average-sized adult). The symptoms of aortic stenosis are due to diminished cardiac output, with pulmonary complications occurring later as the left ventricle fails. Syncope, fatigue, low systolic blood pressure, and faint pulses are common signs and symptoms. Angina occurs frequently in patients with critical aortic stenosis and is often initiated by exertion and relieved by rest. Angina is thought to occur because of thickening of the ventricular wall with reduced perfusion and high intraventricular wall tension. Syncope and "graying out" spells may occur when cerebral perfusion is inadequate. The onset of atrial fibrillation or heart block may precipitate worsening of symptoms. A characteristic murmur occurs during ventricular systole and varies in intensity, progressively getting louder and then diminishing (crescendo–decrescendo). The murmur of aortic stenosis generally radiates to the neck. The heart rate is usually slow to allow for a necessarily long ejection phase, and a prominent S_4 is usually present. Surgical correction is indicated for symptomatic aortic stenosis because medical therapy is not effective.

Aortic Regurgitation

Aortic regurgitation results from an incompetent aortic valve that allows blood to leak back from the aorta into the left ventricle during diastole. Causes of aortic regurgitation are similar to those of mitral regurgitation (see Table 18.5). Valvular incompetence may be secondary to an abnormal aortic valve or to aortic root dilation with widening of the aorta such that the valve leaflets no longer appose. Aortic root dilation is a more common cause of aortic regurgitation than primary valvular disease. The aorta may dilate because of degenerative changes with aging or as a consequence of connective tissue disease. The left ventricle becomes volume overloaded because it contains its usual preload, received from the atrium, plus regurgitant blood from the aorta. The left ventricle compensates for this extra volume work with hypertrophy and dilation. A larger-than-normal stroke volume is thus achieved and produces a high systolic blood pressure. Diastolic blood pressure is generally lower than normal because of rapid runoff of blood into the ventricle. The large stroke volume and rapid decline in diastolic blood pressure result in a bounding peripheral pulsation, and the head may bob with each systole (Fig. 18.19).

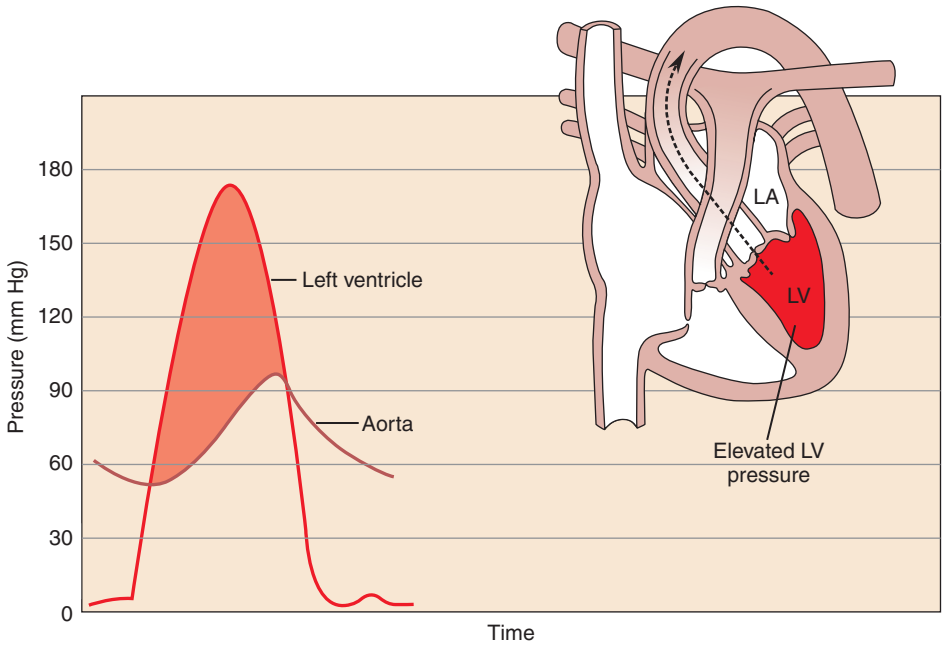


FIG 18.18 Aortic stenosis is characterized by an abnormal left ventricular (LV)–to–aortic pressure gradient (shaded area). *LA,* Left atrium.

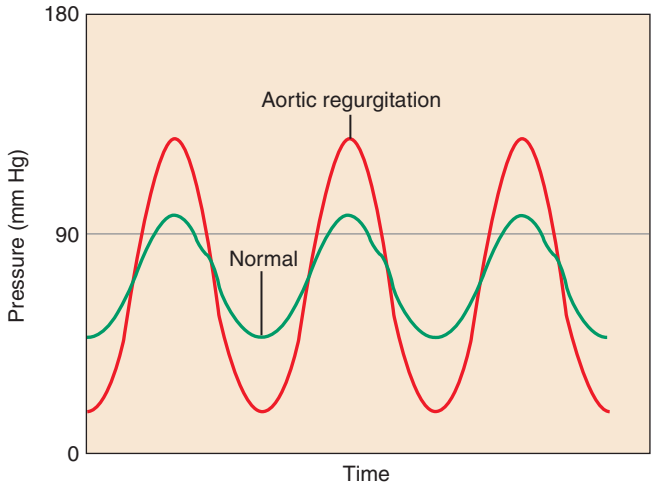


FIG 18.19 Typical arterial pressure fluctuation in aortic regurgitation showing a high systolic pressure and a low diastolic pressure.

Aortic insufficiency is characterized by a high-pitched blowing murmur during ventricular diastole. Patients may complain of palpitations and a throbbing or pounding heart because of the large ventricular stroke volume. The major complication of aortic regurgitation is left-sided heart failure as a result of the high ventricular workload. However, chronic aortic regurgitation is well tolerated for years, and asymptomatic individuals can delay valve replacement surgery. Acute aortic regurgitation is poorly tolerated and necessitates immediate correction.

Diseases of the Endocardium

Rheumatic Heart Disease

Rheumatic heart disease is an uncommon but serious consequence of rheumatic fever. The incidence of rheumatic fever has steadily declined in the United States, but the disease still affects an estimated 15 million people worldwide. Rheumatic fever is an acute inflammatory disease that follows infection with group A β-hemolytic streptococci.

Damage is due to an immune attack on the individual's own tissues. For incompletely understood reasons, antibodies against the streptococcal antigens are also directed against self tissues, possibly because of an immune hypersensitivity reaction resulting from cross-reactivity between streptococcal antigens and certain tissue molecules. Epitopes on the bacterial surface are similar to proteins on cardiac myosin, valve, skin, joint, and brain tissue. Most individuals with group A β-hemolytic streptococcal infection do not develop rheumatic fever, and a number of genetic and immune factors are associated with development of the disease.

The acute infection occurs primarily in children and is accompanied by fever and a sore throat. In only 3% of children with pharyngeal streptococcal infection does rheumatic fever eventually develop. Prompt initiation of antibiotic therapy is often effective in primary prevention of rheumatic fever. Rheumatic fever diffusely affects connective tissue in joints, the heart, and the skin. The central nervous system and kidney are also frequently involved. Inflammation of the heart usually includes all layers and results in carditis. Endocardial inflammation results in valvular swelling, erosions, and clumping of platelets and fibrin on valve leaflets. Scarring and shortening of valvular structures become progressively more severe. The myocardium and pericardium may show signs of rheumatic inflammation; however, if there is no associated valvular inflammation, the diagnosis is unlikely to be rheumatic fever. Other hallmarks of rheumatic fever include joint inflammation, involuntary movements (Sydenham chorea), and a distinctive truncal rash. An elevated antibody titer against streptococcal products (antistreptolysin O, anti-DNase B) may help confirm the diagnosis. Unfortunately, individuals who experience rheumatic fever have a high chance of recurrence if they have another pharyngeal streptococcal infection. Prophylactic antibiotic therapy is recommended for individuals who develop rheumatic fever.

Infective Endocarditis

Infective endocarditis is caused by invasion and colonization of endocardial structures by microorganisms with resulting inflammation. A variety of organisms are known to have an affinity for the endocardium and for the cardiac valves in particular. Valvular lesions include growths of microorganisms enmeshed in fibrin deposits. These growths are called *vegetations.* They may become quite large, interfering with valvular function and predisposing to embolus formation. Common bacterial culprits are several strains of *Streptococcus* and *Staphylococcus aureus.* A requisite for infective endocarditis is invasion of the bloodstream by infective organisms. The portal of entry may be obvious, as with an overt infection, intravenous drug abuse, or invasive surgical or dental procedures. Sometimes the source may be less obvious, such as the gastrointestinal tract or the oral cavity. Once the organism enters the circulation, several factors influence its ability to attack endocardial structures and cause disease.

Acute infective endocarditis may theoretically develop in any individual if host resistance is low, if the organism is highly virulent, and if the bacterial invasion is sufficiently large. Acute infective endocarditis usually affects individuals with previously normal valves and leads to death in a large percentage of patients. The overall mortality rate for infective endocarditis is relatively high; however, mortality rates vary with the type of infective organism. Intravenous drug abusers are particularly susceptible to acute infective endocarditis. *Subacute* infective endocarditis has a more insidious onset and generally affects individuals with some preexisting propensity for valvular colonization. The offending organisms are less virulent. Rheumatic heart disease, congenital heart abnormalities, mitral valve prolapse, calcified valves, and prosthetic valves are important predisposing factors. Immunosuppression and repeated exposure through intravenous drug abuse are other predisposing influences. *S. aureus* and *Staphylococcus epidermidis,* which colonize the skin, are common offenders in intravenous drug users. The valves on the right side of the heart may be infected in this population.

Organisms associated with subacute infective endocarditis usually are not virulent enough to attack normal healthy endocardium but are able to gain a foothold in hearts with some underlying predisposition. Preexisting cardiac disease may allow the formation of platelet–fibrin deposits on the valves because of abnormal or stagnant blood flow patterns. These deposits become the site of organism attachment. Antibodies against the invader may further assist attachment by causing clustering of organisms.

The diagnostic findings in both acute and subacute infective endocarditis are much the same. Large, bulky, bacteria-laden vegetations hang from the heart valves and adjacent endocardial surfaces (Fig. 18.20). In addition to the risk of embolization, vegetations may cause erosion or perforation of the underlying valve leaflet. In acute forms, adjacent myocardium may be eroded and abscessed. With time, valvular vegetations become fibrotic and calcified.

Unfortunately, the clinical features of subacute infective endocarditis are quite nonspecific, with low-grade fever the most consistent sign. Nonspecific fatigue, weight loss, and flulike symptoms may be the only clues. Positive blood cultures may help confirm the diagnosis. In contrast, acute infective endocarditis has a more obvious onset with fever, chills, malaise, and, frequently, a heart murmur. Complications such as valvular insufficiency, myocardial abscess, embolization, and renal disease generally

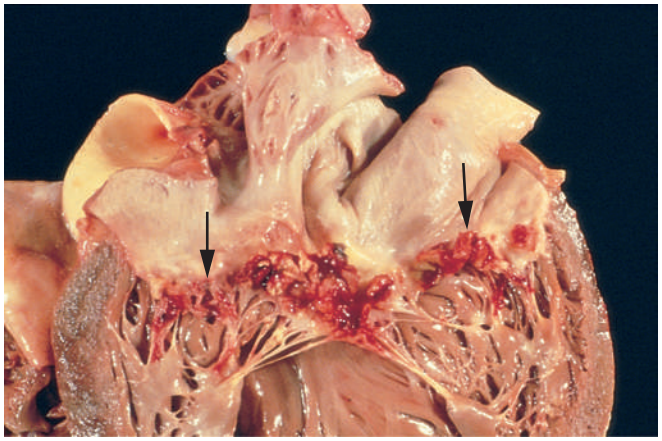


FIG 18.20 Mitral valve endocarditis from subacute bacterial infection with *Streptococcus viridans*. The left ventricle contains numerous abscesses formed by seeding from vegetations traveling in the coronary arteries. (From Kumar V et al, editors: *Robbins and Cotran pathologic basis of disease,* ed 9, Philadelphia, 2015, Saunders, p 560.)

occur early in the course of the disease. The interval between initiation of bacteremia and the onset of symptoms is less than 2 weeks in the majority of cases. Management of the acute and subacute types centers on antibiotic therapy, with surgical replacement of valves when indicated. Prevention through prophylactic antibiotic therapy in individuals at risk is an important consideration.

The endocardium is prey to many other disorders, such as systemic lupus erythematosus (SLE, an immunologic disease), calcium deposition secondary to renal disease, and nonbacterial thrombotic endocarditis secondary to hypercoagulable states associated with cancer.

KEY POINTS

- Valvular disorders are of two primary types. Failure of stenotic valves to open properly causes an abnormal pressure gradient across the valve and increases the pressure work of the heart. Regurgitant valves allow blood to flow backward across the valve and result in extra volume work for the heart.
- Mitral stenosis is characterized by a large left atrial–to–left ventricular pressure gradient during ventricular diastole. Mitral stenosis leads to left atrial and pulmonary congestion.
- Mitral regurgitation is characterized by large V waves in the left atrial pressure tracing and a loud systolic murmur that radiates to the left axilla. Mitral regurgitation increases the work of the left atrium and ventricle and can lead to left-sided heart failure.
- Aortic stenosis results in obstruction to the outflow of blood from the left ventricle. It is characterized by a large left ventricular–to–aortic pressure gradient and a crescendo–decrescendo murmur during systole that radiates to the neck. The extra pressure work can lead to left ventricular hypertrophy and failure.
- Aortic regurgitation is characterized by a high systolic and low diastolic blood pressure and a bounding pulse. The murmur of aortic regurgitation occurs during diastole. Left ventricular failure may result because of the high-volume work.
- Rheumatic heart disease results from immune-mediated damage to the endocardium after group A β-hemolytic streptococcal infection.
- Acute and subacute infective endocarditis results in the growth of bacteria-laden vegetations on heart valves. In addition to valvular erosion and scarring, embolization may occur.

BOX 18.3 Common Etiologic Agents of Myocarditis

Viral
Coxsackie virus/Enterovirus
Cytomegalovirus
Influenza virus
Human immunodeficiency virus

Bacterial
Chlamydia pneumoniae
Streptococcal species
Mycoplasma pneumonia
Neisseria meningococcus
Lyme disease

Fungal
Candida

Protozoal
Trypanosoma cruzi

Parasitic
Schistosomiasis
Trichinosis

Toxins
Anthracyclines
Cocaine
Ethanol

Hypersensitivity
Clozapine
Sulfonamides
Cephalosporins
Penicillins
Tricyclic antidepressants

Autoimmune Activation
Giant cell myocarditis
Inflammatory bowel disease
Sarcoidosis
Systemic lupus erythematosus
Wegener granulomatosis

Data from Cooper LT, Knowlton KU: Myocarditis. In Mann D, Zipes D, Libby P, Bonow R: *Braunwald's heart disease: a textbook of cardiovascular medicine,* ed 10, Philadelphia, 2015, Elsevier, p 1591.

MYOCARDIAL DISEASES

In addition to the diseases already discussed, which secondarily affect the myocardium as a consequence of inadequate blood supply or endocardial infection, two other categories of diseases of heart muscle are *myocarditis* and *cardiomyopathy.* Myocarditis is an inflammatory disorder of the heart muscle characterized by necrosis and degeneration of heart muscle cells. Cardiomyopathy includes several disorders of the heart muscle that may be genetic or acquired but are noninflammatory. The division of these categories is somewhat arbitrary; however, the clinical course of myocarditis is generally acute and stormy, with either recovery or death from cardiac failure occurring weeks to months after the onset of symptoms. In contrast, the cardiomyopathies generally evolve more insidiously over years, with few symptoms until the heart slips into failure.

Myocarditis

Myocarditis is characterized by inflammation, leukocyte infiltration, and necrosis of cardiac muscle cells. Causes of myocarditis are many and include microbial agents, several forms of immune-mediated disease, and several physical agents. The more common causes of myocarditis are listed in Box 18.3. The true incidence of myocarditis is unknown because the diagnosis relies largely on circumstantial evidence.

Most cases of myocarditis in the Northern Hemisphere are associated with viral infections. Cardiac involvement generally appears days or weeks after a viral infection elsewhere in the body. Documenting a viral cause is often impossible, but a rising antibody titer supports the diagnosis. The mechanism of viral myocarditis is incompletely understood. Direct viral cytotoxicity may occur to some extent, and the virus may evoke an immune response directed against the heart. Most investigators currently believe the second mechanism to be most important. In some countries, nonviral organisms are commonly

associated with myocarditis. For example, the protozoan *Trypanosoma cruzi,* which is endemic in areas of Central and South America, infects tens of thousands worldwide, although the incidence has decreased dramatically in recent decades. A myocarditis called *Chagas disease* eventually develops in a number of infected individuals and is an important cause of cardiovascular death in endemic countries.

In some cases of myocarditis, the immune system reaction against the myocardium appears to be the primary cause. Antibodies or activated lymphocytes are formed against heart tissue. Several drugs, including penicillin, tend to evoke a hyperactive immune response in some individuals and may cause an allergic-type reaction that affects the myocardium. Myocarditis accompanies some autoimmune disorders, such as SLE and polymyositis. Toxins and chemical causes of myocarditis include cocaine, chemotherapeutic agents, snake bite and insect venoms, lead, and numerous others. Regardless of the specific cause, inflammation of cardiac muscle is characteristic.

Acute myocarditis is commonly characterized by left ventricular dysfunction or general dilation of all four heart chambers. The ventricular myocardium is "flabby" with patchy or diffuse necrotic lesions. The heart muscle appears inflamed and edematous with WBC infiltrates. Endocardial structures are usually normal. The clinical course of acute myocarditis varies in severity from asymptomatic to rapidly evolving heart failure. Generalized symptoms related to the inflammatory process may be present, as well as electrocardiographic changes caused by myocardial cell death. Common presenting symptoms include fatigue, dyspnea on exertion, and dysrhythmia with associated palpitations. Many persons recover completely, whereas others have progressive disease that is manifested years later as dilated cardiomyopathy. Thus myocarditis and the cardiomyopathic forms of myocardial disease overlap and are difficult to separate. Therapy is supportive and usually includes therapy for heart failure (see Chapter 19). Immunosuppressive therapy may be considered for myocarditis associated with autoimmune disease or hypersensitivity reactions.

Cardiomyopathy

Cardiomyopathies can be classified by etiology or by functional impairments. Recent advances in genetic analysis of cardiomyopathies have led to classification systems that include genetic linkages in addition to morphologic changes in the heart. This has been called the *phenome–genome system of classification.* Those cardiomyopathies with uncertain cause are classified based on their predominant morphologic features, including dilated, infiltrative, hypertrophic, and restrictive forms (Fig. 18.21). The terms *primary cardiomyopathy* for dysfunction of unknown cause and *secondary cardiomyopathy* for myocardial dysfunction of known cause are also in clinical use (Table 18.7). Definitions of primary cardiomyopathy usually exclude hypertensive, ischemic, congenital, valvular, pericardial, and inflammatory myocardial disorders; however, classification of cardiomyopathy continues to evolve as more is understood about the genetic contributions to the condition.

Dilated Cardiomyopathy

Dilated or congestive cardiomyopathy (DCM) is characterized by cardiac failure associated with dilation of one or both ventricular chambers. Numerous factors are suspected in the initiation of dilated cardiomyopathy, including alcohol toxicity, genetic abnormality, pregnancy, and postviral myocarditis. Alcohol and its metabolites are toxic to heart muscle cells and are associated with thiamine and other nutritional deficiencies. *Peripartum cardiomyopathy* is the term applied to cases of dilated cardiomyopathy discovered just before or just after delivery. The etiology is unclear; however, inflammatory factors are implicated, and a high incidence of lymphocytic activation has been reported. A risk of recurrence in subsequent pregnancies has been noted.

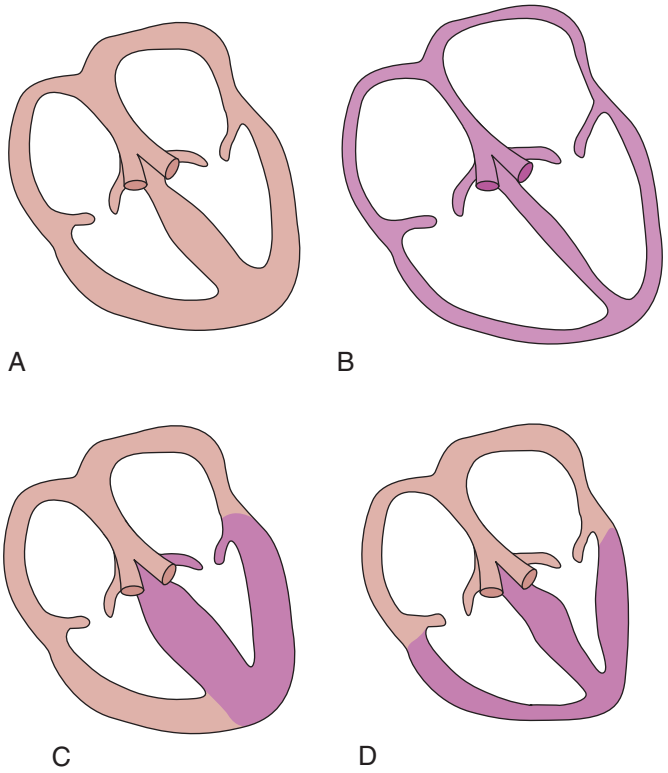


FIG 18.21 The three types of cardiomyopathy. **A,** Normal heart. **B,** Dilated cardiomyopathy demonstrating enlargement of all four chambers. **C,** Hypertrophic cardiomyopathy showing a thickened left ventricle. **D,** Restrictive cardiomyopathy characterized by a small left ventricular volume.

In some cases, dilated cardiomyopathy runs in families and has a presumed genetic basis. At least 20% to 35% of patients with DCM have a first-degree relative with signs of the disease. Inherited genetic defects in the protein structure of the myocardial sarcomere and other cytoskeleton proteins appear to be contributory in most genetic forms of DCM.

Postviral myocarditis is an attractive pathogenic mechanism for dilated cardiomyopathy, as previously discussed. Myocardial biopsy specimens often reveal signs of inflammatory injury; however, progression from acute myocarditis to dilated cardiomyopathy is difficult to document. A variety of other causes have been proposed as well. In fact, *dilated cardiomyopathy* is a bit of a catch-all term invoked to cover cases of dilated congestive failure having no well-defined origin. The clinical picture is one of slowly progressing biventricular heart failure with low ejection fraction (EF).

Hypertrophic Cardiomyopathy

In contrast to dilated cardiomyopathy, hypertrophic cardiomyopathy (HCM) is characterized by a thickened, hyperkinetic ventricular muscle mass. The hypertrophy is often not uniform throughout the heart, and in about 25% of patients the septum is most affected and causes a dynamic aortic outflow obstruction. The left ventricle is usually more involved than the right.

Substantial evidence suggests that this form of cardiomyopathy is transmitted genetically in an autosomal-dominant pattern. Abnormalities of cardiac sarcomere proteins have been identified in several familial forms of the disease. Abnormalities in genes coding for myosin heavy chain, myosin-binding protein C, tropomyosin, and troponin T account for 70% to 80% of HCM cases. The remaining cases are thought to

TABLE 18.7 Classification of the Cardiomyopathies

Disorder	Description
Dilated cardiomyopathy	Dilatation and impaired contraction of left or both ventricles. Caused by familial/genetic, viral and/or immune, alcoholic/toxic, or unknown factors or is associated with recognized cardiovascular disease.
Hypertrophic cardiomyopathy	Left and/or right ventricular hypertrophy, often asymmetric, which usually involves interventricular septum. Mutations in sarcoplasmic proteins cause disease in many patients.
Restrictive cardiomyopathy	Restricted filling and reduced diastolic size of either or both ventricles with normal or near-normal systolic function. Is idiopathic or associated with other disease (e.g., amyloidosis, endomyocardial disease).
Dysrhythmogenic right ventricular	Progressive fibrofatty replacement of right and, to some degree, left ventricular cardiomyopathic myocardium. Familial disease is common.
Unclassified cardiomyopathy	Diseases that do not fit readily into any category. Examples include systolic dysfunction with minimal dilatation, mitochondrial disease, and fibroelastosis.
Specific Cardiomyopathies	
Ischemic cardiomyopathy	Presents as dilated cardiomyopathy with depressed ventricular function not explained by extent of coronary artery obstructions or ischemic damage.
Valvular cardiomyopathy	Presents as ventricular dysfunction that is out of proportion to abnormal loading conditions produced by valvular stenosis and/or regurgitation.
Hypertensive cardiomyopathy	Presents with left ventricular hypertrophy with features of cardiac failure attributable to systolic or diastolic dysfunction.
Inflammatory cardiomyopathy	Cardiac dysfunction as a consequence of myocarditis.
Metabolic cardiomyopathy	Includes a wide variety of causes, including endocrine abnormalities, glycogen storage disease, deficiencies (such as hypokalemia), and nutritional disorders.
General systemic disease	Includes connective tissue disorders and infiltrative diseases such as sarcoidosis and leukemia.
Muscular dystrophies	Includes Duchenne, Becker-type, and myotonic dystrophies.
Neuromuscular disorders	Includes Friedreich ataxia, Noonan syndrome, and lentiginosis.
Sensitivity and toxic reactions	Includes reactions to alcohol, catecholamines, anthracyclines, irradiation, and others.
Peripartal cardiomyopathy	First becomes manifest in peripartum period, but it is likely a heterogeneous group.

Derived from *Circulation* 2005;112:1825–1852. ACC/AHA 2005 Guideline Update for the Diagnosis and Management of Chronic Heart Failure in the Adult. Reprinted with permission. Table: Stages in the Development of HF/recommended therapy by stage. Copyright © 2007 American Heart Association.

have mutations in other cytoskeletal proteins. Although the genetic basis of hypertrophic cardiomyopathy is well established, the events leading from cytoskeletal protein mutations to clinical disease are poorly understood.

There is wide variation in the expression of HCM. It may be asymptomatic or may be associated with symptoms of ventricular outflow obstruction or impaired diastolic filling. Outflow obstruction is particularly problematic when the myocardial hypertrophy is localized in the subaortic septal region. Strenuous activity may precipitate profound outflow obstruction, negligible stroke volume, and sudden death. Other factors contributing to reduced stroke volume are the smaller intraventricular chamber size and a noncompliant ventricle characteristic of diastolic dysfunction with preserved EF. Common symptoms of HCM are dyspnea and angina. Microscopically, the hypertrophied muscle cells appear disorganized and haphazardly oriented into disarray, rather than the usual linear arrangement.

The clinical course of HCM is variable, with most patients experiencing little change in cardiac function over many years. Surgery to thin the septal thickening (myectomy) is rarely performed any longer because this technique has been superseded by drug therapy. In general, drugs that increase myocardial contractility or heart rate are avoided because they further impair diastolic filling and worsen aortic outflow obstruction. β-Adrenergic antagonists and calcium channel–blocking drugs may be used to dampen the hypercontractility. Normal life expectancy is possible. Myocardial ischemia is common, and the risk of sudden cardiac arrest is significant. Patients who die suddenly from HCM usually have fatal dysrhythmias that may occur when sedentary, but are frequently associated with vigorous physical exertion.

Restrictive Cardiomyopathy

Restrictive cardiomyopathy (RCM) is the rarest form and is characterized by a stiff, fibrotic ventricle with impaired diastolic filling. Familial forms are rare, and RCM can frequently be attributed to specific clinical disorders, the most common being amyloidosis—abnormal deposition of amyloid protein in tissues. The mechanisms proposed for development of other cases include sarcoidosis, glycogen storage diseases, iron overload, scleroderma, radiation injury, and exposure to agents that promote fibrosis. Regardless of the specific cause, the myocardium becomes fibrosed, rigid, and noncompliant. The major difficulty is restricted diastolic filling with resultant low stroke volume and heart failure. Exercise intolerance, dyspnea, and weakness may be present. RCM is difficult to manage effectively because no specific therapy is available for most types.

KEY POINTS

- Myocarditis is an inflammatory disorder characterized by scattered necrotic and dead heart muscle cells. Most cases are associated with viral infection. The major complication of myocarditis is left ventricular dysfunction and dilation of the heart chambers, with reduced contractility.
- Cardiomyopathies encompass a number of disorders of heart muscle that may be genetic or acquired. In some cases the exact cause is unknown.
- Dilated cardiomyopathy is characterized by enlargement of both ventricular chambers, reduced contractility, and low ejection fraction (EF).

- Hypertrophic cardiomyopathy (HCM) primarily affects the left ventricle and ventricular septum. Genetic mutations in sarcomere proteins are suspected in most cases. Conditions that increase contractility of the heart (exercise, drugs) can result in obstruction of ventricular outflow and reduced cardiac output. EF is preserved, and diastolic dysfunction is predominant. Patients with HCM have a significant risk of sudden cardiac arrest.
- Restrictive cardiomyopathy (RCM) is characterized by a stiff, fibrotic left ventricle that resists diastolic filling. Decreased cardiac output and left-sided heart failure can result.

PERICARDIAL DISEASES

Pericardial disorders are rarely isolated processes of primary cause; rather, they are sequelae of other disorders such as systemic infection, trauma, metabolic derangement, or neoplasia. Despite the diversity of causative factors, pericardial involvement may be asymptomatic, manifested as an accumulation of fluid in the pericardial sac, or manifested as painful inflammation of pericardial structures.

Pericardial Effusion

An accumulation of fluid in the pericardial sac is called *pericardial effusion.* Normally, the pericardial space contains only 30 to 50 mL of thin, clear fluid. Under pathologic conditions, as much as 500 mL may accumulate. The compositions of the usual types of effusions are as follows:

Serous—a transudate secondary to heart failure or hypoproteinemia
Serosanguineous—a mixture of serous fluid and blood that may follow blunt chest trauma, heart surgery, or CPR
Blood—hemopericardium usually resulting from penetrating trauma to the heart
Purulent – infected fluids containing WBC (pus)

Cardiac Tamponade

The accumulation of pericardial fluid is generally without clinical significance except as an indicator of underlying disease processes. However, if the fluid accumulation is large or occurs suddenly, the life-threatening condition of cardiac tamponade may ensue. *Tamponade* refers to external compression of the heart chambers such that filling is impaired.

Signs and symptoms of cardiac tamponade include reduced stroke volume and compensatory increases in heart rate. Systemic venous congestion occurs because blood is prevented from entering the compressed heart by way of the superior and inferior venae cavae. Venous congestion may be apparent as distended neck veins. Changes in intrathoracic pressure during respiration may have exaggerated effects on cardiac filling. The presence of waxing and waning of blood pressure in synchrony with respiration is called *pulsus paradoxus.* Significant pulsus paradoxus is usually defined as a difference of 10 mm Hg or more in systolic blood pressure between inspiration and expiration. Other manifestations of tamponade include rising filling pressures in the heart chambers, muffled heart sounds, dull chest pain, diminished electrocardiographic amplitude, and a compressed cardiac silhouette on radiographs. The three classic findings in cardiac tamponade—hypotension, distended neck veins, and muffled heart sounds—have been called *Beck's triad.*

Treatment is aimed at relieving the pericardial pressure by aspirating the offending fluid *(pericardiocentesis).* Failure to manage tamponade may result in drastically reduced diastolic filling, cardiovascular collapse, and death. Pericardial effusions occasionally are detected by chance on chest x-rays being performed for other purposes. Nonsymptomatic pericardial effusions may be aspirated or merely monitored and allowed to resolve spontaneously. Treatment is directed at the underlying cause of the effusion.

BOX 18.4 Categories of Pericardial Disease and Selected Specific Causes

Idiopathic
Infectious
Viral (echovirus, coxsackievirus, adenovirus, cytomegalovirus, hepatitis B, infectious mononucleosis, HIV/AIDS)
Bacterial (*Pneumococcus, Staphylococcus, Streptococcus, Mycoplasma,* Lyme disease, *Haemophilus influenzae, Neisseria meningitidis,* and others)
Mycobacteria (*Mycobacterium tuberculosis, Mycobacterium avium-intracellulare*)
Immune-Inflammatory
Connective tissue disease (systemic lupus erythematosus, rheumatoid arthritis, scleroderma, mixed)
Late posttrauma
Drug induced (procainamide, hydralazine, isoniazid, cyclosporine, others)
Neoplastic Disease
Secondary: breast and lung carcinoma, lymphomas, Kaposi sarcoma
Radiation Induced
Early Post–Cardiac Surgery and Post–Orthotopic Heart Transplantation
Hemopericardium
Trauma
Post–myocardial infarction free wall rupture
Device and procedure related: percutaneous coronary procedures, implantable defibrillators, pacemakers, post–dysrhythmia ablation, post–atrial septal defect closure, post–valve repair or replacement
Dissecting aortic aneurysm
Trauma
Blunt and penetrating, post–cardiopulmonary resuscitation

Data from LeWinter M, Hopkins WE: Pericardial diseases. In Mann D, Zipes D, Libby P, Bonow R: *Braunwald's heart disease: a textbook of cardiovascular medicine,* ed 10, Philadelphia, 2015, Elsevier, p 1638.

Pericarditis

Inflammation of the pericardium originates from a variety of causes (Box 18.4). Rarely is the pericardium the primary site of disease. Pericarditis is often categorized as acute or chronic; chronic pericarditis refers to a healed stage of the acute form that results in chronic pericardial dysfunction.

Acute Pericarditis

The majority of cases of acute pericarditis are of uncertain etiology, and most are presumed to be viral. Uncomplicated acute pericarditis typically resolves spontaneously within 2 weeks, and nonsteroidal antiinflammatory drugs and colchicines may be the only therapies needed. A small number of cases may be complicated by significant pericardial effusion or by persistent or recurrent inflammation and may require hospitalization for more thorough diagnostic investigation to determine the specific etiology and more intensive therapy.

The symptoms of acute pericarditis are associated with the systemic effects of inflammation and pericardial damage and include fever, leukocytosis, malaise, and tachycardia. Acute pericarditis almost always presents with chest pain and may be confused with anginal pain. Adhesion and friction between the visceral and parietal pericardial layers cause pain that may radiate to the back and be associated with esophageal

discomfort and dysphagia (difficulty swallowing). Rubbing of the pericardial layers may be heard as a friction rub. The rub can be transient and intermittent and may sound squeaky or like scratchy sandpaper. Epicardial injury from pericarditis may be apparent on the ECG as ST-segment elevation. Treatment is generally symptom oriented and includes medications to relieve pain and minimize inflammation.

In the past, MI was a common cause of pericarditis, but the incidence of both acute post-MI pericarditis and the delayed pericarditis of Dressler syndrome has decreased with the advent of early reperfusion therapy for MI. Acute post-MI pericarditis occurs within a few days of an MI in which the infarction involves the epicardial surfaces of the heart. The inflammation of the necrotic heart muscle extends to adjacent pericardial structures, causing them to become inflamed as well. The degree of pericardial involvement reflects the size of the MI.

Chronic Pericarditis

Healing of an acute form of pericardial inflammation may result in chronic (healed) pericardial dysfunction of two principal kinds: adhesive mediastinopericarditis and constrictive pericarditis. Adhesive mediastinopericarditis is usually a consequence of suppurative or caseous pericarditis or a complication of previous cardiac surgery. It may also follow significant irradiation of the chest. The pericardial sac is destroyed and the external aspect of the heart adheres to surrounding mediastinal structures. The workload of the heart increases significantly because contraction is opposed by the attached surrounding structures.

Constrictive pericarditis may be a result of previous suppurative or caseous pericarditis, commonly secondary to tuberculosis. However, in many cases the cause of pericardial dysfunction is unknown. The pericardial sac becomes dense, nonelastic, fibrous, and scarred. It encases the heart like a stiff cage and impairs diastolic filling. The constrictive process generally occurs slowly and may be quite advanced by the time symptoms occur.

Symptoms may include exercise intolerance, weakness, fatigue, and systemic venous congestion. Treatment is aimed at relieving the constriction by removal of pericardium (pericardectomy) and administration of inotropic agents to improve cardiac contractility.

KEY POINTS

- Large accumulations of pericardial fluid can result in cardiac tamponade. External compression of the heart chambers impairs diastolic filling and results in decreased stroke volume. Distended neck veins, hypotension, muffled heart sounds, pulsus paradoxus, and elevated and equalized intrachamber pressures are indicative.
- Acute pericarditis causes sticking and rubbing of the visceral and parietal pericardial layers. A friction rub, pain radiating to the back, esophageal discomfort, and generalized signs of inflammation are usually present.
- Chronic pericarditis can lead to destruction of the pericardial sac with adhesion of the heart to surrounding mediastinal structures. Cardiac contraction may be impaired.
- Constrictive pericarditis results in a fibrous, scarred pericardium that restricts cardiac filling.

CONGENITAL HEART DISEASES

Congenital heart disease is an abnormality of the heart that is present from birth. A wide variety of defects have been described, and only the pathophysiology of the most common defects will be included here. A brief description of fetal cardiac development is a necessary prelude to a discussion of congenital heart diseases.

Embryologic Development

Development of the heart involves a complex orchestration of formation and resorption of structures. Abnormalities in the development of four important heart structures are at the root of most of the common heart defects: (1) development of the atrial septum, (2) development of the ventricular septum, (3) division of the main outflow tract (truncus arteriosus) into the pulmonary and aortic arteries, and (4) development of the valves. Each of these processes is briefly reviewed.

The primitive heart begins as an enlarged tube much like a blood vessel. The tube has three layers. The inner luminal layer is thin and composed of endothelial cells. This layer will eventually line the inner chambers of the heart and valves. The outermost layer is also thin and is called the *myoepicardial mantle.* The outer mantle will form the epicardial and muscular structures of the heart. In between these two thin layers of cells is a thick layer of gelatinous substance called *cardiac jelly.* Cardiac jelly is the precursor to endocardial cushion tissue, which is important in the formation of membranes in the heart, including the septa that separate the four chambers of the heart.

By day 23, the heart tube begins to beat. The tube folds on itself to form an asymmetric loop structure (Fig. 18.22). One bulge of the loop forms a primitive single atrium, another forms the future left ventricle, and a third forms the future right ventricle and common ventricular outflow tract (truncus arteriosus). The common atrium is divided into right and left atria by growth of the interatrial septum. Atrial septation occurs in several steps. First, the *septum primum* is formed passively in the superior surface by an indentation caused by the overlying truncus arteriosus (Fig. 18.23). Next, the superior and inferior endocardial cushions grow and extend toward each other. These flaps of tissue overlap but do not fuse so that blood can pass through from the right atrium to the left atrium. A reverse in the direction of flow, as occurs at birth, normally would push the flap shut and close the hole (see Pediatric Considerations box). This flaplike opening, called the *ostium secundum,* remains open throughout fetal life and is later called the *foramen ovale.*

Septal formation between the ventricles follows a similar pattern. The lower portion of the interventricular septum is formed by circular

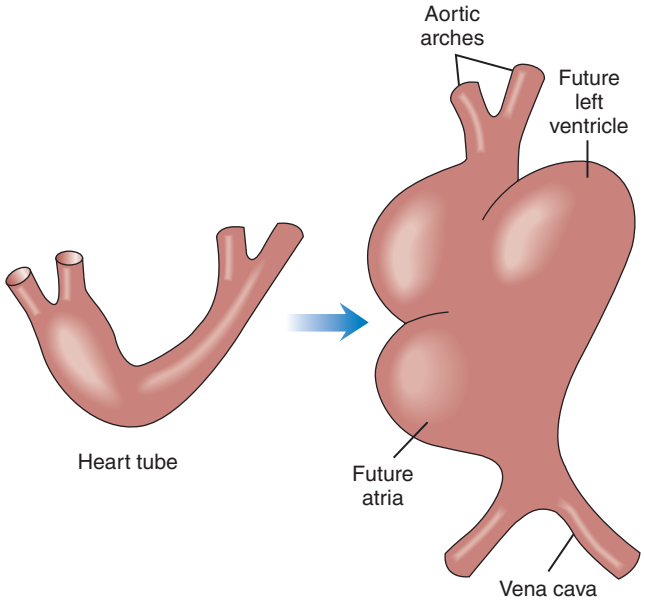


FIG 18.22 Asymmetric loop structure in early embryonic development of the heart.

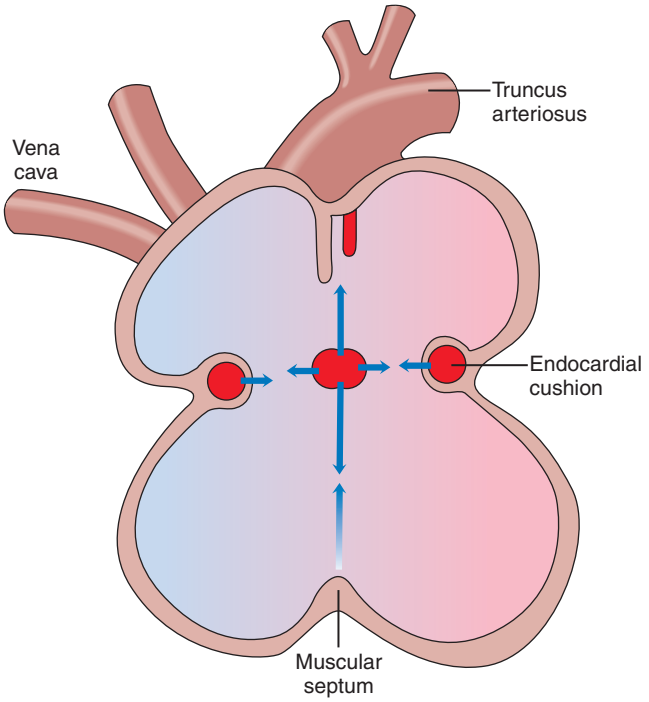


FIG 18.23 Formation of the intracardiac septa.

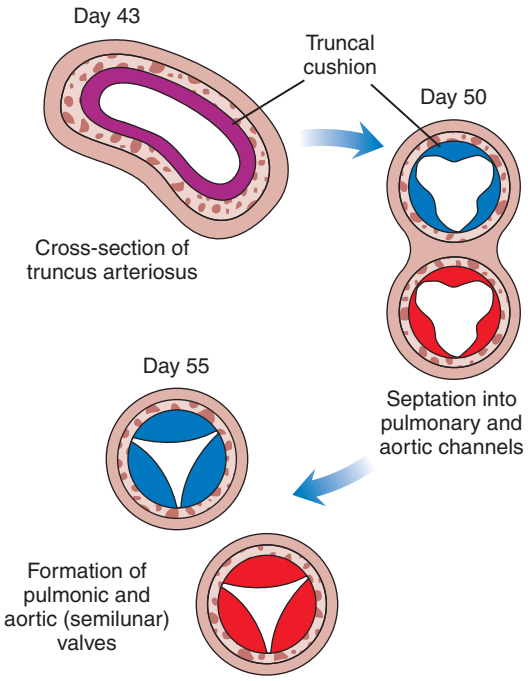


FIG 18.24 Septation of the truncus arteriosus and formation of the semilunar valves.

growth and fusing of the muscular ventricular walls. Then the muscular septum proliferates upward toward the atria. The inferior endocardial cushion tissue also grows downward to meet the uplifting muscular septum (see Fig. 18.23).

At about the same time that the atrial and ventricular septal structures are being elaborated, the common ventricular outflow tract, the truncus arteriosus, is divided into the pulmonary and aortic channels. This process is accomplished by growth and eventual fusion of mounds of endocardial cushions located in the wall of the truncus arteriosus. The truncal endocardial cushions go on to form the semilunar valves as well (Fig. 18.24).

The atrioventricular septum and valves are similarly formed by growth and fusion of the right and left lateral cushions. Superior and inferior cushions also contribute to formation of the septum between the atria and ventricles. Leaflets of the valves are initially formed by lumps of cushion material, which are replaced by muscle tissue from the ventricular wall. The muscle tissue also forms the chordae tendineae and papillary muscle structures. Eventually, the muscle cells of the leaflets and chordae tendineae are replaced by tough, fibrous connective tissue.

When the embryonic heart is fully developed, two important passageways still permit blood flow to bypass the lungs (Fig. 18.25). The foramen ovale lies between the left and right atria and allows blood to bypass the right ventricle. Blood flows right to left through the atrial opening because the pressure in the left atrium is low. High resistance of the deflated lungs also causes right ventricular pressure to be high, which impedes right ventricular filling. The other important structure is the ductus arteriosus, a channel that connects the pulmonary artery and the aorta. Blood flows from the pulmonary artery into the aorta during fetal life because of high vascular resistance in the collapsed lungs. Both these communications generally close after birth when the lungs inflate and the resistance on the right side of the heart falls. Clamping the umbilical cord also serves to increase systemic vascular resistance, which further augments the reverse in pressure gradient, with left heart pressures now exceeding those on the right.

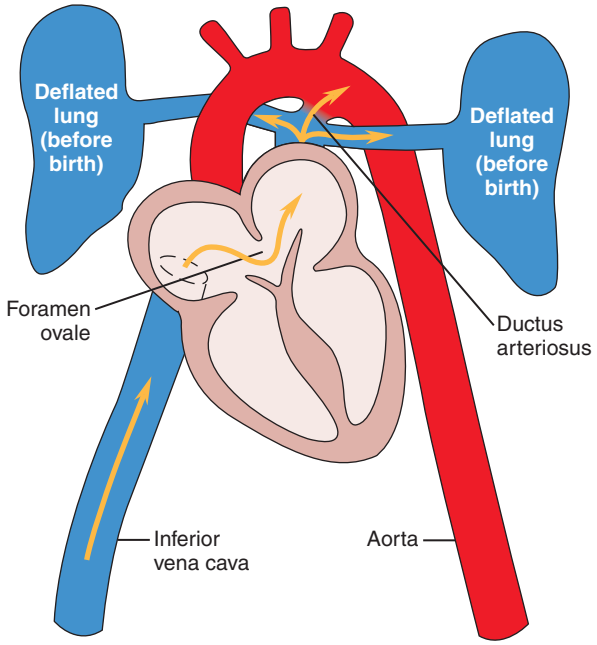


FIG 18.25 Fully developed embryonic heart showing the foramen ovale and ductus arteriosus. These structures allow blood to bypass the pulmonary circulation during fetal life.

Etiology and Incidence of Congenital Heart Disease

Congenital heart disease is the most common heart disorder in children, with an overall incidence of about 1% to 5% of all live births. The most common heart defects are listed in Table 18.8, with approximate frequencies of occurrence. In the majority of cases, the cause of the heart defect is unknown. Multifactorial inheritance with both genetic and environmental influences is probable. Abnormalities in several genes that

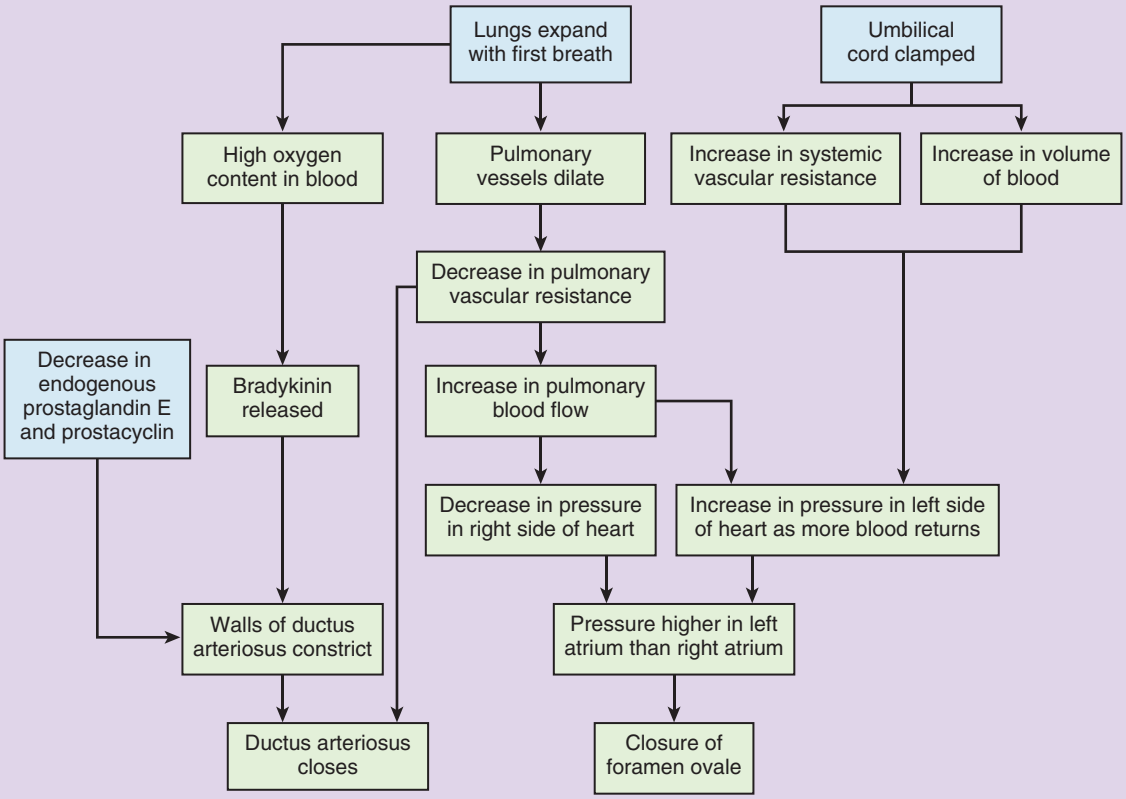
PEDIATRIC CONSIDERATIONS

Changes in the Newborn Heart

When a newborn takes his or her first breath, several changes occur. The alveoli of the lungs expand with the inspired oxygen dilating the pulmonary vessels, which decreases pulmonary vascular resistance. Pulmonary blood flow increases because of this dilation. The increased pulmonary blood flow is from the right side of the heart. Because of this increased flow, the pressures in the right atrium, right ventricle, and pulmonary arteries are decreased. Simultaneously, the umbilical cord is clamped, which increases systemic vascular resistance and increases blood volume. The pressure in the left side of the heart increases as more blood returns to the left atrium from the pulmonary veins related to the clamping. Because the pressure gradient has shifted with higher pressure in the left side of the heart, the circulation of blood through the fetal shunts is reversed.

Foramen ovale closure occurs because of the decreased blood flow from the placenta related to cord clamping. The blood flow from the placenta holds the foramen ovale open before birth. The change in the pressure gradient closes the foramen ovale at or soon after birth. When this closure occurs, blood from the right ventricle flows entirely into the pulmonary circulation.

Increased oxygen concentration of the blood is the critical factor in closure of the ductus arteriosus. Bradykinin is signaled to be released by high oxygen concentration in blood from the initial aeration of the lungs. Bradykinin has contractile effects on smooth muscle and forces the ductus arteriosus walls to constrict. The secretion of endogenous prostaglandin E and prostacyclin, which maintain the patency of the ductus arteriosus during gestation, is decreased after birth, which further diminishes the opening of the ductus arteriosus. In addition, decreased pulmonary vascular resistance decreases the blood flow from the ductus arteriosus. Functional closure of the ductus arteriosus occurs 4 days after birth, but may be delayed in preterm or ill infants.



code for transcription factors increase the risk for congenital anomalies (Table 18.9). Very few cases of congenital malformation can be clearly attributed to environmental factors. Maternal rubella during the first trimester of pregnancy is the best documented environmental cause of heart defects. A large number of cardiac teratogens are suspected from animal studies, including hypoxia, ionizing radiation, and heavy alcohol consumption.

Less than 15% of congenital heart defects can be attributed to genetic or chromosomal abnormalities. A twofold to tenfold increase in the incidence of congenital heart defects is seen in siblings. And several heart defects also occur more frequently in males. Monozygotic twin pregnancies have double the incidence of heart defects compared with singleton pregnancies, but usually only one of the pair is affected even though their genotypes are identical. Thus a complex interplay between genetic and environmental influences is probable and as yet poorly understood.

Pathophysiology of Congenital Heart Disease

The many forms of congenital heart anomalies result in two primary pathologies: *shunts* and *obstructions*. A shunt denotes an abnormal path of blood flow through the heart or great vessels. The shunt may be further characterized as *right-to-left* or *left-to-right* to indicate the direction of abnormal blood flow. Right-to-left shunts allow unoxygenated blood from the right side of the heart to enter the left side and systemic circulation without first passing through the lungs. Infants with right-to-left shunting of blood generally have some degree of cyanosis because of the decreased oxygen content of the arterial blood (cyanotic defect). Conversely, a left-to-right shunt occurs when

oxygenated blood from the left side of the heart or aorta flows back into the right side to be recirculated through the lungs. The blood reaching the systemic circulation is oxygenated and the infant is not cyanotic (acyanotic defect). However, the right side of the heart has an increased workload because of the extra shunt blood. In time, the overload of the right side of the heart can result in right ventricular hypertrophy and high right-sided heart pressures. A left-to-right shunt may then progress to a more dangerous right-to-left shunt when right heart pressures exceed left heart pressures. Congenital disorders causing abnormal blood flow through the heart include atrial septal defect, ventricular septal defect, patent ductus arteriosus, tetralogy of Fallot, transposition of the great arteries, truncus arteriosus, and tricuspid atresia.

Some heart anomalies produce obstructions to blood flow because of abnormal narrowings. Stenosis or atresia (failure to develop) of valves and coarctation of the aorta are the most common obstructive defects. Obstructions do not result in cyanosis but generally increase the workload of the affected chamber. Heart failure is a potential consequence of congenital heart defects and presents differently in infants and children than in adults (Box 18.5).

TABLE 18.8 Relative Frequency of Occurrence of Cardiac Malformations at Birth

Disease	Percentage
Ventricular septal defect	42
Atrial septal defect	10
Pulmonic stenosis	8
Patent ductus arteriosus	7
Tetralogy of Fallot	5
Coarctation of the aorta	5
Atrioventricular septal defect	4
Aortic stenosis	4
Complete transposition of the great arteries	4
Persistent truncus arteriosus	1
Anomalous pulmonary venous connection	1
Tricuspid atresia	1

Based on 44 published studies. Adapted from Hoffman JI, Kaplan S: The incidence of congenital heart disease, *J Am Coll Cardiol* 2002;39(12):1890.

In addition to being classified according to pathologic features as obstructions or shunts, heart defects are classified according to the clinical manifestation of cyanosis. Acyanotic disorders include the obstructive disorders and left-to-right shunts. The cyanotic category includes abnormalities causing right-to-left shunts. Specific heart defects are described further and follow this categorization.

Acyanotic Congenital Defects

Atrial Septal Defect

During the third to fifth week of fetal development, the left and right atria are separated by flaps of tissue that become the atrial septum. The foramen ovale remains patent during intrauterine life such that blood may pass from the right to the left atrium and bypass the uninflated and nonfunctional lungs. The foramen ovale normally remains open in utero because pressure on the right side of the heart is higher than that on the left. With birth, however, the pressure gradient reverses as the lungs inflate and greatly reduce pulmonary vascular resistance. The higher left-sided pressure forces the flap shut, and fusion of the foramen ovale membrane normally occurs. The majority of atrial septal defects occur at the location of the foramen ovale. The abnormal septal opening may be of variable size. Small defects (1 cm) are well tolerated. Even larger atrial septal defects may be asymptomatic for many years as long as the shunt flow is left to right and therefore acyanotic (Fig. 18.26).

The long-term increase in pulmonary blood flow may eventually lead to pulmonary hypertension, right ventricular hypertrophy, and a reversal of the shunt to a right-to-left pattern. Cyanosis, respiratory

BOX 18.5 Features of Heart Failure in Infants

Poor feeding and failure to thrive
Respiratory distress (mainly tachypnea)
Rapid heart rate (160–180 beats/min)
Pulmonary rales or wheezing
Cardiomegaly and pulmonary edema on radiogram
Hepatomegaly (peripheral edema unusual)
Gallop sounds
Color (ashen pale or faintly cyanotic)
Excessive perspiration
Diminished urine output

TABLE 18.9 Selected Examples of Gene Defects Associated With Congenital Heart Disease

Disorder	Genes	Gene Product Function
Nonsyndromic:		
ASD or conduction defects	NKX2.5	Transcription factor
ASD or VSD	GATA4	Transcription factor
Tetrology of Fallot	ZFPM2 or NKX2.5	Transcription factor
Syndromic:		
Alagille syndrome – pulmonary artery stenosis or tetralogy of Fallot	JAG1 or NOTCH2	Signaling proteins or receptors
Char syndrome – PDA	TFAP2B	Transcription factor
CHARGE syndrome – ASD, VSD, PDA, or hypoplastic right side of the heart	CHD7	Helicase-binding protein
DiGorge syndrome – ASD, VSD, or outflow tract obstruction	TBX1	Transcription factor
Holt-Oram syndrome – ASD, VSD, or conduction defect	TBX5	Transcription factor
Noonan syndrome – pulmonary valve stenosis, VSD, or hypertrophic cardiomyopathy	PTPN11, KRAS, SOS1	Signaling proteins

ASD, Atrial septal defect; *PDA,* patent ductus arteriosus; *VSD,* ventricular septal defect;
From Kumar V et al: *Robbins and Cotran pathologic basis of disease,* ed 9, Philadelphia, 2015, Saunders, p 532.

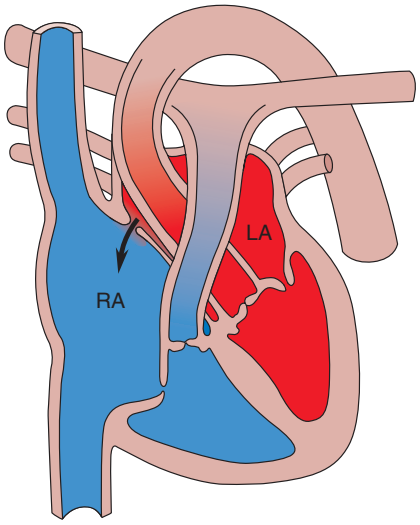


FIG 18.26 Atrial septal defect. Blood flow through the defect is usually left to right and produces an acyanotic shunt.

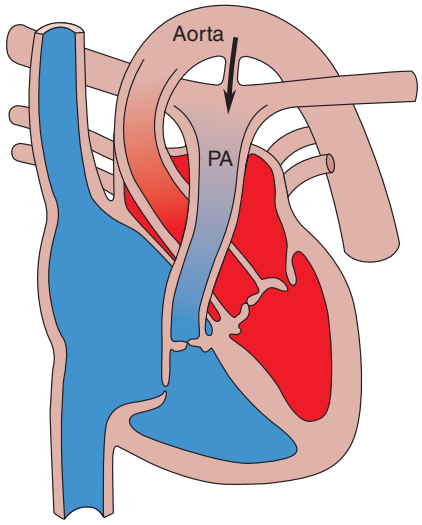


FIG 18.28 Patent ductus arteriosus. Blood flow through the ductus is usually from the aorta to the pulmonary artery and produces an acyanotic shunt.

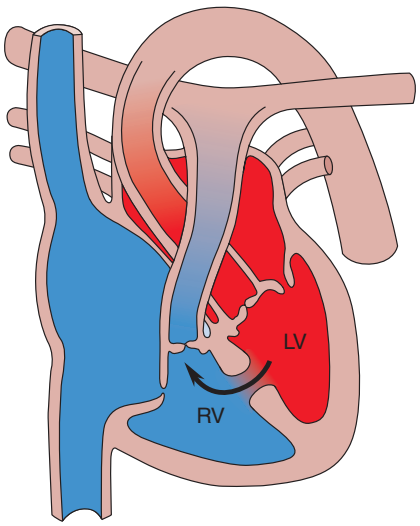


FIG 18.27 Ventricular septal defect. Blood flow through the defect is usually left to right and produces an acyanotic shunt.

difficulty, and right-sided heart failure may ensue. Large or symptomatic atrial septal defects are commonly repaired surgically early in life, before pulmonary complications occur.

Ventricular Septal Defect

A ventricular septal defect is the most common congenital cardiac anomaly. It is frequently associated with other cardiac defects such as tetralogy of Fallot, transposition of the great arteries, and atrial septal defects. The ventricular septum develops between the fifth and sixth weeks of fetal life as the membrane derived from the endocardial cushion fuses with the muscular septum.

The majority of ventricular septal defects are located in the membranous septum, very close to the bundle of His. As with atrial septal defects, the functional significance depends largely on the size of the defect. The shunt is initially left to right because left-sided heart pressures are higher (Fig. 18.27). With the increase in pulmonary blood flow, pulmonary hypertension and right ventricular hypertrophy may result and cause a reversal of the shunt.

Large ventricular septal defects may be apparent at birth because of rapidly developing right-sided heart failure and a loud systolic murmur. Large, symptomatic defects in infants or moderate defects in older children are repaired surgically to avoid progression to pulmonary vascular disease. Small ventricular septal defects in infants are generally not immediately repaired because of the tendency of such defects to close spontaneously.

Patent Ductus Arteriosus

The ductus arteriosus is a normal channel between the pulmonary artery and the aorta that remains open during intrauterine life (Fig. 18.28). Within 1 to 2 days after birth, the ductus arteriosus closes functionally, and within a few weeks it closes permanently. The ductus arteriosus allows blood to flow from the pulmonary artery into the aorta, thus bypassing the lungs. Low oxygen tension and local production of prostaglandins appear to be important in maintaining patency of the channel during fetal life. After birth, flow through the ductus arteriosus switches to left to right because of the higher pressure in the aorta. This change in flow direction brings oxygenated blood through the ductus arteriosus and stimulates it to close. A reduction in prostaglandin E production after birth appears to contribute to constriction and closure. In many cases the reason for abnormal continued patency of the ductus arteriosus after birth is not well understood. Conditions that cause low blood oxygen tension may contribute to continued patency.

Most often a patent ductus arteriosus has no clinical significance early in life because the shunt is left to right and no cyanosis is evident. Surgical management is usually delayed because these defects tend to close spontaneously. Prostaglandin inhibitors may be given to induce closure of the defect. Continued patency of the ductus arteriosus is usually obvious because of a harsh, grinding systolic murmur and often a systolic thrill (vibration). Surgical closure of the patent ductus arteriosus is done as soon as it becomes evident that spontaneous closure is unlikely. As with other left-to-right shunt disorders, uncorrected patent ductus arteriosus results in pulmonary hypertension complicated by respiratory and right-sided heart failure. Eventual reversal of the shunt to a right-to-left pattern results in cyanosis. Because the ductus is usually located

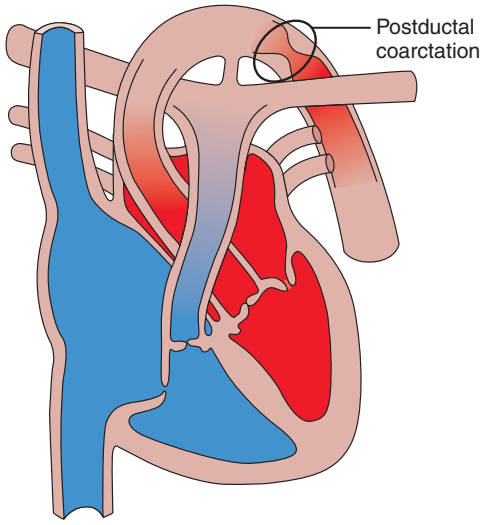


FIG 18.29 Coarctation of the aorta. The arterial narrowing can produce a weaker pulse in the lower extremities.

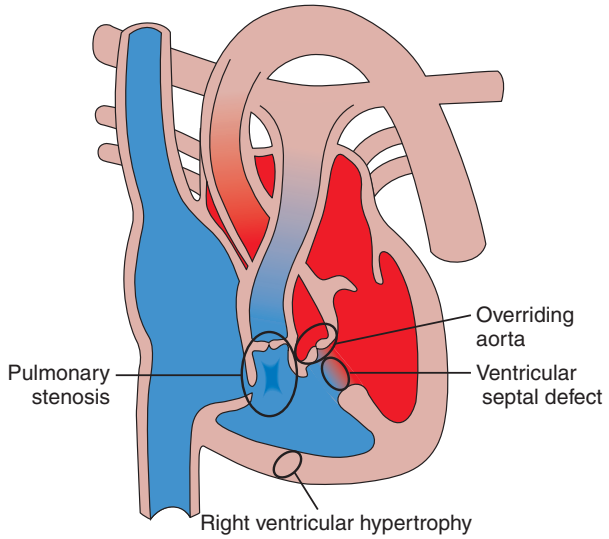


FIG 18.30 Tetralogy of Fallot showing the four characteristic abnormalities: pulmonary stenosis, ventricular septal defect, overriding aorta, and right ventricular hypertrophy. Tetralogy of Fallot is a cyanotic defect.

distal to the origin of the subclavian artery, the lower extremities may show cyanosis, whereas the upper extremities remain pink.

Coarctation of the Aorta

Coarctation refers to a narrowing or stricture that may impede blood flow. Coarctation of the aorta is a common heart defect that affects males three to four times more frequently than females. Narrowing of the aorta may occur anywhere along its length; however, in most cases the coarctation is located just before or just after the ductus arteriosus (Fig. 18.29). Preductal coarctation (proximal to the ductus arteriosus) is usually more severe and often associated with other anomalies. In some instances, the aortic stricture is so severe that blood flow to the lower part of the body must be maintained solely by flow through the ductus arteriosus. This situation results in a very high workload for the right side of the heart and may lead to heart failure in the early neonatal period. Blood supply to the arms and head is unaffected because these arteries arise proximal to the stricture. Postductal coarctation is generally less severe and may remain unrecognized until adulthood.

The upper extremities typically have an elevated blood pressure, whereas the lower extremities have weak pulses and low blood pressure. An important part of assessment of the newborn is comparison of pulses in the upper and lower extremities to assess for symmetry. All types of coarctation are usually accompanied by systolic murmurs and ventricular hypertrophy. The stricture can be repaired surgically by resection of the narrowed region. If left untreated, significant coarctation may lead to congestive heart failure, intracranial hemorrhage, or aortic rupture.

Pulmonary Stenosis or Atresia

Isolated pulmonary stenosis and atresia are included in the category of acyanotic defects because they do not themselves result in cyanosis. However, they often occur in conjunction with other anomalies that allow survival into the neonatal period. The other defects may allow shunting of blood and result in cyanosis. In pulmonary atresia, no communication is found between the right ventricle and the lungs so that blood must enter the lungs by first traveling through a septal opening and then through a patent ductus arteriosus. The right ventricle is typically underdeveloped (hypoplasia), and the atrial septal defect is large. Pulmonary stenosis can be mild to severe, depending on the extent of narrowing of the pulmonic valve. Pulmonary stenosis is usually due to abnormal fusion of the valvular cusps. Right ventricular hypertrophy occurs secondary to the high ventricular afterload caused by the narrowed outflow opening. Isolated pulmonary stenosis is easily corrected by surgery; however, the prognosis depends in large part on the health of the right ventricle.

Aortic Stenosis or Atresia

Congenital aortic atresia is rare and not compatible with survival. However, depending on its severity, aortic stenosis is correctable and associated with a good prognosis. Aortic stenosis may involve the valvular cusps or the subvalvular fibrous ring just below the cusps. The narrowed aortic outflow tract results in a high left ventricular afterload, which causes the left ventricle to enlarge. A prominent systolic murmur is usually apparent. Surgical replacement is the definitive treatment if the stenosis is severe, progresses, or becomes symptomatic.

Cyanotic Congenital Defects

Tetralogy of Fallot

The four defining features of tetralogy of Fallot are (1) a ventricular septal defect, (2) an aorta positioned above the ventricular septal opening (overriding aorta), (3) pulmonary stenosis that obstructs right ventricular outflow, and (4) right ventricular hypertrophy (Fig. 18.30).

The severity of the symptoms is related primarily to the degree of pulmonary stenosis. The heart is generally enlarged because of the extensive right ventricular hypertrophy. Even if the condition is untreated, individuals with tetralogy of Fallot may live into adulthood. The defect often results in cyanosis because the overriding aorta receives unoxygenated blood from the right side of the heart as well as oxygenated blood from the left side. The degree of cyanosis depends on the amount of blood received from the right side, which in turn depends on the degree of pulmonic obstruction. Surgical correction of the defects is usually recommended because prolonged noncorrective management carries the risk of infective endocarditis and secondary polycythemia.

Transposition of the Great Arteries

In the most common form of transposition of the great arteries, the aorta arises from the right ventricle and the pulmonary artery arises from the left ventricle (Fig. 18.31). This anomaly results in the formation

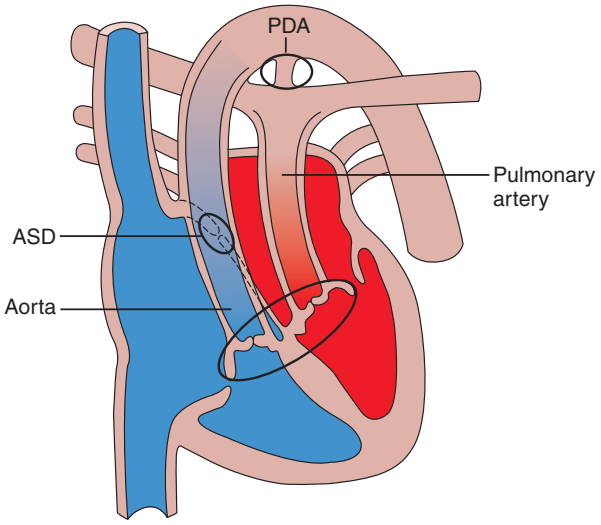


FIG 18.31 Transposition of the great arteries. Two separate circulations are formed, which is incompatible with life unless mixing of blood occurs through other defects. *ASD,* Atrial septal defect; *PDA,* patent ductus arteriosus.

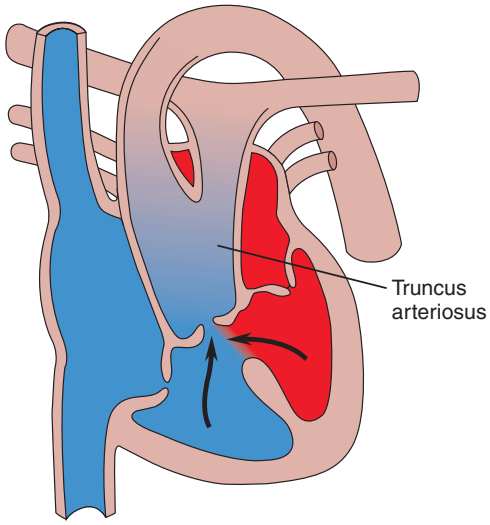


FIG 18.32 Truncus arteriosus is a cyanotic defect. Failure of septation results in a common outflow tract from the ventricles. A ventricular septal defect is also present.

of two separate, noncommunicating circulations. The right side of the heart receives blood from the systemic circulation and recirculates it through the body by way of the aorta. Blood reaching the body has not passed through the lungs and is therefore not oxygenated. The left side of the heart receives oxygenated blood from the lungs and then recirculates it through the lungs by way of the pulmonary artery. Unless some mixing of these separate circulations takes place through other heart defects, such as septal defects, transposition is not compatible with life.

Nearly all infants who survive the neonatal period have an interatrial opening, and most also have a patent ductus arteriosus. A good deal of mixing must be maintained after birth for the infant to survive. Surgery may be directed at improving the mixing of systemic and pulmonary blood by enlarging or creating openings in the heart. Corrective surgery in which the aorta and pulmonary arteries are excised from the heart and sutured to the opposite ventricular outflow tract is the treatment of choice. The coronary arteries must also be reimplanted into the new left ventricular outflow tract in this procedure.

Truncus Arteriosus

Truncus arteriosus is a congenital malformation in which failure of the pulmonary artery and aorta to separate results in formation of one large vessel that receives blood from both the right and left ventricles (Fig. 18.32). A large ventricular septal defect and a single valvular structure are present and lead to the single large artery. Mixing of blood from the right and left sides of the heart results in systemic cyanosis. The amount of blood entering the systemic versus the pulmonary circulation depends on the degree of vascular resistance in the two systems. Abnormally high pulmonary blood flow may progress to pulmonary hypertension and right ventricular hypertrophy. Increased pulmonary resistance causes the cyanosis to become more severe as more venous blood enters the systemic circulation. Surgical correction is required for survival.

Tricuspid Atresia

Absence of the tricuspid valve is almost always associated with underdevelopment of the right ventricle and an atrial septal defect. Circulation is maintained by the defect, which allows blood to bypass the right ventricle. A patent ductus arteriosus is required to perfuse the lungs. In some cases, a concomitant ventricular septal defect is present and may allow some blood to pass into the right ventricle and enter the pulmonary circulation. Cyanosis is present from birth, and the mortality rate is high. Surgical correction is required for survival.

KEY POINTS

- Different congenital heart anomalies result in two primary pathologic processes: (1) shunting of blood through abnormal pathways in the heart or great vessels and (2) obstruction to blood flow because of abnormal narrowing.
- Disorders that result in left-to-right shunting of blood or obstruction to flow are generally acyanotic. These disorders include atrial septal defect, ventricular septal defect, patent ductus arteriosus, coarctation of the aorta, and pulmonary and aortic stenosis or atresia.
- Disorders that result in right-to-left shunting of blood result in cyanosis. These disorders include tetralogy of Fallot, transposition of the great arteries, truncus arteriosus, and tricuspid atresia.

SUMMARY

A variety of disease processes may interfere with the heart's ability to provide the body with oxygenated blood. Among these processes are CHD, valvular and endocardial diseases, myocardial diseases, pericardial diseases, and congenital heart defects. CHD includes several clinical forms: stable angina pectoris, ACS (unstable angina, MI), chronic ischemic heart disease, and sudden cardiac arrest. Stenotic coronary lesions obstruct blood flow to the myocardium and result in these ischemic clinical syndromes. Distinction between unstable angina and MI relies on the presence of serum biomarkers of myocardial damage. Unstable angina is ischemia without cellular death and therefore does not result

in biomarker elevation. MI, on the other hand, is associated with the death of myocardial cells and subsequent release of intracellular components. MI may be complicated by dysrhythmias and failure of the heart to pump efficiently.

Valvular disorders are of two types: those that impede flow because of stenosis and those that allow regurgitation because of failure to close completely. The general consequence of valvular disorders is increased myocardial workload secondary to high afterload (stenosis) or high preload (regurgitation). The heart may eventually decompensate and proceed to heart failure. The endocardial diseases, rheumatic heart disease and infective endocarditis, also primarily affect the heart valves and create stenosis and regurgitation.

Disorders of the myocardium include myocarditis, which is an inflammatory process, and cardiomyopathy, which is a noninflammatory process, usually of genetic cause. Most cases of myocarditis are viral; however, it is the immune system's response to the virus that appears to cause myocardial damage. Myocarditis results in a dilated, flabby heart with decreased pumping efficiency. The cardiomyopathies are a diverse group of disorders that may be classified as primary (having unknown cause) and specific (caused by a known disease process). Cardiomyopathies include a dilated form, a hypertrophic form, and a restrictive form. The primary problem in the dilated form is poor contractility of all heart chambers. The hypertrophic form may cause left ventricular outflow obstruction that interferes with cardiac output and increases left ventricular strain. Dysfunction in the restrictive form is caused by poor diastolic filling as a result of a stiff, fibrosed ventricular chamber.

Pericardial disorders include accumulations of fluid in the pericardial sac and acute and chronic forms of pericarditis. Pericardial fluid may be serous, serosanguineous, infectious, or frank blood. Pericardial accumulations are usually of little consequence except as indicators of underlying pathophysiologic processes. However, if the accumulation is large or rapid, it may compress the heart and interfere with diastolic filling—a process called *cardiac tamponade.* Pericarditis refers to inflammation of the pericardium. It is usually secondary to other disease processes. Pericardial inflammation generally causes pain and may be associated with a friction rub. Chronic pericarditis can cause erosion of the pericardial sac such that the epicardial layer of the heart may become fused to other mediastinal structures. Alternatively, chronic pericarditis may cause the pericardial sac to become fibrotic and noncompliant such that it restricts expansion of the heart during diastolic filling.

A number of heart disorders may be present at birth and can be categorized as obstructions or shunts and as cyanotic or acyanotic. In general, disorders that allow unoxygenated blood from the right heart to enter the systemic circulation (right-to-left shunt) cause cyanosis. Examples of cyanotic defects include tetralogy of Fallot, transposition of the great arteries, truncus arteriosus, and tricuspid atresia. Examples of acyanotic defects are coarctation of the aorta, atrial and ventricular septal defects, and patent ductus arteriosus.

All heart diseases discussed in this chapter may be complicated by heart failure. Heart failure occurs when the pumping efficiency of the heart is decreased such that cardiac output is subnormal. It is often accompanied by congestion of the lungs or the systemic venous system. Heart failure is discussed in Chapter 19.

RESOURCES

American Heart Association: *Heart disease and stroke statistics—2017 update*, Dallas, TX, 2017, The Association.

Arbustini E, et al: The MOGE(S) classification for a phenotype–genotype nomenclature of cardiomyopathy: endorsed by the world heart federation. *J Am Coll Cardiol* 62:2046–2072, 2013.

Kong MH, Fonorow GC, Peterson ED, et al: Systematic review of the incidence of sudden cardiac death in the United States. *J Am Coll Cardiol* 57:794–901, 2011.

Kumar V, Abbas A, Aster J, editors: *Robbins and Cotran pathologic basis of disease*, ed 9, Philadelphia, 2015, Saunders.

Mann D, Zipes D, Libby P, Bonow R: *Braunwald's heart disease: a textbook of cardiovascular medicine*, ed 10, Philadelphia, 2015, Elsevier.

US Preventive Services Task Force: Statin use for the primary prevention of cardiovascular disease in adults: US preventive services task force recommendation statement. *JAMA* 316(19):1997–2007, 2016.

Yusuf S, Bosch J, Dagenais G, et al: HOPE-3 investigators. Cholesterol lowering in intermediate-risk persons without cardiovascular disease. *N Eng J Med* 374(21):2021–2031, 2016.

19

Heart Failure and Dysrhythmias: Common Sequelae of Cardiac Diseases

Benjamin J. Miller and Jacquelyn L. Banasik

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- What are the common predisposing factors for development of heart failure?
- How does heart failure with systolic dysfunction differ from heart failure with preserved systolic function?
- How do the compensatory responses triggered in heart failure work to restore cardiac output, and how might they lead to remodeling and progression?
- What are the clinical manifestations of heart failure?
- How are preload, afterload, and contractility managed therapeutically in the patient with heart failure?
- What are the characteristic electrocardiographic features of the common cardiac dysrhythmias?
- What is the clinical significance and usual treatment of each of the common cardiac dysrhythmias?

CHAPTER OUTLINE

Heart Failure, 412
Pathogenesis and Diagnosis, 412
Systolic Dysfunction with Low Ejection Fraction, 412
Diastolic Dysfunction with Preserved Ejection Fraction, 412
Compensatory Mechanisms, Remodeling, and Progression, 413
Sympathetic Nervous System Activation, 413
Increased Preload, 413
Myocardial Remodeling and Progression, 415
Clinical Manifestations, 417
Left-Sided Heart Failure, 417
Right-Sided Heart Failure, 418
Biventricular Heart Failure, 420
Class and Stage of Heart Failure, 420
Treatment, 421
Cardiac Dysrhythmias, 421
Dysrhythmia Mechanisms, 421
Automaticity, 421
Triggered Activity, 421
Reentry, 422
Dysrhythmia Analysis, 423
Normal Sinus Rhythm, 423
Abnormal Rates of Sinus Rhythm, 424
Sinus Tachycardia, 424
Sinus Bradycardia, 424
Sinus Arrhythmia, 424
Sinus Arrest, 424
Abnormal Site of Impulse Initiation, 425
Escape Rhythms, 425
Atrial Dysrhythmias, 425
Junctional Dysrhythmias, 426
Ventricular Dysrhythmias, 427
Conduction Pathway Disturbances, 428
Disturbances of Atrioventricular Conduction, 428
Abnormal Conduction Pathways, 429
Intraventricular Conduction Defects, 429
Treatment, 431

Heart failure (HF) and cardiac dysrhythmias (arrhythmias) may occur in association with cardiac diseases from a number of different causes. *Heart failure* refers to the inability of the heart to maintain sufficient cardiac output to optimally meet metabolic demands of tissues and organs, and is the end stage of many cardiac diseases. HF involves multiple organ systems and is a progressive syndrome. If the contracting and relaxing abilities of the heart are impaired, then blood flow to the systemic circulation will be reduced, and congestion of blood can occur in the pulmonary venous circulation. In patients with HF, these symptoms of fluid overload sometimes are described as *congestive heart failure* (CHF). This chapter includes the chronic forms of HF; acute HF is discussed in Chapter 20 because it commonly results in cardiogenic shock. Disturbances in electrical activity of the heart may signify underlying pathophysiologic processes and may also lead to insufficient cardiac output. Neither HF nor dysrhythmia is a primary cardiac disease; therefore underlying pathophysiologic processes must be investigated.

HF is a significant cardiac disorder affecting about 5.1 million Americans. More than 670,000 new cases are diagnosed in the United States each year, with an incidence of 10 per 1000 population after age 65. HF is a common reason for hospitalization in patients older than 65 years. The increasing incidence and hospitalization rates of HF reflect aging of the U.S. population, as well as better treatment and an improved survival rate after myocardial infarction (MI).

HEART FAILURE

Pathogenesis and Diagnosis

A large number of cardiac disorders, including most of those discussed in Chapter 18, can lead to the development of HF. Coronary heart disease (CHD) and hypertension (HTN) are associated with the majority of HF cases. CHD contributes to HF progression through mechanisms that include endothelial dysfunction, ischemia, and infarction. CHD and HTN interact to increase the risk of HF. Less common causes of HF include dilated cardiomyopathy, congenital heart defects, valvular disorders, respiratory diseases, anemia, and hyperthyroidism. The diagnosis of HF is based on the presence of a constellation of signs and symptoms that are characteristic of the syndrome. However, different sets of criteria are in use, including the Framingham Criteria and Minnesota Heart Failure Criteria. Commonly used criteria for identifying HF include the presence of dyspnea, pulmonary rales, cardiomegaly, pulmonary edema, S_3 heart sound, and tachycardia, although many other criteria may be applied. No single diagnostic test is available for HF. The diagnosis should be based on a thorough medical history and physical examination.

Chronic HF is classified in various ways. It may be categorized as right sided, left sided or biventricular according to the ventricles that are failing. HF also is classified by the ability of the heart to eject blood from the ventricles. Those with a low ejection fraction (EF) have significant systolic dysfunction, called *HF with low EF (HFlEF)*. EF is calculated by dividing stroke volume by end-diastolic volume. A normal EF is 60% to 80%. Patients with systolic failure have characteristically low EFs (<35%). Patients with EF greater than 50% do not have significant systolic dysfunction and are categorized as *HF with preserved EF (HFpEF)*. Patients diagnosed with HF who have an EF greater than 50% generally have HF symptoms because of poor diastolic relaxation. In general, patients with low EF but no congestive symptoms have a survival rate about the same as those with preserved EF who have congestive symptoms. The highest mortality occurs in patients with both low EF and congestive symptoms. The overall mortality for HF is high, with about 50% of patients dying within 5 years of diagnosis.

Systolic Dysfunction With Low Ejection Fraction

Patients with systolic dysfunction have reduced myocardial contractility evidenced by a low EF and a reduced d*P*/d*t* during ventricular systole. The d*P*/d*t* is a measure of inotropy—how quickly the ventricle can develop a forceful contraction. The nature of the impaired contractility is only partially understood; however, myocyte loss, mechanical derangements of myocardial cells, and dysregulation of neurohormones are believed to be critical elements.

Impaired contractility attributable to MI is a common cause of HF. MI, with cell death and loss of contractile elements, reduces the heart's contractile force. The degree of pump failure is related to the amount of heart muscle lost. In patients with HF, myocardial cells are also subject to high rates of apoptosis or programmed cell death. Apoptosis can be triggered by excessive stimulation by certain neurohormones and by ischemia. Over time, the loss of myocardial cells contributes to reduced contractility. In severe systolic HF, the EF may fall below 15% or 20%. In general, the prognosis worsens as EF decreases.

Inadequate supplies of oxygen to the contracting cells may impair contractility because each myosin cross-bridge cycle requires a molecule of adenosine triphosphate (ATP). When ATP production is low, fewer cross-bridge cycles are completed with each contraction, which results in a reduced EF. Ischemia also impairs diastolic relaxation because removal of calcium ions from the cytoplasm is an energy-requiring process that may be slowed with insufficient production of ATP (see Chapter 17).

β_1-Receptor down-regulation is thought to be an important mechanism of impaired systolic function. Chronic overexcitation of cardiac β_1 receptors by sympathetic neurotransmitters (e.g., norepinephrine [NE]) leads to a reduction in β_1 receptor function and results in a myocardium that is less responsive to sympathetic stimulation and adrenergic drug therapy. β_1-receptor–blocking agents have been shown to improve EF and to reduce mortality, lending support to the view that chronic excessive sympathetic nervous system (SNS) activation is detrimental to cardiac function.

Diastolic Dysfunction With Preserved Ejection Fraction

CHD and HTN are associated with development of diastolic dysfunction, just as they are with development of systolic failure. Why the same disease risk factors result in different cardiac dynamics in different individuals is not completely known. HF with preserved EF is more common in women, the elderly, and those with no history of MI. Diastolic failure is a disorder of myocardial relaxation. In this condition, the left ventricle is excessively noncompliant and does not fill effectively. Two separate functional processes normally occur during the diastolic relaxation phase: The first is an energy-requiring process (lusitropy) that removes free calcium ions from the cytoplasm by pumping them back into the sarcoplasmic reticulum and across the cell membrane into the extracellular fluid. Removal of calcium ions inhibits cross-bridge formation and allows the thick and thin filaments of the sarcomere to passively slide apart. Ischemia, with subsequent ATP deficiency, interferes with the efficiency of calcium ion removal and can impair the active phase of diastolic relaxation.

The second process is passive stretch of the ventricular myocardium to accommodate filling. Passive compliance of the ventricle can be decreased by deposition of fibrin and collagen during scar formation or by hypertrophic thickening of the ventricular wall. Both active and passive processes may be impaired together and are difficult to distinguish clinically.

The hallmark of HF with preserved EF is that the patient exhibits clinical manifestations of HF, including pulmonary congestion, and edema formation, but has a normal or near-normal EF (usually defined as greater than 50%), indicating absence of significant systolic impairment. Because prognosis and treatment recommendations may differ, an echocardiogram to measure EF is recommended in all patients with HF. A comparison of the left ventricular pressure–volume loops in systolic and diastolic dysfunction is shown in Fig. 19.1. Systolic failure is characterized by higher-than-normal diastolic volume and low EF, whereas the pressure–volume loop in diastolic failure indicates poor compliance with a lower diastolic volume at a higher-than-normal pressure. When the EF is between 35% and 50%, the patient cannot be specifically assigned into the low EF or preserved EF categories.

KEY POINTS

- Heart failure (HF) is the end stage of many cardiac disorders. It occurs when the heart is unable to provide sufficient cardiac output to meet normal metabolic functions of the body.
- The risk factors for HF include myocardial ischemia from coronary artery disease, hypertension (HTN), and cardiomyopathies.

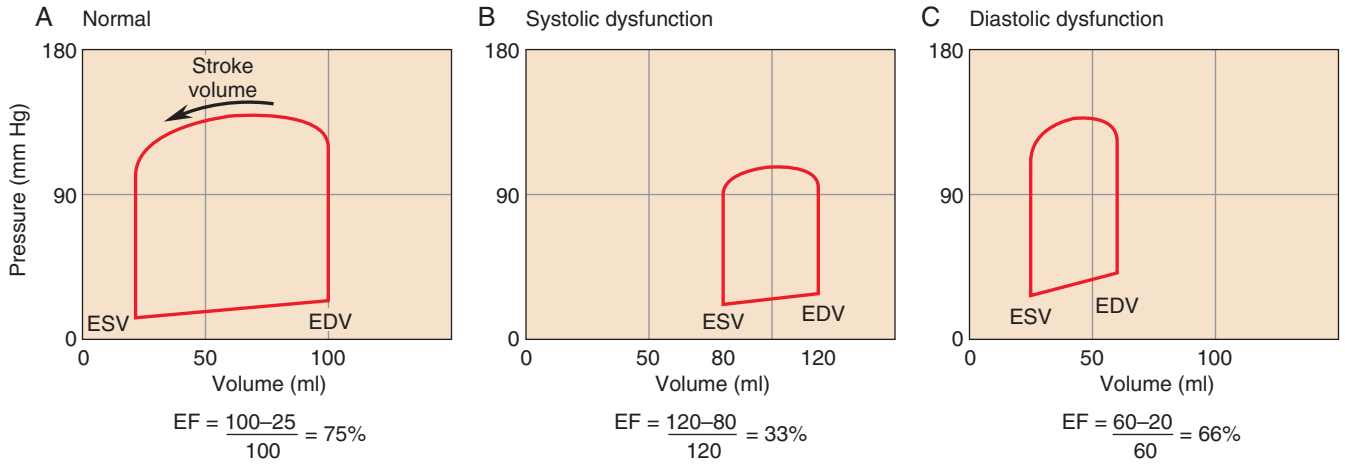


FIG 19.1 Comparison of the left ventricular pressure–volume loop in a **(A)** normal heart, **(B)** systolic dysfunction, and **(C)** diastolic dysfunction. Note that end-diastolic pressure is higher than normal in both systolic and diastolic failure, but end-diastolic volume is lower in diastolic dysfunction. *EDV,* End-diastolic volume; *EF,* ejection fraction; *ESV,* end-systolic volume.

- Impaired contractility resulting in systolic failure is frequently associated with HF symptoms. The biochemical basis of impaired contractility involves loss of cardiac muscle cells, β_1-receptor down-regulation, reduced adenosine triphosphate (ATP) production, and altered calcium ion regulation.
- In about half of HF patients, systolic function is preserved and diastolic dysfunction predominates. HF with preserved ejection fraction (EF) is particularly likely to develop in the elderly, in women, and in those without a history of myocardial infarction (MI).
- Left ventricular pressure–volume loops characterize the differences in systolic and diastolic dysfunction. High diastolic volume and reduced EF indicate systolic failure, whereas diastolic failure is characterized by higher diastolic pressures at lower volumes.

Compensatory Mechanisms, Remodeling, and Progression

When the heart fails to provide adequate cardiac output to meet tissue demands, a number of compensatory mechanisms are triggered. In the short term, these mechanisms are helpful in restoring cardiac output toward normal levels, but in the long term, they may be detrimental to cardiac structure and function. Much of the current management of HF is aimed at attenuating the harmful consequences of these compensatory responses. Three main compensatory mechanisms are activated in HF: SNS activation, increased preload, and myocardial hypertrophy (Fig. 19.2).

Sympathetic Nervous System Activation

Sympathetic activation of the heart is partly a result of baroreceptor reflex stimulation. The baroreceptors (pressoreceptors), located in the aorta and carotid arteries, detect a fall in pressure because of diminished stroke volume and transmit this information to the central nervous system (CNS). The CNS increases activity in the sympathetic nerves to the heart, resulting in increased heart rate and contractility. However, because of impaired contractile ability, the failing heart may have reduced responsiveness to sympathetic activation. Sympathetic activation also causes venoconstriction, which redistributes blood and increases cardiac preload. Sympathetic constriction of arterioles helps maintain blood pressure when cardiac output is reduced. Specialized cells in the kidney called *juxtaglomerular cells* also receive SNS stimulation when cardiac output falls. The juxtaglomerular cells release renin and initiate the renin–angiotensin–aldosterone cascade, leading to salt and water retention by the kidney. Sympathetic activation is an early and immediate compensatory response to insufficient cardiac output.

Sympathetic activation is an effective means for increasing cardiac output in an acute process, such as volume depletion. However, in HF, sympathetic activation becomes a chronic process that is ultimately deleterious. A major problem with excessive sympathetic activation is that afterload on the left ventricle can be increased significantly. A high afterload increases cardiac workload and may decrease stroke volume. Therefore treatment of high blood pressure, if present, is important for improving cardiac function in patients with HF.

Drugs that block β_1 receptors have been advocated in the management of HF to inhibit the cardiac effects of sympathetic activation. Many clinicians had been reluctant to use β-blockers in patients with HF because these drugs are negative inotropes and have the potential to reduce cardiac output. In HF, where cardiac output is already low, the use of a negative inotrope would seem to be contraindicated. However, several randomized clinical trials have reported an improved mortality rate in patients receiving certain β_1-blockers, and they are now recommended as standard therapy in most HF guidelines. Long-term SNS stimulation of the heart may contribute to HF progression and remodeling of the cardiac tissue. Remodeling is a process of myocyte loss, hypertrophy of remaining cells, and interstitial fibrosis (Fig. 19.3). The remodeled tissue is less functional and may predispose to worsening failure and cardiac dysrhythmias.

Increased Preload

Increased preload in the cardiac chambers is initially a consequence of reduced EF with a resultant increase in residual end-systolic volume. Subsequently, decreased cardiac output to the kidney reduces glomerular filtration, resulting in fluid conservation. In addition, the renin–angiotensin–aldosterone system (RAAS) is activated because of reduced blood flow to the kidney and SNS activation of the juxtaglomerular cells. Angiotensin II (AII) and aldosterone enhance sodium and water reabsorption by the kidney, contributing to an elevated blood volume. Increased preload is a compensatory mechanism that enhances the ability of the myocardium to contract forcefully. An enlarged chamber volume causes the myocardial fibers to lengthen during diastole, which results in greater fiber shortening during contraction (Frank–Starling mechanism). The diastolic length of the muscle fibers is thought to determine the number of effective cross-bridge cycles that can be

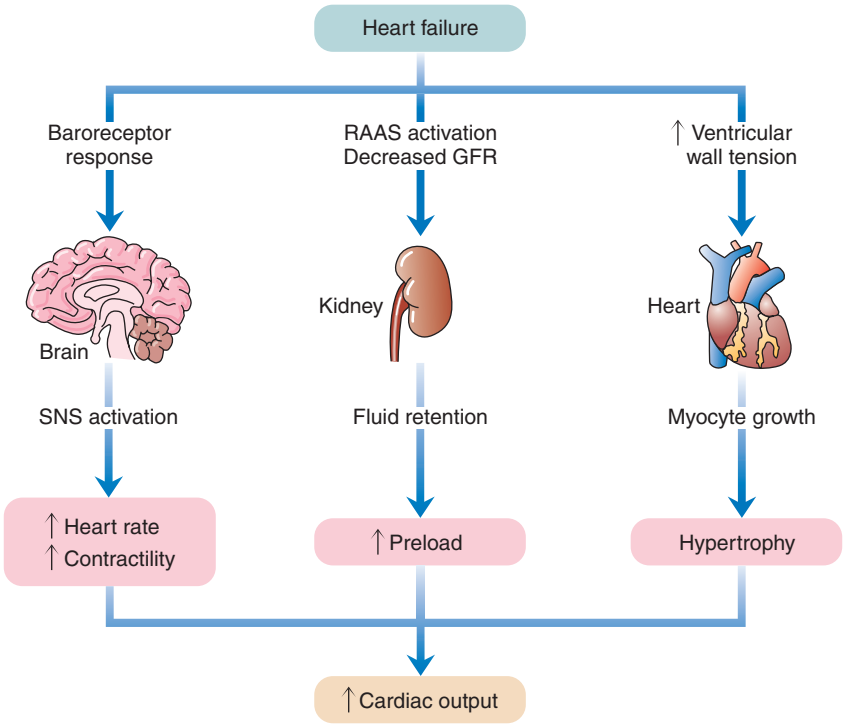


FIG 19.2 Major compensatory mechanisms in heart failure that act to restore cardiac output. *GFR,* Glomerular filtration rate; *RAAS,* renin–angiotensin–aldosterone system; *SNS,* sympathetic nervous system.

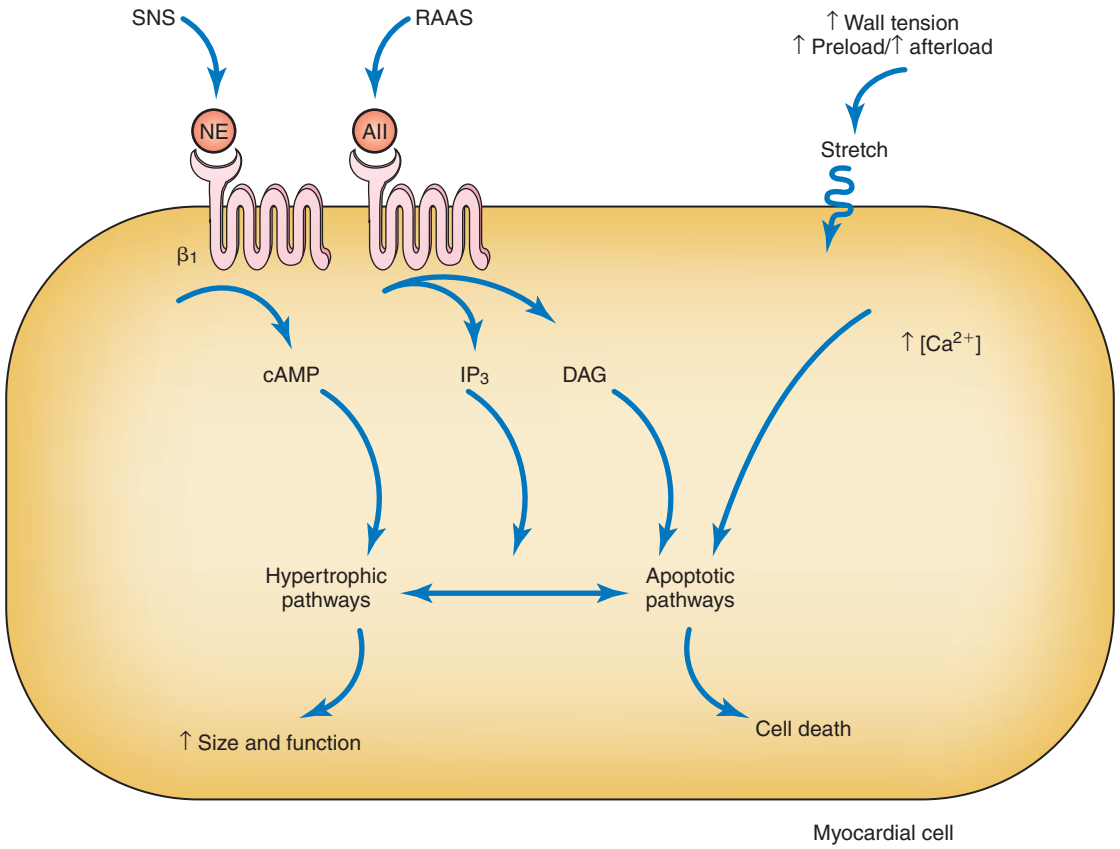


FIG 19.3 Mechanisms of ventricular remodeling in heart failure. Activation of β1 receptors and AII receptors along with stretch of the cell membrane trigger signaling cascades. Under some conditions these triggers lead to effective hypertrophy and an increase in size and function, and in others they trigger apoptotic cell death. *AII,* Angiotensin II; *cAMP,* cyclic adenosine monophosphate; *DAG,* diacylglycerol; *IP3,* 1,4,5-inositol trisphosphate; *NE,* norepinephrine; *RAAS,* renin–angiotensin–aldosterone system; *SNS,* sympathetic nervous system.

accomplished during systole (see Chapter 17). Thus up to a point, an increase in the volume or preload of the heart will result in a greater force of contraction (Fig. 19.4). The cardiac function curve flattens out at a certain point, and minimal benefit is obtained despite increasing preload. Patients with systolic failure have a cardiac function curve that is flat and shifted to the right of normal. Thus they require a higher preload to achieve a given stroke volume. However, patients with HF often retain so much volume that their hearts are functioning on the flat part of the curve. These patients benefit from preload reduction, which will decrease systemic and pulmonary congestive symptoms and cardiac workload with little or no reduction in cardiac output. Diuretics are commonly used to achieve moderate preload reduction.

Myocardial Remodeling and Progression

Hypertrophy of cardiac muscle cells is the third mechanism of compensation and generally takes much longer to occur than preload enhancement or sympathetic activation. Hypertrophy appears to result, in part, from a chronic elevation of myocardial wall tension. Wall tension may be high as a result of increased diastolic blood volume (high preload) or as a consequence of high systolic pressures generated in the chamber to overcome high afterload. The relationship between myocardial wall tension and intrachamber pressure and diameter is described by the law of Laplace:

$$\text{tension} = (\text{transmural pressure} \times \text{radius})/\text{wall thickness}$$

When the ventricular chamber enlarges and pressures increase, more tension is created in the ventricular muscle wall (Fig. 19.5). The development of high systolic pressures in the ventricle may be necessary to overcome a high afterload, such as occurs with arterial HTN and aortic valve stenosis. The hypertrophy of contractile elements in the myocardium increases the heart's pumping force and helps reduce the wall tension of the heart toward normal levels. In general, an increase in chamber diameter because of excessive preload is thought to contribute to eccentric hypertrophy in which the muscle fibers elongate. High afterload results in concentric hypertrophy in which the muscle fibers grow in diameter and thicken the ventricular wall (Fig. 19.6).

Neurohormones, which include NE and angiotensin (AII), also have hypertrophic effects on the heart. Circulating AII levels are higher than normal in HF because of poor kidney perfusion, which triggers production of AII through the RAAS. In HF, AII is also produced locally in the heart. AII binds to the angiotensin type 1 (AT1) receptor on cardiac myocytes to activate genes in various growth pathways. Initially, hypertrophy may help the heart compensate for acute loss of myocardial tissue from MI or help maintain cardiac function during chronic HTN. But over time, the signals that promote hypertrophy are thought to trigger a type of ventricular remodeling that contributes to progression of HF. Pathologic remodeling includes loss of myocardial cells through apoptosis and production of fibrous changes in the heart that stiffen the ventricles and contribute to diastolic failure (Fig. 19.7). In addition to NE and AII, a number of other factors and immune cytokines have been implicated in cardiac remodeling (Box 19.1). Many of these factors are under investigation as possible targets for drug therapy.

Evidence for a role of AII in remodeling comes from drug studies in which AII production is inhibited or the actions of AII are blocked at the AT1 receptor. An important enzyme in the pathway of AII production is angiotensin-converting enzyme (ACE). Drugs called ACE inhibitors (ACEIs) have been developed to inhibit the activity of this enzyme and prevent formation of AII. Another drug class, angiotensin type 1 receptor blockers (ARBs), binds to this AII receptor and blocks the intracellular actions of AII. ARBs were developed as a more selective means of RAAS

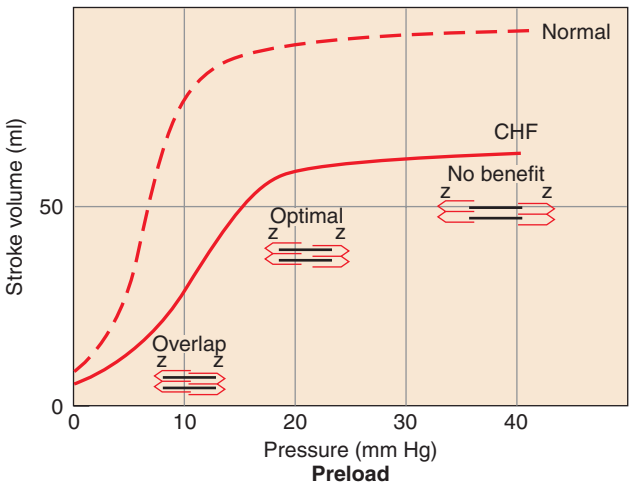


FIG 19.4 Effect of increased preload on sarcomere length and stroke volume. Systolic failure results in a shift of the curve to the right and a dampening of maximal stroke volume. A greater preload is required to achieve a given stroke volume compared with the normal ventricle. *CHF,* Congestive heart failure.

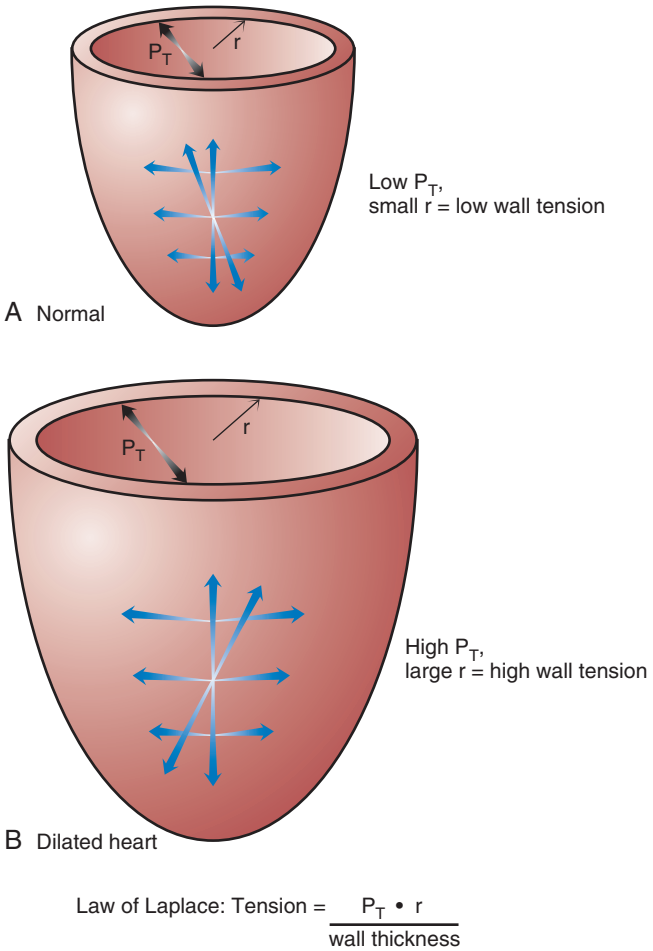


FIG 19.5 Mechanism of myocardial hypertrophy attributable to increased ventricular wall tension. According to the law of Laplace, an increase in chamber radius or pressure will increase wall tension. The hypertrophic response increases wall thickness and helps relieve wall tension. **A,** Heart with normal radius (r) and intraventricular pressure. **B,** Heart with enlarged chamber and high intraventricular pressure. P_T, Transmural pressure.

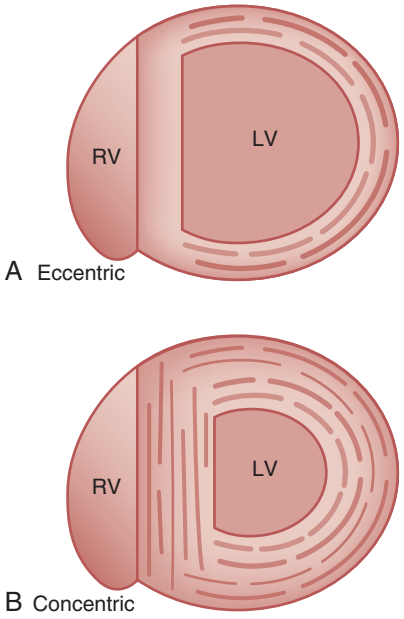


FIG 19.6 Forms of ventricular hypertrophy. **A,** Eccentric, in which muscle fibers grow in length and the chamber diameter increases. **B,** Concentric, in which muscle fibers grow in diameter and the ventricular wall becomes thicker. *LV,* Left ventricle; *RV,* right ventricle.

BOX 19.1 Examples of Neurohormones and Cytokines That Are Dysregulated in Heart Failure

- Norepinephrine
- Epinephrine
- Renin activity
- Angiotensin II
- Aldosterone
- Vasopressin
- Neuropeptide Y
- Vasoactive intestinal peptide
- Prostaglandin
- A and B natriuretic peptide
- Endothelin
- Beta endorphin
- Calcitonin gene-related peptide
- Growth hormone
- Cortisol
- Tumor necrosis factor (TNF) alpha
- Neurokinin A
- Substance P
- Adrenomedullin
- Other immune cytokines

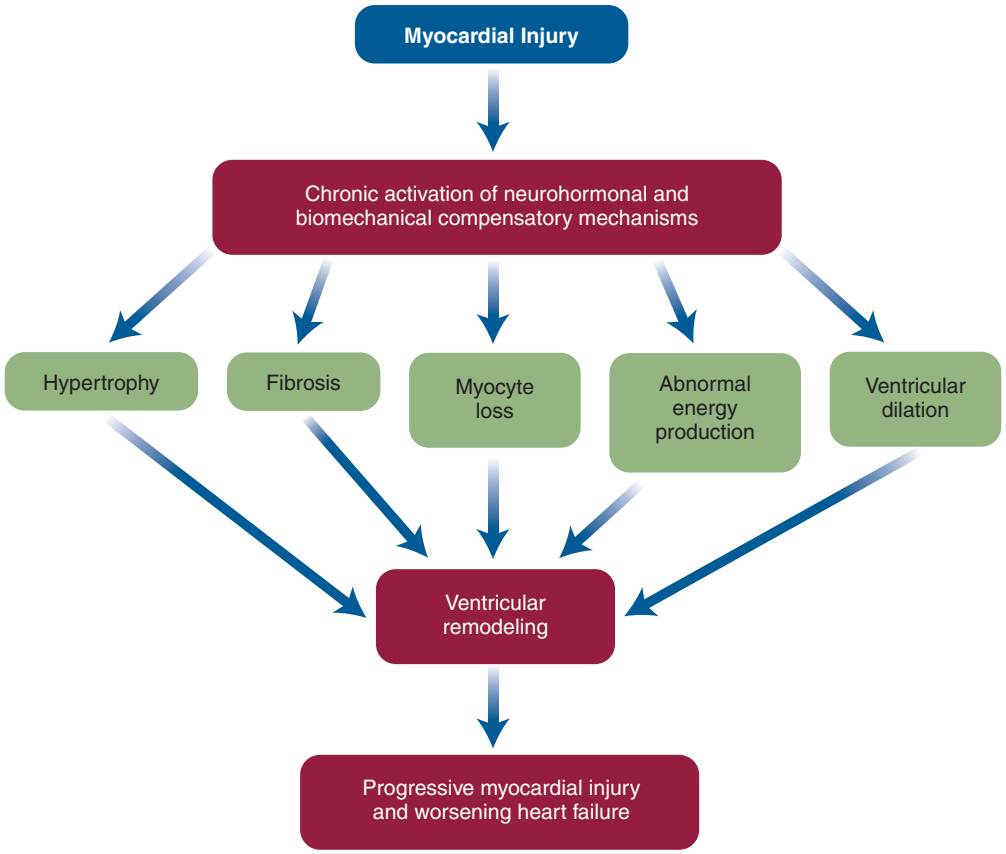


FIG 19.7 Mechanisms of cardiac remodeling that contribute to heart failure progression.

inhibition and to improve the safety and tolerability profile of ACEIs. Both ACEIs and ARBs have been shown to significantly reduce HF mortality, and both drug classes are used as standard therapy in HF. In general, ARBs are prescribed if a patient is intolerant of ACEIs. Drugs that block β_1 receptors and antagonists of aldosterone have also been shown to improve mortality in HF (see "Treatment" section).

In summary, enhanced preload and cardiac hypertrophy may allow a heart to compensate for reduced ventricular function for an extended period. Unfortunately, these compensatory mechanisms, which serve to restore cardiac output to the tissues, also result in an increase in myocardial work and oxygen requirements and appear to cause pathologic remodeling. Progression and decompensation may occur when the primary disease plus the superimposed burdens of compensation and neurohormonal dysregulation overwhelm the heart's ability to function. The focus of therapy for HF is to maintain a state of compensation by minimizing cardiac work while optimizing cardiac output and preventing or delaying ventricular remodeling.

KEY POINTS

- Compensatory mechanisms are activated in heart failure (HF) in an attempt to improve cardiac output. Unfortunately, these responses also increase myocardial workload and may perpetuate the HF. Treatment is aimed at attenuating the harmful effects of the compensatory responses and neurohormonal dysregulation.
- Sympathetic activation is an early response to reduced cardiac output. Sympathetic nervous system (SNS) activation increases heart rate, contractility, arterial vasoconstriction, and renin release. The failing heart generally has reduced responsiveness to SNS neurotransmitters because of β_1-receptor down-regulation.
- Decreased cardiac output reduces kidney perfusion and leads to activation of the renin–angiotensin–aldosterone system (RAAS) and volume retention. Extra blood volume increases cardiac preload. Higher preload results in more forceful ejection of blood from the heart (Frank–Starling law) and improves cardiac output.
- Cardiac hypertrophy is stimulated by elevated myocardial wall tension and the growth-promoting actions of neurohormones, such as NE and AII. Hypertrophy adds contractile filaments and improves contractile force.
- The mechanisms that enable the heart to compensate for reduced stroke volume are detrimental in the long term. Excessive neurohormones, volume overload, and high wall tension contribute to abnormal ventricular remodeling. Gradually, the ventricle loses myocytes and accumulates fibrotic tissue. The remaining myocytes are usually hypertrophied and less efficient. These processes lead to progression of HF over time.

Clinical Manifestations

The clinical presentation of HF differs depending on which ventricle (left, right, or both) is failing to pump blood adequately. Left ventricular failure is the most common presentation of HF. Because of circulatory dynamics, left ventricular failure often leads to right ventricular failure—a condition termed *biventricular failure.* The etiologic process, clinical manifestations, and management of isolated right ventricular failure differ substantially from those for left ventricular and biventricular failure. Recall that the right side of the heart receives blood from the systemic venous circulation and pumps blood into the pulmonary system, whereas the left side of the heart receives blood from the pulmonary circulation and delivers it to the systemic arterial system (Fig. 19.8). Insufficient cardiac pumping is manifested by poor cardiac output, called *forward failure,* and by congestion of blood behind the pumping chamber, called *backward failure.* The clinical manifestations of left and right ventricular failure differ as a result of the anatomic location of the "backward" or congestive processes, but the forward effects of low cardiac output are the same.

The *forward* effects of HF are due to insufficient cardiac output with diminished delivery of oxygen and nutrients to peripheral tissues and organs. Inadequate perfusion of the brain may lead to restlessness, mental fatigue, confusion, anxiety, and impaired memory. Generalized fatigue, activity intolerance, and lethargy may be present.

Reduced perfusion of the kidney results in a decline in urine output (oliguria) with subsequent fluid retention. Activation of the RAAS contributes to conservation of sodium and water by the kidney and may also cause blood vessel constriction. Constriction of blood vessels serves to maintain blood pressure and redistribute reduced cardiac output to vital organs. However, this vasoconstriction also increases afterload, so the damaged left ventricle must generate more force to pump the same volume of blood. Depending on how much the afterload increases, the damaged left ventricle may not be able to pump sufficient quantities of blood into the circulation. If renal blood flow becomes severely limited, the patient with left ventricular failure may develop kidney failure. Forward failure also results in activation of the SNS because of the baroreceptor reflex. Sympathetic activation contributes to blood vessel constriction and helps maintain blood pressure in the face of reduced cardiac output; however, as with AII, SNS activation increases left ventricular afterload. SNS activation results in a compensatory increase in heart rate that may augment cardiac output to some extent, but also raises myocardial ATP consumption.

Left-Sided Heart Failure

Left-sided HF is most often associated with left ventricular infarction and systemic HTN. The *backward effects* of left-sided HF may produce dramatic clinical symptoms attributable to pulmonary dysfunction (Fig. 19.9). Ineffective pumping of the left ventricle results in an accumulation of blood within the pulmonary circulation. As hydrostatic pressure builds within the pulmonary veins and capillaries, fluid is forced from the capillaries into interstitial and alveolar spaces, causing edema. Pulmonary congestion and edema are associated with a number of clinical findings (Fig. 19.10). Dyspnea, or breathlessness, occurs early in the progression of left-sided HF and may be considered the cardinal symptom. Difficulty breathing may be exacerbated by activity (dyspnea on exertion), lying down (orthopnea and paroxysmal nocturnal dyspnea), and blood volume expansion from excessive salt or fluid intake. Orthopnea and paroxysmal nocturnal dyspnea are due in part to a redistribution of blood volume from the periphery to the heart when the individual lies down. The failing left ventricle is unable to effectively pump extra volume, and pulmonary congestion is worsened. The severity of orthopnea may be quantified by the degree of head elevation (e.g., number of pillows) used to relieve dyspnea. Paroxysmal nocturnal dyspnea refers to intermittent attacks of severe dyspnea during the night and is a most distressing form of orthopnea. The individual experiences a feeling of suffocation and panic at not being able to overcome the dyspnea. Sitting or standing helps relieve the dyspnea because blood pools in the extremities, reducing pulmonary hydrostatic pressure and congestion.

Clinical signs of pulmonary congestion include cough, respiratory crackles (rales), hypoxemia, and high left atrial pressure (LAP). Cough results from bronchial irritation associated with congestion. In severe cases, sputum may be blood tinged, from breakage of fragile capillaries, and frothy, from fluid buildup in the alveoli. The severity of pulmonary edema can be estimated from the location of crackles within the lung fields. Crackles are abnormal sounds caused by the movement of air through partially fluid-filled alveoli. Edema fluid collects in dependent lung fields because of gravity and progressively moves up the lung as more edema fluid accumulates. For example, in mild pulmonary edema, crackles might be heard with a stethoscope only at the base of the upright lung, but with increasing severity they become apparent in the

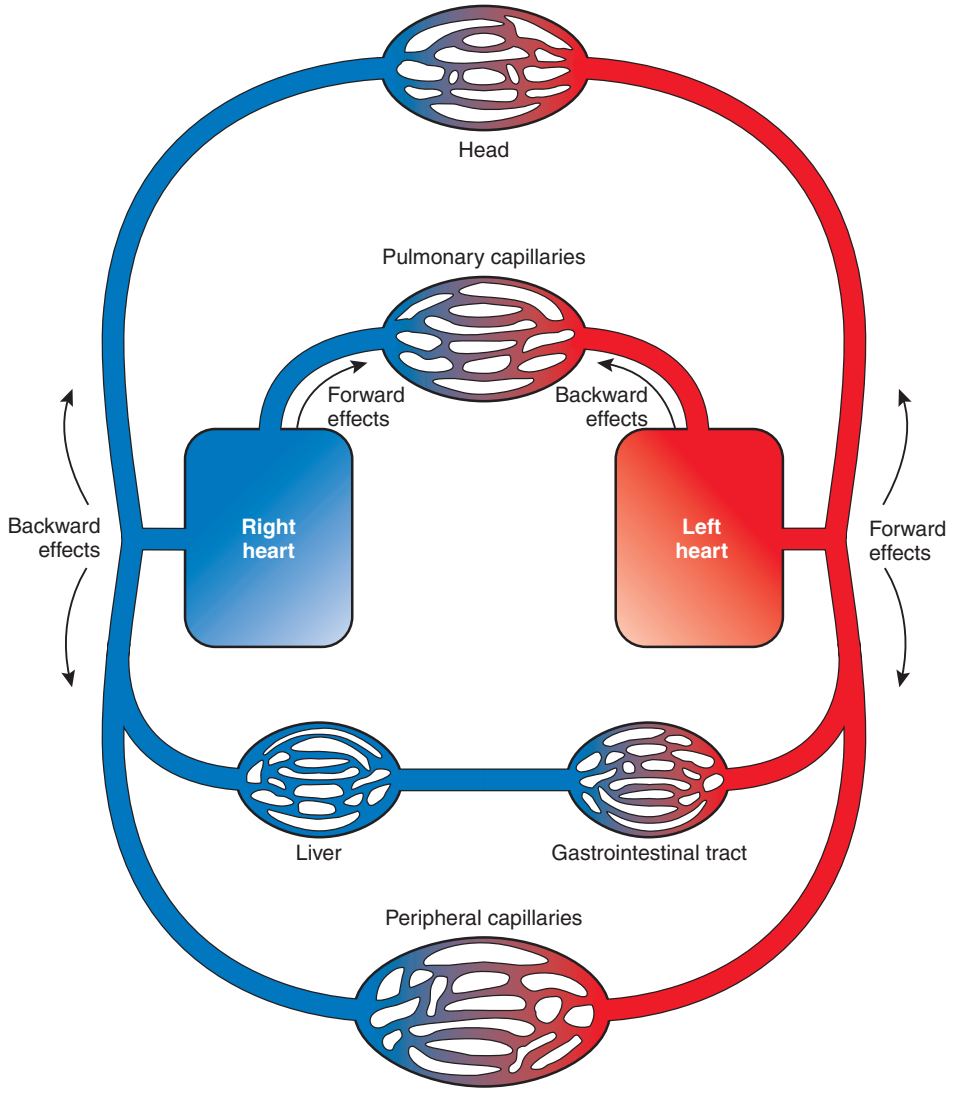


FIG 19.8 Systemic and pulmonary circulations viewed as separate but interdependent systems.

lower third to half of the lung. Fluid in the alveoli and interstitial spaces also interferes with alveolar–capillary gas exchange and results in some degree of hypoxemia. Hypoxemia may be detected by arterial blood gas analysis or pulse oximetry and may be apparent clinically as cyanosis. Cyanosis refers to a blue coloration of the skin typically seen around the mouth (circumoral cyanosis) and results from the presence of significant amounts of desaturated hemoglobin in the blood. Cyanosis is a late sign and is clinically evident only when a large amount (about 5 g/dL) of hemoglobin is deoxygenated (less than or equal to 75% saturated).

Elevated LAP is a common finding in left-sided HF because of excessive blood volume and the compensatory responses of atrial dilation and hypertrophy. Atrial pressure can be estimated by inserting a balloon-tipped catheter (Swan–Ganz) into the pulmonary artery. If LAP acutely increases to 25 mm Hg (normal is 4 to 12 mm Hg), increased capillary filtration leads to pulmonary edema. Patients with chronic elevations in LAP associated with chronic HF are more resistant to developing acute pulmonary edema and may not experience symptoms until pressures approach 40 mm Hg. On x-ray, findings of fluid overload include an enlarged heart and engorged pulmonary capillaries and lymphatic vessels.

Acute cardiogenic pulmonary edema is a life-threatening condition associated with left ventricular failure that severely impairs gas exchange, producing dramatic signs and symptoms. The patient exhibits severe dyspnea and anxiety, and a bolt-upright posture is usually assumed in order to maximize respiratory effort. Bubbly crackles may be heard all the way up the lung from the bases to the apices, and pink frothy sputum may be expectorated or well up from the trachea into the nose and mouth. Anxiety and hypoxemia contribute to tachycardia, which may worsen the pumping efficiency of the failing heart. Cyanosis and symptoms of tissue hypoxia are usually apparent. The immediate treatment is aimed at reducing the fluid volume in the lungs and supporting oxygenation.

Right-Sided Heart Failure

Because the right and left ventricles function in series, left ventricular failure eventually increases the workload on the right ventricle. Consequently, the right ventricle may fail. The etiology of right ventricular failure must include all the causes of left ventricular failure. Isolated right ventricular failure is rare and is usually a consequence of right ventricular infarction or pulmonary disease. Pulmonary disorders that result in increased pulmonary vascular resistance impose a high afterload

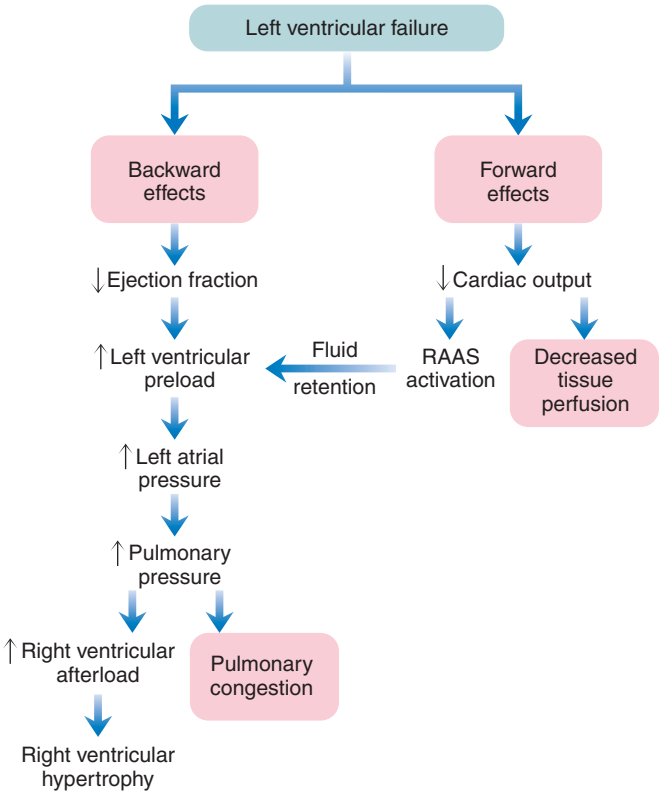


FIG 19.9 Pathophysiologic process of isolated left-sided heart failure, showing backward and forward effects. *RAAS,* Renin–angiotensin–aldosterone system.

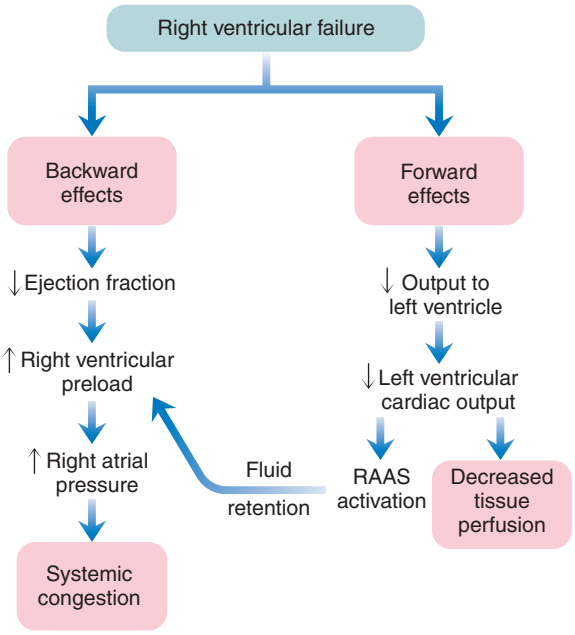


FIG 19.11 Pathophysiologic process of isolated right-sided heart failure, showing backward and forward effects. *RAAS,* Renin–angiotensin–aldosterone system.

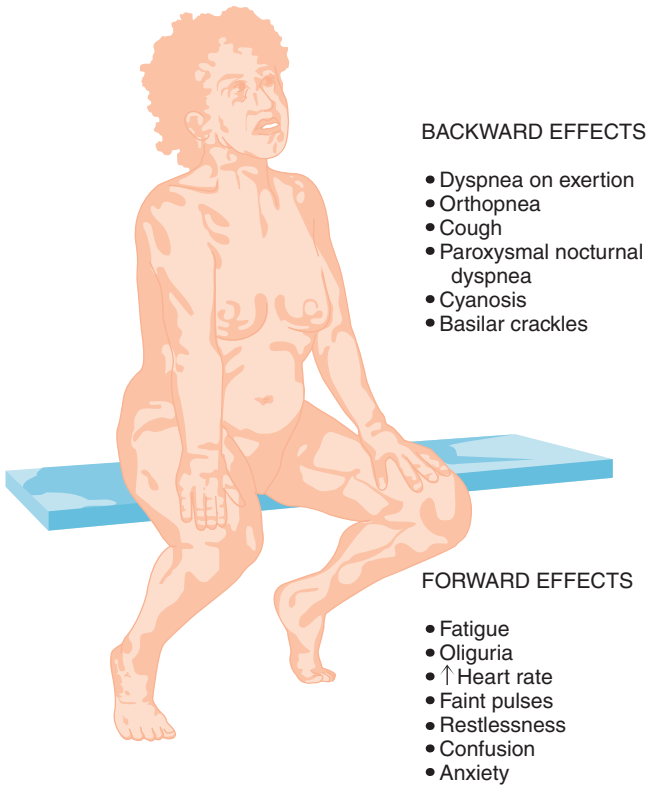


FIG 19.10 Clinical manifestations of isolated left-sided heart failure.

on the right ventricle. The resultant right ventricular hypertrophy, called cor pulmonale**,** may progress to right ventricular failure as the lung disease worsens.

Any lung disorder that decreases the total cross-sectional area of the lung vasculature can increase pulmonary vascular resistance and produce right ventricular strain. Hypoxemia, for example, causes the pulmonary arterioles to constrict, which increases pulmonary resistance. Constriction or blockage of the vascular bed, such as occurs with pulmonary HTN or pulmonary embolus, similarly reduces the cross-sectional area of the pulmonary vasculature and leads to increased pulmonary resistance. If the increase in pulmonary resistance and right ventricular workload occurs gradually, the right ventricle can compensate by increasing preload and hypertrophy. However, the thin musculature of the right ventricle has limited ability to adjust to acute changes in workload, as would occur with a right ventricular infarction or large pulmonary embolus.

As with left-sided HF, congestion of blood occurs behind the failing right ventricle because of inefficiency of the pump. The *backward effects* of right-sided HF are due to congestion in the systemic venous system (Fig. 19.11). Systemic venous congestion results in impaired function of the liver, portal system, spleen, kidneys, peripheral subcutaneous tissues, and brain (Fig. 19.12).

The liver is usually somewhat increased in size and weight, but individual hepatocytes may show signs of atrophy and necrosis attributable to chronic passive congestion. Impedance to blood flow through the liver may cause hydrostatic pressure to build in the portal system, leading to edema formation in the peritoneal cavity (ascites). Increased pressure in the portal system is reflected back to the spleen and gastrointestinal tract. The spleen is generally enlarged (congestive splenomegaly), and gastrointestinal symptoms such as anorexia and abdominal discomfort may be present.

Increased systemic venous pressure causes congestion of the kidneys, which contributes to the decreased glomerular filtration and fluid retention. Fluid retention may be perpetuated by the congested liver, which is unable to metabolize plasma aldosterone normally. Excess

fluid volume and venous congestion caused by right-sided HF result in subcutaneous edema. Edema is usually particularly apparent in the lower extremities or sacral area if the patient is supine.

Drainage of venous blood from the head and neck by way of the superior vena cava is also impeded by right-sided HF. The jugular veins may be abnormally distended, and mental functioning may be impaired. The hepatojugular reflux test can be done to assess the severity of right-sided HF. The liver is manually compressed, causing a sudden increase in venous blood returning to the right heart, while jugular neck veins are observed for sudden distention. In the absence of right-sided HF, the sudden increase in venous return would enter the heart unimpeded, and no neck vein distention would be apparent.

Biventricular Heart Failure

In many cases, HF is not isolated to one side of the heart. Biventricular failure is most often a result of primary left ventricular failure that has progressed to right-sided HF. With biventricular failure, cardiac output is reduced and pulmonary congestion exists as a result of left-sided HF, as well as systemic venous congestion attributable to right-sided HF.

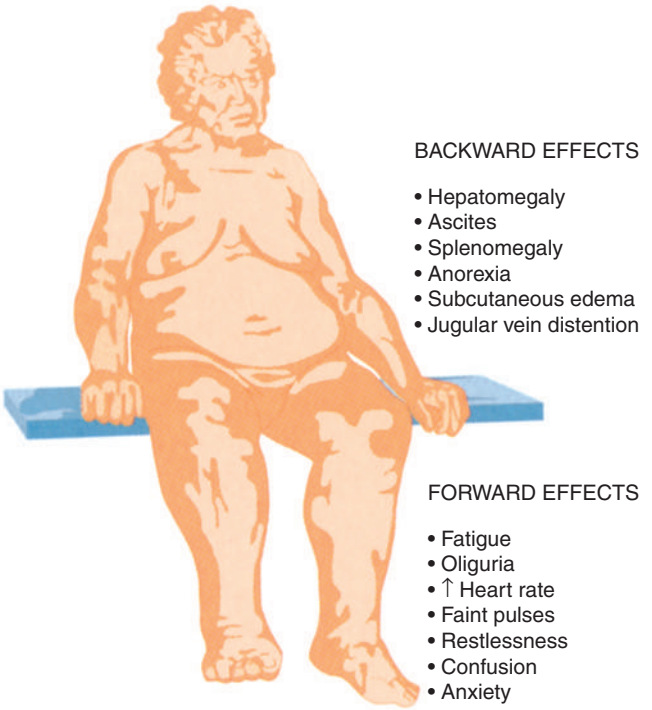


FIG 19.12 Clinical manifestations of isolated right-sided heart failure.

Class and Stage of Heart Failure

A variety of clinical criteria have been used to aid in the diagnosis of HF, but to date, no diagnostic test has been universally accepted. A commonly suggested tool for detecting HF in previously undiagnosed individuals is called FACES of HF: *f*atigue, *a*ctivity limitation, *c*ongestion, *e*dema, *s*hortness of breath. These are classic findings in HF and, if present, indicate a need for further diagnostic assessment. Traditionally, a patient with suspected HF would be diagnosed by x-ray and echocardiography. Echocardiography is the gold standard for evaluation of HF and is able to measure chamber size, volume, and dynamics during diastole and systole, as well as determine EF. A blood test for B-type natriuretic peptide (BNP) may also be used to help identify patients with HF. BNP is synthesized by left ventricular myocytes under conditions of increased left ventricular wall stress. A significant correlation between the amount of plasma BNP or the BNP precursor, N-terminal pro-BNP (NT-proBNP), and the severity of HF has been documented.

The severity of symptoms can be used to assign an HF class (New York Heart Association classes [NYHA] I to IV). Another classification scheme has been proposed by the American Heart Association (AHA) to allow inclusion of patients at high risk for HF but whose disease is not yet symptomatic. By including this pre-HF group, efforts aimed at prevention may be instituted in more patients. Class and stage are used to determine prognosis, therapy, and monitoring in HF. These two classification schemes are compared in Table 19.1.

KEY POINTS

- The clinical manifestations of heart failure (HF) are characterized by the effects of forward failure (reduced cardiac output) and backward failure (congestion behind the pumping chamber).
- Insufficient cardiac output may be manifested as confusion, fatigue, tachycardia, reduced urine output, and poor peripheral circulation.
- Left-sided HF is characterized by pulmonary congestion, which may manifest with dyspnea, orthopnea, crackles, cough, pulmonary edema, and hypoxemia.
- Right-sided HF is characterized by systemic venous congestion, which may manifest with jugular vein distention, hepatomegaly, splenomegaly, and peripheral edema.
- Left-sided HF frequently leads to development of right-sided HF. With biventricular failure, congestive signs and symptoms are found in both the pulmonary and the systemic venous circulation.
- HF is diagnosed by signs and symptoms, x-ray findings, and echocardiographic findings. A plasma BNP or NT-proBNP level may be used to diagnose HF in patients with shortness of breath.
- The severity of signs and symptoms is used to assign a HF class or stage.

TABLE 19.1 Comparison of ACC/AHA Stages of Heart Failure and New York Heart Association Classes

ACC/AHA Stage	Description	Clinical Clues	NYHA Class
A	Patients at high risk of developing HF	Coronary artery disease, hypertension, diabetes, dyslipidemia, family history of cardiomyopathy	Not applicable
B	Patients who have structural heart disease but have never manifested signs or symptoms of HF	Left ventricular hypertrophy (by ECG or echo), valvular disease, past myocardial infarction	I
C	Patients who have current or previous symptoms of HF	Dyspnea, fatigue, exercise intolerance, prior HF hospitalization	II–III
D	Patients with advanced structural heart disease and marked symptoms of HF at rest	End stage, awaiting transplant, receiving palliative care	IV

Adapted from Hunt SA et al: American College of Cardiology; American Heart Association: 2009 focused update incorporated into the ACC/AHA 2005 guidelines for the diagnosis and management of heart failure in adults: summary article, *J Am Coll Cardiol* 2009;53(15):e3–e62.

Treatment

Therapy for HF is aimed at improving cardiac output while minimizing congestive symptoms and cardiac workload. These objectives are obtained by manipulating preload, afterload, and contractility. When possible, specific treatment is undertaken to correct the underlying cause of the HF. Use of medications shown to improve mortality are standards of care for all HF patients who are able to tolerate them.

Despite the large number of pharmacologic agents being used in the management of HF, only a few have been associated with significant improvement in mortality risk, particularly ACEIs, ARBs, aldosterone antagonists, and certain β_1-blockers.

A better understanding of the underlying molecular mechanisms of HF is needed to improve pharmacologic management, particularly for patients with preserved EF who have been underrepresented in studies. Many patients with HF have signs and symptoms of elevated preload attributable to an expanded intravascular volume and a reduced EF. According to the Frank–Starling law, an elevated preload is desirable to enhance systolic shortening and improve cardiac output. Unfortunately, high preload exacerbates congestive symptoms and adds to the workload of an already damaged heart. Thus the aim of therapy is to optimize preload so that congestive symptoms are minimized but cardiac output is not compromised. The right ventricle is particularly sensitive to reductions in preload, and care must be taken to avoid a significant drop in right ventricular output when intravascular volume is decreased. Drugs, such as diuretics, may be administered to reduce intravascular volume. Diuretics promote the excretion of fluid by increasing renal blood flow, blocking sodium and chloride reabsorption, or both. Patients may also be instructed to modify salt and fluid intake.

With the exception of digitalis, positive inotropic agents are avoided in chronic HF because they are associated with higher mortality. However, when improving cardiac output is necessary to avoid shock, positive inotropic agents may be used (see Chapter 20). Positive inotropes work by increasing the availability of intracellular calcium ions during systole. Drugs that mimic SNS effects, such as NE, isoproterenol, dobutamine, and dopamine, may be used to improve cardiac output, but have a potential for dramatically increasing myocardial oxygen consumption.

Digitalis or a related cardiac glycoside may be used to manage HF symptoms. Cardiac glycosides directly inhibit the sodium–potassium pump present in the cell membrane of all cells. This results in an increase in intracellular sodium accumulation and a decrease in the gradient for sodium entry into the cell. A diminished sodium gradient slows the sodium-dependent calcium pump that normally removes intracellular calcium. This allows more calcium to enter the sarcoplasmic reticulum, thus strengthening myocardial contraction. Digitalis also slows the heart rate through parasympathetic system activation and promotes sodium and water excretion through improved cardiac output to the kidney. Depletion of serum potassium (hypokalemia) may potentiate digitalis toxicity. Unlike the other positive inotropic agents, digitalis does not appear to increase mortality. Digitalis also does not improve mortality and is not recommended as routine therapy.

Patients with enlarged hearts and conduction delays may benefit from pacemakers, which help synchronize ventricular contraction. A wide QRS complex is the usual indication for resynchronization therapy in the HF patient. Pacing electrodes are placed in the atrium and both ventricles to allow coordinated depolarization of the heart muscle. Many patients experience significant improvement in congestive symptoms and activity tolerance with resynchronization.

In addition to the use of ACEIs, ARBs, β_1-blockers, and aldosterone antagonists for their mortality benefits, treatment goals include controlling etiologic factors, such as HTN, ischemia, and rhythm abnormalities. Measures to assess and improve symptoms and quality of life are paramount.

KEY POINTS

- Important aims of treatment are to improve cardiac output, minimize congestive symptoms, and prevent progression.
- ACEIs, ARBs, aldosterone antagonists, and some β_1-blocking agents have been shown to improve mortality risk in patients with heart failure (HF).
- Diuretics alleviate congestive symptoms and reduce cardiac workload by reducing preload.
- Myocardial contractility may be improved acutely by positive inotropic drugs, such as β-agonists; however, long-term use is associated with higher mortality. An exception is digitalis, which has positive inotropic effects, does not appear to alter mortality, and may improve symptoms in some patients.
- Resynchronization of ventricular depolarization with pacemakers may improve contraction in patients with wide QRS complexes.
- Efforts to improve etiologic factors, such as atherosclerosis, HTN, and atrial dysrhythmias, are recommended.
- Numerous neurohormonal and cytokine factors are dysregulated in HF and continue to be investigated as potential targets of drug therapy.

CARDIAC DYSRHYTHMIAS

Dysrhythmia or arrhythmia refers to a cardiac rhythm abnormality affecting impulse generation or conduction. A normal heartbeat is initiated at an appropriate rate in the sinoatrial (SA) node and follows a consistent pathway of depolarization through the atria, atrioventricular (AV) node, His–Purkinje system, and, finally, the ventricular myocardium (see Chapter 17). Electrical depolarization of the heart is normally followed by atrial and then ventricular muscular contraction. A number of factors may lead to disturbances in heartbeat, including hypoxia, electrolyte imbalance, trauma, inflammation, and drugs. Dysrhythmias are significant for two reasons: (1) they indicate an underlying pathophysiologic disorder and (2) they can impair normal cardiac output. Dysrhythmias can be categorized into three major types: abnormal rates of sinus rhythm, abnormal sites (ectopic) of impulse initiation, or disturbances in conduction pathways.

Dysrhythmia Mechanisms

Disorders of impulse generation result from abnormalities in the *rate* of impulse generation from a normal pacemaker or from impulse generation from an abnormal *(ectopic)* site. Abnormal automaticity and triggered activity are the two mechanisms most commonly cited for dysrhythmias of impulse generation. Abnormalities of impulse conduction include conduction blocks and reentry phenomena.

Automaticity

Failure to repolarize to normal resting membrane potential or abnormal plasma membrane leakiness to sodium or calcium ions at rest (phase 4) results in a shift in the resting membrane potential toward threshold, generating an action potential. Ischemia and subsequent ATP deficiency reduce the cell's ability to control electrolyte flux across the cell membrane. Electrolyte imbalance, particularly hypokalemia, contributes to abnormal automaticity. Alterations in spontaneous calcium ion flux from the sarcoplasmic reticulum and the subsequent triggering of calcium influx across the plasma membrane may also contribute to abnormal automaticity (see Chapter 17).

Triggered Activity

Triggered activity occurs when an impulse is generated during or just after repolarization because of a depolarizing oscillation of the membrane potential (Fig. 19.13). Early afterdepolarizations occur during the relative refractory period of phase 3. Some of the voltage-gated calcium channels

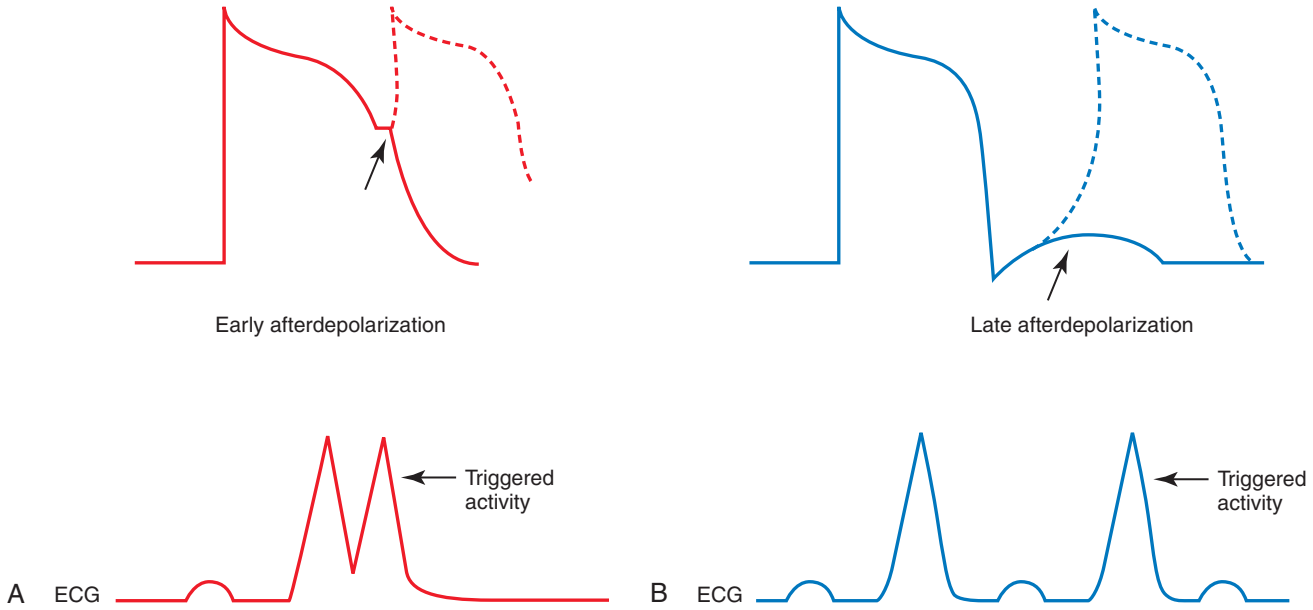


FIG 19.13 Mechanisms of triggered activity. **A,** Early afterdepolarization in ventricular cell showing triggered action potential during the repolarization phase. The electrocardiogram (ECG) shows an R wave occurring on top of the T wave (R-on-T phenomenon). **B,** Late or delayed afterdepolarization occurs after the repolarization phase has been completed and results in an early beat after the T wave. If the late afterdepolarization does not reach threshold, no triggered beat will occur.

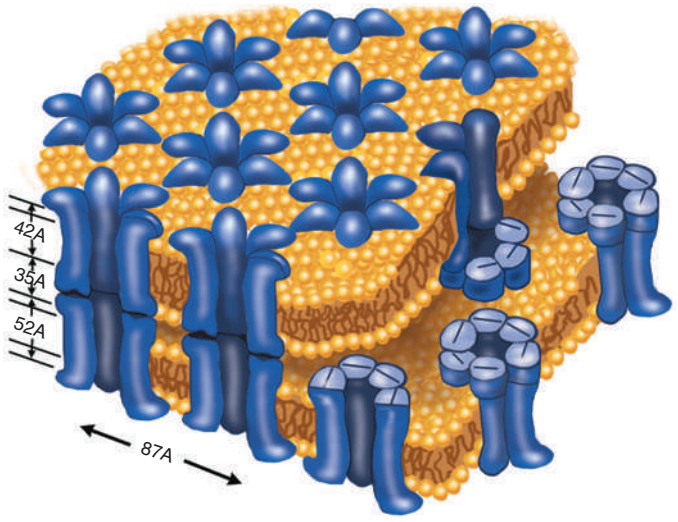


FIG 19.14 Connexons form the gap junction pores in the intercalated disks that connect cardiac myocytes together and allow them to function as a syncytium. Genetic abnormalities in the structure of connexons may contribute to arrhythmia formation by altering the rate of conduction through them. (From Saffitz JE: Cell-to-cell communication in the heart *Cardia/Rev* 3.86, 1995.)

are thought to reopen during phase 3 and trigger another impulse (see Fig. 19.13A). Delayed or late afterdepolarizations occur after the repolarization phase is complete and are seen as oscillating depolarizing waves on the electrocardiogram (ECG) (see Fig. 19.13B). If the delayed afterdepolarization reaches threshold, it will trigger an action potential. Digitalis toxicity and excessive catecholamine stimulation may contribute to this mechanism. Delayed afterdepolarizations are thought to occur because calcium ions are spontaneously released from the sarcoplasmic reticulum after repolarization. An increase in intracellular free Ca^{2+} concentration during phase 4 can trigger Ca^{2+} influx across the plasma membrane and release more Ca^{2+} from the sarcoplasmic reticulum, resulting in an action potential. A number of genetic abnormalities in intracellular calcium handling have been associated with triggered activity mechanisms.

Reentry

Reentry is thought to be the culprit in most tachydysrhythmias, including atrial and ventricular tachycardia, flutter, and fibrillation. Reentry is a complex process in which a cardiac impulse continues to depolarize in a part of the heart after the main impulse has finished its path and the majority of the fibers have repolarized. If the errant impulse proceeds slowly enough, it may eventually meet with nonrefractory cells and initiate an extra, ectopic cardiac depolarization. Reentry processes are produced when electrical conduction in a portion of the heart is abnormally slowed (functional) or has an unusually long pathway (anatomic). A number of theories have been proposed to describe the generation and conduction of reentry depolarizations, including specific subsets of cells that continue to produce waves of activation or more general involvement of the myocardial tissue in propagating waves of conduction. The rate of conduction is controlled in part by the density and structure of gap junctions that connect the cardiac cells together. Recall that a wave of depolarization moves from cell to cell through these gap junctions; however, different regions of the heart may have different gap junction properties that predispose to different conduction rates. Mutations in the genes that code for gap junction proteins (connexins) may increase the predisposition to dysrhythmias (Fig. 19.14). Reentry depolarizations may occur as complex spiral waves in which the activating wavefront follows, or "chases," its repolarizing tail (Fig. 19.15). If the wavefront encounters only refractory tissue, the reentrant process suddenly terminates. Myocardial ischemia, excessive catecholamines, and electrolyte abnormalities predispose to reentry mechanisms.

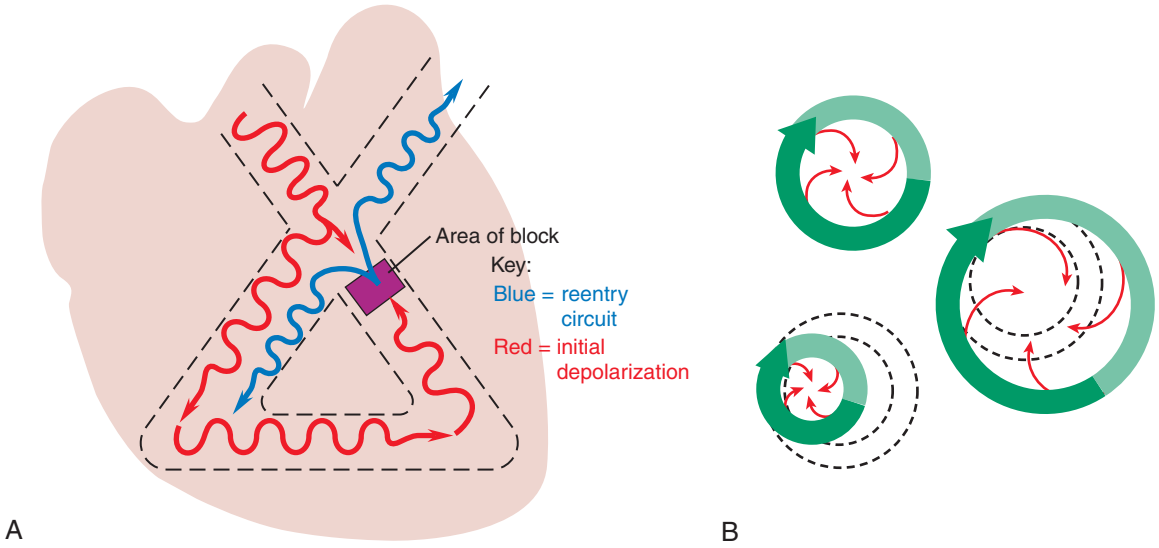


FIG 19.15 Mechanism of reentry. **A,** A wave of depolarization that travels slowly or by an abnormal pathway may encounter myocardium that has had time to recover and can restimulate it. **B,** This may result in an extra beat, or the depolarization may continuously "chase its tail" in a circuit, causing defibrillation. The circuit can be relatively fixed or can wander into various "wavelets."

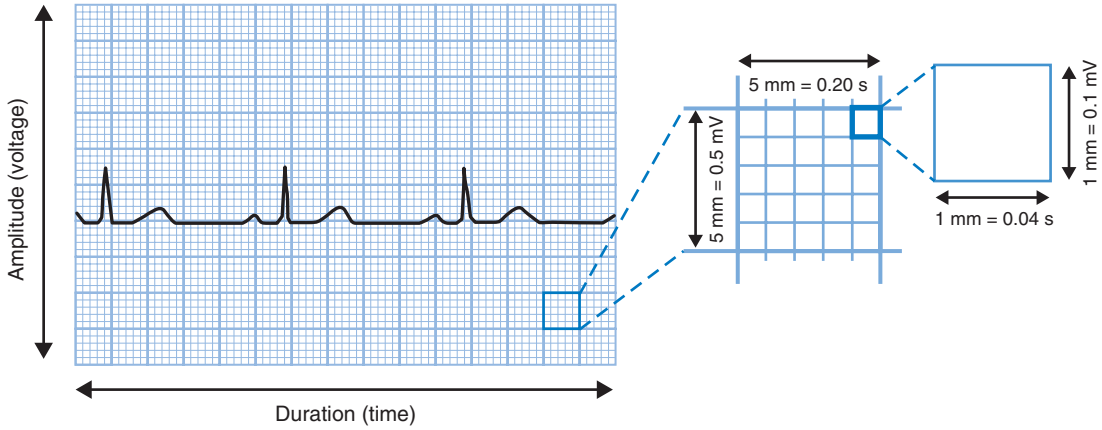


FIG 19.16 Electrocardiographic strip showing the markings for measuring amplitude and duration of waveforms, using a standard recording speed of 25 mm/sec.

Dysrhythmia Analysis

ECG recording paper is specifically designed to allow easy measurement of waveform amplitude and duration (Fig. 19.16). Each small box on the ECG paper represents an amplitude of 0.1 mV and a duration of 0.04 second (paper speed at 25 mm/sec). Larger boxes are also marked on the paper and correspond to 0.5 mV in amplitude (five small boxes) and 0.2 second in duration (five small boxes). These markings allow measurement of waveform amplitude, duration, and heart rate. Rhythm strips presented in this chapter are from a single lead only (usually lead II), but it should be emphasized that thorough ECG interpretation often requires several leads to provide different views of electrical conduction through the heart. (Lead placement is discussed in Chapter 17.)

Normal Sinus Rhythm

Before one proceeds to the interpretation of dysrhythmias, the features of normal sinus rhythm must be understood. Normal sinus rhythm is generally defined as an impulse rate between 60 and 100 per minute that begins in the sinus node and follows the normal conduction pathway.

TABLE 19.2 Electrocardiographic Characteristics of Normal Sinus Rhythm

Characteristic	Findings
Rhythm	Regular, PP intervals and RR intervals may vary as much as 3 mm and still be considered regular
Rate	60–100 beats/min
P waves	One P wave preceding each QRS
PR interval	0.12–0.20 sec, constant
QRS duration	0.04–0.10 sec, constant
QT interval	0.40 sec (varies with rate)

Characteristics of normal sinus rhythm are listed in Table 19.2. The rhythm shown in Fig. 19.17 is regular; there is a P wave for every QRS complex; the PR, QRS, and QT intervals are of normal duration; and there are no "funny-looking" beats.

There are several methods for determining heart rate using the rhythm strip. The easiest but least accurate method is to count the number of QRS complexes within 6 seconds and multiply by 10. ECG

FIG 19.17 Normal sinus rhythm (rate, 64/min).

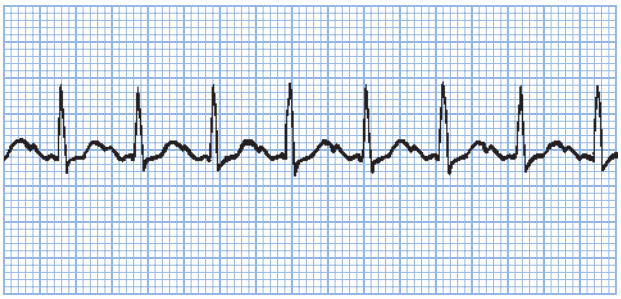


FIG 19.18 Sinus tachycardia (rate, 150/min).

paper has 3-second marks along the top that can be used to determine a 6-second interval. A more accurate method for determining heart rate is to count the number of small boxes between complexes. The number of boxes is divided by 1500 to determine heart rate because there are 1500 small boxes per minute (1500 × 0.04 second = 60 seconds). Neither of these methods is accurate with irregular rhythms, so heart rate must be calculated for a longer interval, usually 1 minute. With this understanding of rate calculation and methods to measure the duration and amplitude of waveforms, one can analyze dysrhythmias.

KEY POINTS

- Dysrhythmias are initiated by three types of depolarizing mechanisms: abnormal automaticity, triggered activity from afterdepolarizations, and reentrant circuits.
- Measurement of electrocardiogram (ECG) waveform amplitude, duration, and frequency is necessary to analyze cardiac rhythms. ECG paper is marked in small boxes representing 0.1 mV of amplitude and 0.04 second of duration.
- Normal sinus rhythm is characterized by regular PP and RR intervals, a rate of 60 to 100 beats/min, and normal PR (0.12 to 0.20 second) and QRS (0.04 to 0.10 second) intervals.

Abnormal Rates of Sinus Rhythm

Sinus Tachycardia

Sinus tachycardia is an abnormally fast heart rate of more than 100 beats per minute (Fig. 19.18). A number of factors, including sympathetic activation, decreased parasympathetic activity, fever, hyperthyroidism, pain, increased metabolism, low blood pressure, and hypoxia, can lead to sinus tachycardia, making it a common dysrhythmia. Sinus tachycardia often is a compensatory response to increased demand for cardiac output or reduced stroke volume. Treatment is aimed at correcting the underlying cause. In some instances, however, the rate can become so high that ventricular filling is impaired and cardiac output is compromised. The heart rate at which this occurs will vary depending on age and cardiac function. Sympatholytic agents or calcium channel–blocking agents may then be indicated.

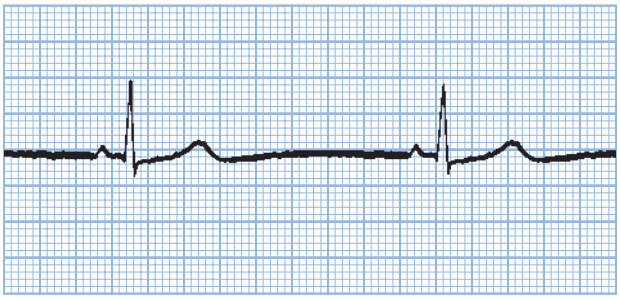


FIG 19.19 Sinus bradycardia (rate, 35/min).

Sinus Bradycardia

Traditionally a heart rate of less than 60 beats per minute is called *bradycardia;* however, lower rates are commonly encountered in physically trained individuals. Sinus bradycardia results from slowed impulse generation by the sinus node in response to increased parasympathetic activity, sleep, drugs, increased stroke volume, or acute HTN (baroreceptor reflex). Important features of sinus bradycardia are shown in Fig. 19.19. Sinus bradycardia may be a normal finding in well-conditioned individuals who have large resting stroke volumes. Abnormal parasympathetic activation can result from pain (vasovagal response), carotid sinus massage, endotracheal suctioning, and the Valsalva maneuver (bearing down). Slow heart rates may be well tolerated by some individuals and not require treatment. If the slow heart rate precipitates low cardiac output, it is usually treated with sympathomimetic or parasympatholytic drugs.

Sinus Arrhythmia

A degree of variability in the heart rate, or sinus arrhythmia, is a normal finding associated with fluctuations in autonomic influences and respiratory dynamics. Sinus arrhythmia can be particularly pronounced in children. Sinus arrhythmia must be differentiated from a sinus node irregularity called *sick sinus syndrome,* in which alternating periods of sinus bradycardia and tachycardia occur (Fig. 19.20). Sick sinus syndrome may necessitate implantation of a permanent pacemaker. Sinus arrhythmia is a normal finding and thus requires no treatment.

Sinus Arrest

The absence of impulse initiation in the heart results in electrical *asystole.* It is characterized by a flat ECG lacking recognizable waveforms (Fig. 19.21). Electrical asystole results in mechanical asystole and zero cardiac output. An escape rhythm from a slower pacemaker will generally begin to fire after several seconds of sinus arrest. Sinus arrest may result from MI, electrical shock, electrolyte disturbances, acidosis, and extreme parasympathetic activity. Prolonged complete electrical asystole is unlikely, and fine ventricular fibrillation may be the underlying rhythm. Sinus arrest may be treatable with a cardiac pacemaker.

KEY POINTS

- Sinus tachycardia (more than 100 beats/min) usually occurs from sympathetic activation of the heart. Sympathetic nervous system (SNS) activation may be compensatory (e.g., occurring in the setting of low blood pressure, low cardiac output, or hypoxemia) or may be due to pain and anxiety.
- Sinus bradycardia (less than 60 beats/min) usually occurs in response to parasympathetic activity. Bradycardia is treated if the slow heart rate precipitates inadequate cardiac output.
- Sinus arrhythmia is usually normal and more pronounced in young persons than in older adults.
- Sinus arrest may lead to prolonged intervals of electrical asystole and zero stroke volume until another pacemaker begins to fire. An artificial pacemaker may be required.

FIG 19.20 A, Sinus arrhythmia is a normal finding that may be particularly pronounced in children. **B,** Sick sinus syndrome. The strip shows alternating periods of tachycardia and bradycardia.

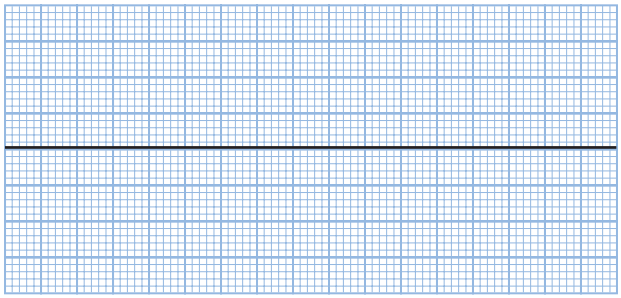


FIG 19.21 Electrical asystole.

FIG 19.22 Junctional escape rhythm (rate, 59/min).

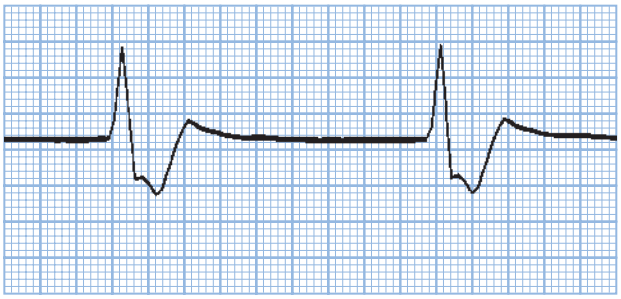


FIG 19.23 Ventricular escape rhythm (rate, 33/min).

Abnormal Site of Impulse Initiation

Initiation of a cardiac impulse at a site other than the SA node occurs primarily for two reasons. First, SA node failure may allow a slower pacemaker to take over. Takeover by a slower pacemaker is called an *escape rhythm*. Second, enhanced excitability, triggered activity, or reentrant circuits may cause premature depolarization and override the SA node.

Escape Rhythms

Escape beats can originate in the AV nodal region or in the ventricular Purkinje fibers. A junctional escape rhythm originates in the AV node, has a rate of 40 to 60 beats per minute, and has a normal QRS configuration (Fig. 19.22). A ventricular escape rhythm originates in the Purkinje fibers, has a rate of 15 to 40 beats per minute, and is characterized by an abnormally wide QRS complex on the ECG (Fig. 19.23). An important clue to identifying escape rhythms is the absence of normal P waves and PR intervals. After the impulse is generated in the Purkinje or nodal cell, it can be conducted backward to the atria (retrograde P wave). Thus a P wave, if present, may be inverted and located before, during, or after the QRS complex. Escape rhythms are usually poorly tolerated because they are slow and associated with decreased cardiac output. Failure of the sinus node can be managed with a pacemaker.

Atrial Dysrhythmias

Premature atrial complexes and tachycardia. Premature atrial complexes (PACs) originate in the atria but not at the SA node. The PAC occurs earlier than normal, is preceded by a P wave, and has a normal QRS configuration (Fig. 19.24). P waves preceding the PAC usually have a different shape (morphology) than the sinus beats. Sometimes the PAC is not conducted through the AV node to the ventricle and is not followed by a QRS complex (nonconducted P wave). Isolated or rare PACs are not clinically significant. However, frequent PACs may indicate an underlying pathophysiologic process and may be precursors to more serious dysrhythmias. Paroxysmal focal atrial tachycardia is a

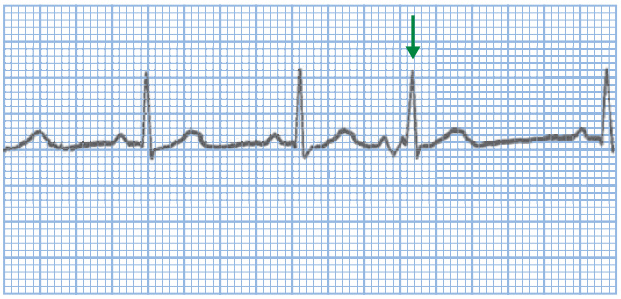


FIG 19.24 Premature atrial complex *(arrow).* Note early P wave and different P wave morphology.

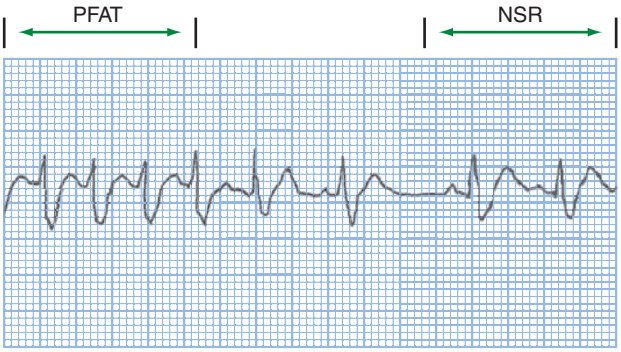


FIG 19.25 Paroxysmal focal atrial tachycardia (PFAT) followed by transition to normal sinus rhythm (NSR).

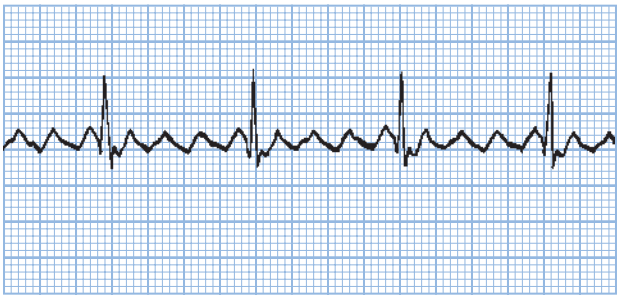


FIG 19.26 Atrial flutter with four atrial depolarizations to one ventricular depolarization.

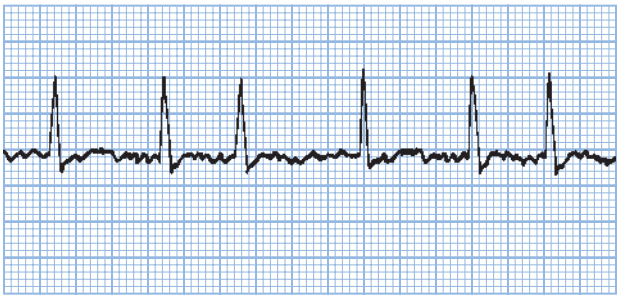


FIG 19.27 Atrial fibrillation showing an irregularly irregular ventricular response.

burst of atrial complexes resembling several PACs in a row (Fig. 19.25). The rhythm is regular at a usual rate of 130 to 240 beats per minute. It may be difficult to distinguish this rhythm from sinus tachycardia; however, differences in P wave configuration are usually apparent. The period of atrial tachycardia may last for minutes, hours, or days and can result in ischemia. Patients may perceive atrial tachycardia as palpitations and may experience chest pain. Focal atrial tachycardia can occur in persons with no underlying heart disease in response to emotional stress or drugs. An episode may start as a PAC that has an abnormally slow conduction time through the atria and AV node. This is thought to allow the wave of depolarization to reexcite previously depolarized cells, resulting in reentry and perpetuation of the abnormal rhythm.

Atrial flutter and fibrillation. Atrial flutter is typically manifested by a rapid atrial rate of 240 to 350 beats per minute and a characteristic sawtooth pattern of atrial depolarizations (Fig. 19.26). There is overlap in the mechanism of atrial tachycardia and atrial flutter, and several types of flutter have been described. These are commonly categorized according to atrial rate; type I (typical) has rates of 240 to 350 beats per minute, and type II has rates in excess of 350 beats per minute. The QRS configuration is normal; however, some of the atrial depolarizations do not conduct through the AV node, resulting in a slower ventricular rate. The ventricular rate may be irregular if there is a variable block or may be regular if there is a uniform block, such as 2:1 or 3:1. Reentry is the probable mechanism for typical atrial flutter. Persons exhibiting atrial flutter usually have underlying heart disease, fluid overload, or atrial ischemia.

Atrial fibrillation is a completely disorganized and irregular atrial rhythm accompanied by an irregular ventricular rhythm of variable rate (Fig. 19.27). The atrial impulses appear as small, squiggly waves of various sizes and shapes. Atrial fibrillation is sustained by multiple reentrant "wavelets" that continually change in size and direction. The majority of atrial depolarizations are blocked at the AV node, with few reaching the ventricles and initiating ventricular contraction. Atrial fibrillation causes the atria to quiver rather than to contract forcefully. This allows blood to become stagnant in the atria and may lead to formation of thrombi. Atrial fibrillation may occur intermittently or be sustained in the long term. Patients with chronic atrial fibrillation often are treated with anticoagulant medications to prevent atrial clot formation. Atrial fibrillation is a significant risk factor for cerebrovascular stroke. Patients with HF may experience more symptoms when they are in atrial fibrillation because the usual "atrial kick" that normally adds 15% to 20% more blood to the ventricle before systole is lost and therefore cardiac output may be reduced. Cardioversion with an electrical shock to the chest is commonly used to manage atrial fibrillation. Numerous antidysrhythmic agents can be used to convert atrial fibrillation to sinus rhythm or control the ventricular response rate, including calcium channel blockers, β-blockers, digitalis, and amiodarone.

Junctional Dysrhythmias

Premature junctional complexes can be initiated in two junctional zones: in the area just proximal to the AV node, where atrial fibers enter, or in the area just distal to the AV node, where nodal fibers enter the bundle of His. The impulse spreads upward into the atrium, causing a P wave, and downward into the ventricle, causing a normally configured QRS complex. The P wave may precede, follow, or be buried in the QRS complex. Premature junctional beats have the same clinical significance as PACs and are generally well tolerated.

Junctional tachycardia is a rapid junctional discharge in the range of 70 to 140 beats per minute (Fig. 19.28). The rhythm resembles a series of junctional premature beats, with P waves preceding, following, or buried in the QRS complexes. Differentiation of the electrocardiographic pattern produced by junctional tachycardia from that produced by atrial tachycardia is often difficult, and the term *supraventricular tachycardia* may be used for both.

FIG 19.28 Junctional tachycardia (rate, 108/min). Note that P waves follow the QRS waves because of retrograde depolarization spreading from the atrioventricular node to the atria. This rhythm may also be called *supraventricular tachycardia.*

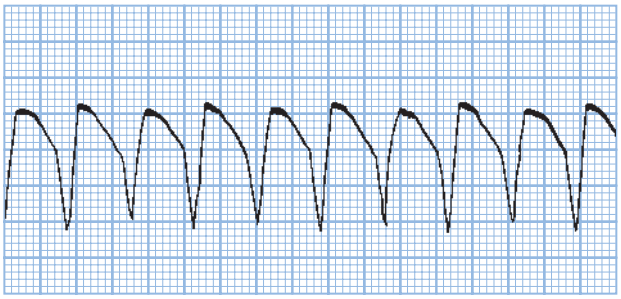


FIG 19.30 Ventricular tachycardia (rate, 178/min).

FIG 19.29 Premature ventricular complex.

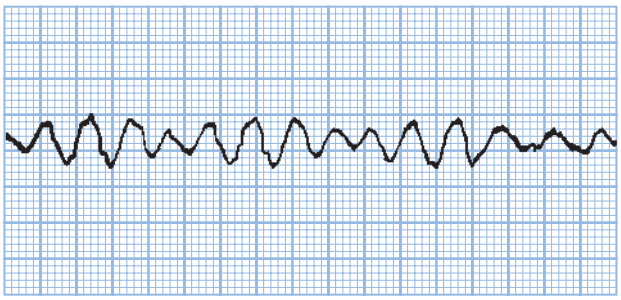


FIG 19.31 Ventricular fibrillation.

Ventricular Dysrhythmias

Premature ventricular complexes. Premature ventricular complexes (PVCs) arise from the ventricular myocardium. The impulse depolarizes the ventricles but does not activate the atria or depolarize the sinus node. Thus the normal rhythm of sinus discharge is not disturbed. The normal sinus impulse is generally buried in the bizarre-looking QRS complex from the premature ventricular beat. The sinus impulse does not result in a QRS complex because the ventricles are refractory from the premature depolarization. The next sinus beat occurs just when it would have occurred normally if there had been no premature beat. Thus the interval between the sinus beat preceding the premature beat and the sinus beat following the premature beat is twice the regular interval (Fig. 19.29). This is known as a *compensatory pause* and helps confirm the diagnosis of PVCs. The QRS of the premature complex is prolonged (more than 0.10 second) and bizarre in appearance. The T wave is usually in a direction opposite to the main QRS deflection. Premature ventricular beats are commonly associated with coronary artery disease, drug overdose, and electrolyte disturbances—particularly hypokalemia and hypomagnesemia. The clinical significance depends in part on the frequency of the premature beats. The PVCs may occur at regular intervals, such as bigeminy (every other beat) or trigeminy (every third beat). With high frequency, cardiac output may be compromised. Frequent PVCs may be managed with antidysrhythmic drugs, such as amiodarone. However, prophylactic use of antidysrhythmic drugs in patients with asymptomatic disease is not recommended. In some groups (e.g., after MI), certain antidysrhythmics have been linked to higher mortality.

Ventricular tachycardia. Ventricular tachycardia consists of three or more consecutive ventricular complexes at a rate greater than 100 beats per minute (Fig. 19.30). The rhythm is fairly regular, and the complexes generally have the same configuration (monomorphic). With rapid rates, it may be difficult to distinguish the QRS complexes from the ST segments and T waves, and the ECG depicts a series of large, wide, undulating waves. The sinus node usually continues to discharge independently of the ventricular rhythm, and P waves, if seen, are not associated with the QRS complexes.

Reentry is the probable mechanism of ventricular tachycardia in most cases, although automaticity and triggered activity have also been implicated. Ventricular tachycardia is often associated with myocardial ischemia and infarction. Damage to the myocardium alters conduction times and conduction pathways, which sets the stage for reentry loops. High catecholamine levels and an abnormal electrolyte balance may contribute to the dysrhythmogenesis.

Ventricular tachycardia is a serious dysrhythmia that is nearly always indicative of significant heart disease. It may be fatal unless it is successfully and rapidly managed. Ventricular tachycardia may compromise cardiac output, resulting in loss of consciousness. Treatment consists of administration of antidysrhythmic drugs and, if necessary, cardiopulmonary resuscitation and electrical cardioversion.

Ventricular fibrillation. Ventricular fibrillation is a rapid, uncoordinated cardiac rhythm that results in ventricular quivering and lack of effective contraction. The rhythm is generally easily identified, particularly when assessment of the patient indicates absence of pulse and loss of consciousness. The ECG is rapid and erratic, with no identifiable QRS complexes (Fig. 19.31). Ventricular fibrillation results in death if not reversed within minutes.

The same conditions that result in ventricular tachycardia may cause ventricular fibrillation. A critically timed premature beat or accelerating ventricular tachycardia may be the precursor to ventricular fibrillation. The ventricular depolarization is thought to be fractionated into a number of localized reentrant currents within the myocardial mass. The uncoordinated depolarizations are sustained because of variability in conduction velocities and refractory periods.

Ventricular fibrillation must be rapidly identified and managed with cardiopulmonary resuscitation and defibrillation with electrical current. Defibrillation differs from cardioversion in that the administration of current is not synchronized with the R wave and the amount of energy

delivered is greater (200 to 350 J). The earlier the defibrillation is performed, the better the chance for successful resuscitation. In some instances, the ventricular fibrillation pattern is very fine and is similar to the tracing seen in atrial arrest. Defibrillation is still indicated. Defibrillation and cardiopulmonary resuscitation are usually followed by administration of antidysrhythmic drugs.

KEY POINTS

- Failure of the sinoatrial (SA) node to generate impulses may result in a junctional or ventricular escape rhythm. These rhythms are slow and may be poorly tolerated. Absence of P waves is important in determination of escape rhythms.
- In most cases, premature beats and ectopic rhythms are attributed to reentry mechanisms. Reentry circuits may be established when portions of the heart have abnormal conduction rates or pathways. Enhanced automaticity and triggered activity are alternative mechanisms for generation of ectopic complexes.
- Atrial dysrhythmias include Premature atrial complexes (PACs), tachycardia, flutter, and fibrillation. Atrial dysrhythmias are usually well tolerated unless the ventricular response rate is significantly altered.
- Junctional tachycardias are difficult to distinguish from atrial tachycardias, and they are often regarded together as supraventricular tachycardias.
- Frequent premature ventricular complexes (PVCs), ventricular tachycardia, and ventricular fibrillation are associated with a significant fall in cardiac output and must be rapidly diagnosed and managed.

Conduction Pathway Disturbances

Disorders of cardiac impulse conduction include delays, blocks, and abnormal pathways. Cardiac ischemia and infarction commonly are associated with conduction blocks and delays, whereas abnormal pathways are usually congenital.

Disturbances of Atrioventricular Conduction

A disturbance in conduction between the sinus impulse and its associated ventricular response has been called *atrioventricular block*. The conduction may be abnormally slowed or completely blocked. The AV block results from a functional or pathologic defect in the AV node, bundle of His, or bundle branches. Three categories of AV block have traditionally been described: first-degree block, second-degree block (which includes types I and II), and third-degree (complete) block. These AV conduction disorders are associated with different pathologic processes and clinical implications.

First-degree block is generally identified by a prolonged PR interval (more than 0.20 second) on the ECG (Fig. 19.32). The rhythm remains regular, and each P wave is associated with a QRS complex. First-degree block is a common finding and may occur in the absence of organic heart disease. Drugs and organic heart disorders, such as myocardial ischemia and congenital heart defects, may cause first-degree block. First-degree block is generally monitored but is not actively managed except to alleviate the underlying cause if possible.

Second-degree block is diagnosed when some of the atrial impulses are not conducted to the ventricles. Two types of second-degree block are identified by the pattern of nonconducted impulses. Type I (Mobitz type I, Wenckebach) is associated with progressively lengthening PR intervals until one P wave is not conducted (dropped beat). The pattern repeats, causing the QRS complexes to occur in groups. The PP intervals are constant, whereas the RR intervals vary (Fig. 19.33). Type I second-degree block is usually due to reversible ischemia of the AV node, often associated with acute MI. The ischemic node is slow to recover after each depolarization, resulting in a progressively longer nodal delay until one impulse is not conducted. This gives the AV node time to recover, and the next atrial impulse is conducted more quickly, with a nearly normal PR interval, beginning the cycle again. Treatment is rarely required. If the block progresses to a type II block, a pacemaker may be required.

Type II second-degree block is identified by the presence of nonconducted P waves (dropped beats) with a consistent PR interval (Fig. 19.34). The QRS complex is usually, but not always, wide (0.12 second or greater). Type II block is generally associated with pathologic lesion of the bundle of His, the right bundle branch, or both. It is the bundle branch block that causes the QRS complexes to be abnormally wide. Type II second-degree block is less common than type I but is more serious. It is usually associated with anterior septal MI or fibrosis of the conduction system. Type II block may progress to complete heart block with slow ventricular escape rhythm and poor cardiac output. Type II block may also result in severe bradycardia because of the number of dropped beats. Symptomatic type II block may require implantation of a pacemaker.

Third-degree block may occur as a result of a pathologic lesion of the AV node, bundle of His, or bundle branches. No impulses are conducted from the atria to the ventricles, and a junctional or ventricular

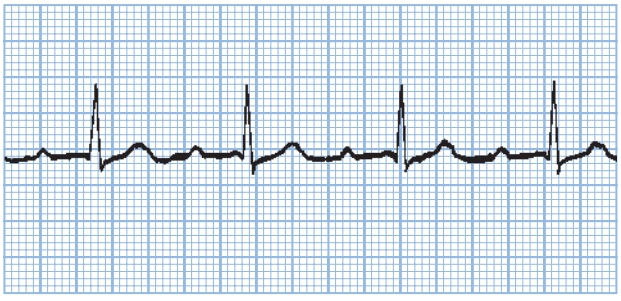


FIG 19.32 First-degree atrioventricular block. PR interval, 0.32 second.

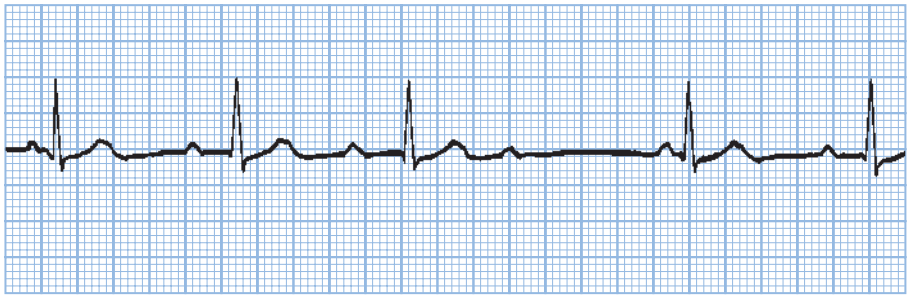


FIG 19.33 Second-degree atrioventricular block, type I (Wenckebach, Mobitz type I). Note the progressive lengthening of the PR interval until one P wave is not conducted (dropped).

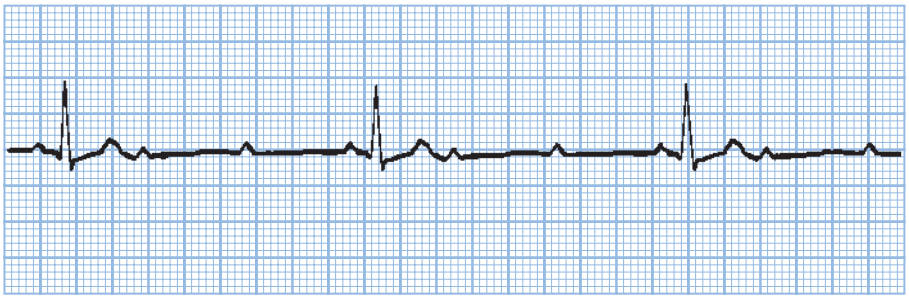


FIG 19.34 Second-degree atrioventricular (AV) block, type II (Mobitz type II). Every third P wave is followed by a QRS complex. The other P waves are not conducted through the AV node. The PR interval on conducted impulses is constant.

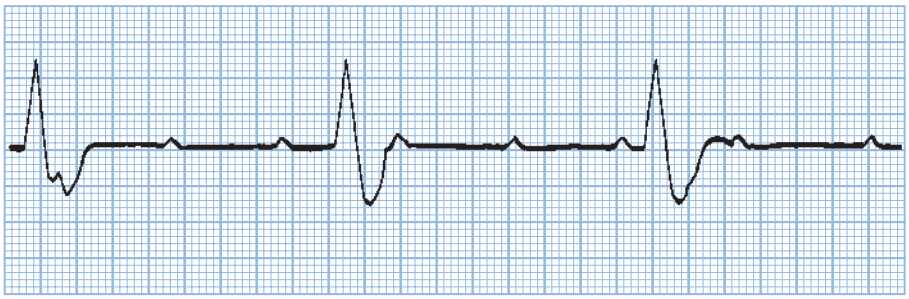


FIG 19.35 Complete third-degree atrioventricular block. Note that there is no relationship between P waves and QRS complexes because the atria and ventricles are depolarizing independently.

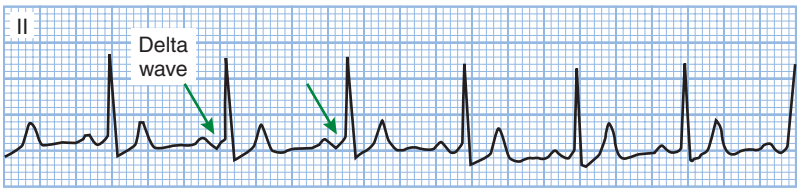


FIG 19.36 Electrocardiogram in lead II from a patient with Wolff–Parkinson–White syndrome. Note the slurred upstroke of the R wave (delta wave).

escape rhythm is evident. The ECG shows regularly occurring P waves that are totally independent of the ventricular rhythm (Fig. 19.35). If the QRS complex is narrow, the block is most likely in the AV node, proximal to the bundle of His. A prolonged QRS interval (more than 0.12 second) indicates pathology distal to the bundle of His, within the bundle branches. The severity of symptoms is determined primarily by the heart rate, with slower rhythms being more serious. A pacemaker is generally required.

Abnormal Conduction Pathways

Some individuals have congenital abnormalities of the cardiac conduction system called *accessory pathways*. These extra conduction tracts provide alternative pathways for depolarization of the heart, resulting in abnormally early ventricular depolarizations following atrial depolarizations. The best known of these preexcitation syndromes is Wolff–Parkinson–White syndrome. This syndrome is caused by accessory pathways that originate in the atria, bypass the AV node, and enter a site in the ventricular myocardium. This results in more rapid activation of the ventricle, a short PR interval, initial slurring of the QRS (δ wave), and a wide QRS complex (Fig. 19.36). The accessory pathway may provide a mechanism for reentry and the development of supraventricular tachycardia. Identification and treatment of individuals with preexcitation syndromes are desirable to prevent symptoms of supraventricular tachycardia and to reduce the possibility of deterioration of the rhythm to atrial or ventricular fibrillation. Antidysrhythmic agents and measures to interrupt the pathway, such as vagal stimulation or ablation, may be used.

Intraventricular Conduction Defects

Abnormal conduction of impulses through the intraventricular bundle branches is called *bundle branch block*. The two primary bundles are the right bundle branch, which supplies the right ventricle, and the left bundle branch, which supplies the left ventricle. The left bundle branch is further divided into three fascicles: anterior, posterior, and septal (Fig. 19.37). These supply the anterior, posterior, and septal portions of the left ventricle, respectively. Slowed or obstructed conduction occurring in one or more of these bundles results in abnormal ventricular depolarization and wide, bizarre-appearing QRS complexes. Bundle branch blocks are best detected with ECG leads V_1 and V_6.

Right bundle branch block may be present in almost any form of heart disease. It is occasionally found in individuals with no clinical evidence of heart disease. Right bundle branch block can progress to complete heart block in some cases. The electrocardiographic pattern is indicative of blocked conduction to the right ventricle such that the left ventricle depolarizes first, then spreads to the right ventricle. Right bundle branch block is classically associated with a late R wave in lead V_1 and an S wave in V_6. These changes are compared with the normal V_1 and V_6 in Fig. 19.38.

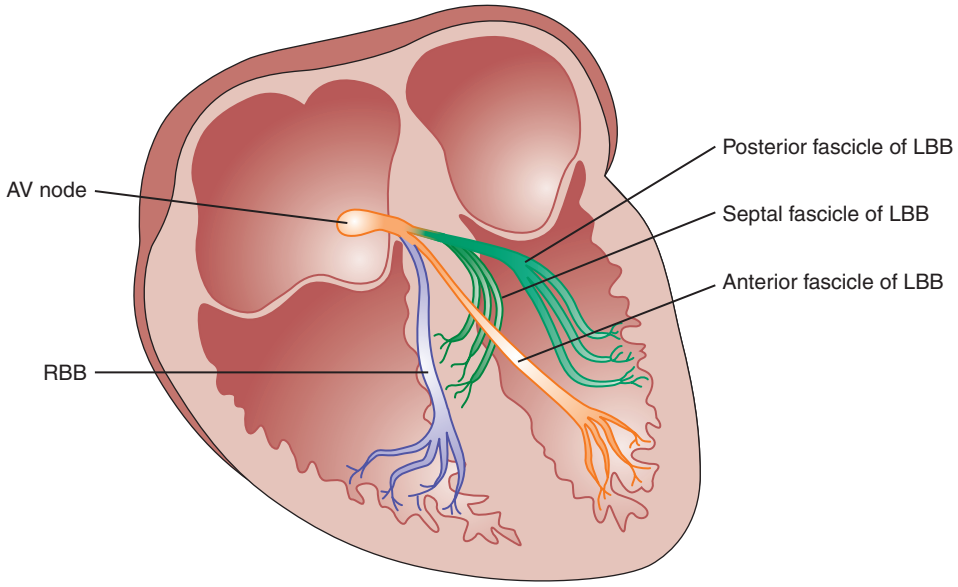


FIG 19.37 The right bundle branch (RBB) innervates the right ventricle. The left bundle branch (LBB) has three divisions: the posterior, septal, and anterior fascicles. *AV,* Atrioventricular.

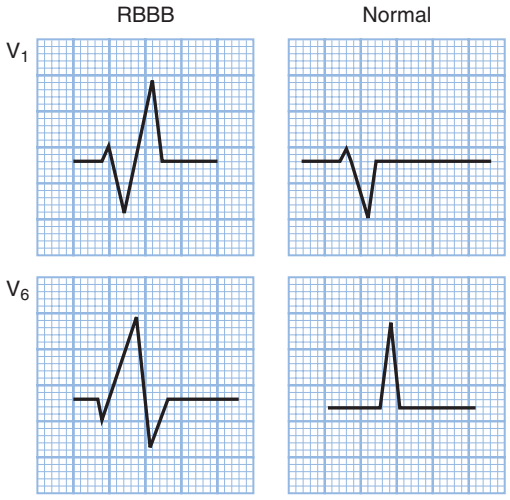


FIG 19.38 Right bundle branch block pattern. Note late R wave in V_1 and abnormal S wave in V_6.

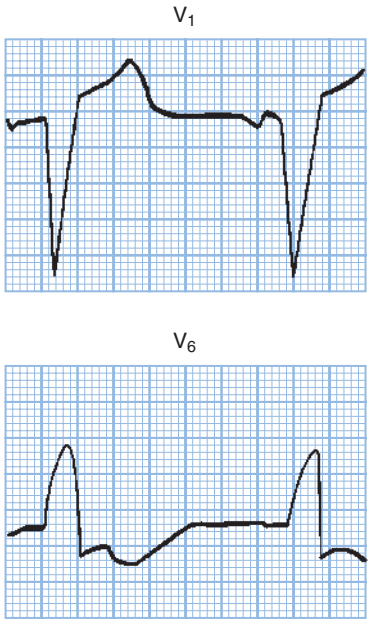


FIG 19.39 Left bundle branch block pattern. Note wide S wave in V_1 and wide R wave in V_6.

Left bundle branch block causes a delay in left ventricular depolarization. The right ventricle is activated first through the right bundle branch, followed by right-to-left activation of the septum and, finally, left ventricular activation. The QRS complex is abnormally wide (more than 0.12 second) but has a nearly normal deflection pattern in V_1 and V_6. In V_1, the small R wave normally associated with septal depolarization is absent, and V_6 consists of a wide R wave (Fig. 19.39).

Left anterior fascicular block is also called *anterior hemiblock.* Impaired conduction in the anterior fascicle causes the posterior aspect of the left ventricle to be activated first, followed by spread through the left ventricular myocardium in an upward and leftward direction. The electrocardiographic pattern shows small initial R waves followed by large S waves in leads II and III. The duration of the QRS complex is within normal limits.

Left posterior fascicular block (hemiblock) is due to a block in the posterior fascicle of the left bundle, which causes the anterior left ventricle to be activated first, followed by spread in a downward and rightward direction. The electrocardiographic findings include Q wave in leads II, III, and aV_F and R wave in leads I and aV_L. These electrocardiographic findings may mimic ventricular hypertrophy or inferolateral MI, making recognition difficult.

Slowed or obstructed conduction may occur simultaneously in more than one bundle or fascicle, leading to the terms *bifascicular block* and *trifascicular block.* For example, a right bundle branch block occurring in conjunction with a left posterior hemiblock is called a *bilateral* or *bifascicular block.* Trifascicular block refers to a bifascicular bundle block (most commonly right bundle branch block with left anterior hemiblock) in addition to a first-degree block (prolonged PR interval). The prolonged

PR interval is usually due to incomplete block in the left posterior fascicle. Complete trifascicular block would make it impossible for a supraventricular depolarization to activate the ventricles and would be a third-degree or complete heart block.

KEY POINTS

- Disturbances of atrioventricular (AV) conduction are generally referred to as *AV blocks*. First-degree block is characterized by a prolonged PR interval and usually requires no treatment.
- Two types of second-degree block have been identified. Type I (Wenckebach) is characterized by progressive prolongation of the PR interval until one P wave is not conducted. Type I block is associated with AV nodal ischemia. Type II second-degree block is identified by a rhythm showing a consistent PR interval with some nonconducted P waves. This block is more serious because it has a tendency to progress to complete AV (third-degree) block.
- Third-degree or complete heart block is diagnosed when there is no apparent association between atrial and ventricular conduction. This rhythm is serious because it is typically associated with slow ventricular rhythm and poor cardiac output.
- Accessory conduction pathways are suspected in persons exhibiting preexcitation syndromes such as Wolff–Parkinson–White syndrome. Severe tachycardias and other reentrant rhythms may occur.
- Disturbances of intraventricular conduction (bundle branch blocks) are characterized by wide, bizarre-looking QRS complexes. Any of the three ventricular fascicles may be affected (right bundle, left anterior fascicle, or left posterior fascicle).

Treatment

Dysrhythmias are generally treated if they produce significant symptoms or are expected to progress to a more serious level. A number of antidysrhythmic drugs have proved effective in managing many dysrhythmias; however, most have also been shown to cause dysrhythmias (prodysrhythmic). These drugs alter the properties of ion movement across cardiac membranes and affect automaticity as well as the rate and duration of depolarization and repolarization. The major electrophysiologic classes of antidysrhythmic (antiarrhythmic) compounds are summarized in Table 19.3. Treatment may also include measures to improve cardiac output, including pacemakers and drugs to improve contractility and blood pressure. Dysrhythmias causing severely reduced cardiac output, such as severe bradycardia, asystole, ventricular tachycardia, and ventricular fibrillation, require cardiopulmonary resuscitation until an effective cardiac rhythm is established.

Ablation procedures may be effective in eliminating a focus of dysrhythmia generation if one can be identified. An electrophysiologic study is done to evoke and analyze the dysrhythmia, followed by interruption (ablation) of the area generating it. Ablation is accomplished with high-frequency radio waves (radioablation) or by surgical excision. The electrophysiologic study requires insertion of electrodes directly into the heart by way of a venous or arterial catheter. The electrodes are used to record activity in specific locations and to deliver electric shocks to initiate or terminate an abnormal rhythm. This test is useful in assessing responses to drug therapy and in identifying risk for sudden cardiac arrest. Those at high risk may benefit from insertion of implantable defibrillators that detect lethal rhythms and apply an electric shock to convert the rhythm.

SUMMARY

HF may result from a number of cardiac and noncardiac disorders that diminish myocardial contractility or impose an excessive workload on the heart. HF is a clinical diagnosis based on characteristic signs and symptoms. Decreased cardiac output to the tissues results in decreased renal blood flow, fluid retention, activity intolerance, and mental fatigue. Backward effects are due to congestion of blood behind the ineffectively pumping ventricle. With left-sided HF, the congestion is located in the lungs and produces a number of signs and symptoms, including dyspnea, orthopnea, hypoxemia, crackles, and frank pulmonary edema. Isolated right-sided HF causes congestion in the systemic venous system leading to congestion and dysfunction of the liver, spleen, and kidney, as well as peripheral subcutaneous edema and distended neck veins. In the early stages HF may be asymptomatic, and attention to risk factors and early structural abnormalities of the heart may allow early intervention. Two types of HF have been described based on EF. Those with low EF are commonly described as having systolic failure. Those with preserved EF, typically defined as greater than 50%, usually have diastolic failure. Many patients with low EF have both systolic and diastolic dysfunction.

Three major compensatory mechanisms operate to maintain cardiac output in the failing heart: (1) sympathetic activation, (2) increased preload, and (3) cardiac muscle cell hypertrophy. Unfortunately, these mechanisms also increase myocardial workload and oxygen requirements and may trigger neurohormonal dysregulation and ventricular remodeling associated with HF progression and decompensation. Progression of HF is related to myocardial remodeling characterized by myocyte loss and myocardial fibrosis. Therapies that slow the remodeling process may slow the progression of HF. The primary aims of therapy are to improve cardiac output, minimize congestive symptoms and cardiac workload, and slow the detrimental remodeling process.

Dysrhythmia refers to an abnormality of electrical impulse generation or conduction. Dysrhythmias may occur in association with a number of cardiac and noncardiac disorders. Disturbances in electrical activity of the heart can indicate underlying pathophysiologic processes but are not themselves primary medical diseases. Dysrhythmias are significant because they can signal underlying pathophysiologic disorders and can disrupt normal cardiac output. Treatment for dysrhythmias centers on maintaining adequate cardiac output, providing antidysrhythmic drugs as needed, and diagnosing and managing the underlying pathologic process.

TABLE 19.3 Classification of Drug Actions Based on Modification of Vulnerable Parameter

Mechanism	Dysrhythmia	Vulnerable Parameter (Effect)	Drugs (Effect)
Automaticity			
Enhanced normal	Inappropriate sinus tachycardia	Phase 4 depolarization (decrease)	β-Adrenergic blocking agents
	Some idiopathic ventricular tachycardias		Na^+ channel blocking agents
Abnormal	Atrial tachycardia	Maximum diastolic potential (hyperpolarization)	Muscarinic receptor subtype 2 (M_2) agonists
		Phase 4 depolarization (decrease)	Ca^{2+} or Na^+ channel blocking agents; M_2 agonists
	Accelerated idioventricular rhythms	Phase 4 depolarization (decrease)	Ca^{2+} or Na^+ channel blocking agents
Triggered Activity			
Early afterdepolarization (EAD)	Torsades de pointes	Action potential duration (shorten)	β-Adrenergic agonists, vagolytic agents (increase rate)
		EAD (suppress)	Ca^{2+} channel blocking agents; Mg^{2+}; β-Adrenergic blocking agents
Delayed afterdepolarization (DAD)	Digitalis-induced dysrhythmias	Calcium overload (unload)	Ca^{2+} channel blocking agents
		DAD (suppress)	Na^+ channel blocking agents
	Right ventricular outflow tract ventricular tachycardia	Calcium overload (unload)	β-Adrenergic blocking agents
		DAD (suppress)	Ca^{2+} channel blocking agents; adenosine
Na^+ Channel–Dependent Reentry			
Long excitable gap	Typical atrial flutter	Conduction and excitability (depress)	Types IA, IC Na^+ channel blocking agents
	Circus movement tachycardia in Wolff–Parkinson–White syndrome (WPW)	Conduction and excitability (depress)	Types IA, IC Na^+ channel blocking agents
	Sustained uniform ventricular tachycardia	Conduction and excitability (depress)	Na^+ channel blocking agents
Short excitable gap	Atypical atrial flutter	Refractory period (prolong)	K^+ channel blocking agents
	Atrial fibrillation	Refractory period (prolong)	K^+ channel blocking agents
	Circus movement tachycardia in WPW	Refractory period (prolong)	Amiodarone, sotalol
	Polymorphic and uniform ventricular tachycardia	Refractory period (prolong)	Type IA Na^+ channel blocking agents
	Bundle branch reentry	Refractory period (prolong)	Type IA Na^+ channel blocking agents; amiodarone
	Ventricular fibrillation	Refractory period (prolong)	K^+ channel blocking agents
Ca^{2+} Channel–Dependent Reentry			
	Atrioventricular nodal reentrant tachycardia	Conduction and excitability (depress)	Ca^{2+} channel blocking agents
	Circus movement tachycardia in WPW	Conduction and excitability (depress)	Ca^{2+} channel blocking agents
	Verapamil-sensitive ventricular tachycardia	Conduction and excitability (depress)	Ca^{2+} channel blocking agents

From the Task Force for the Working Group on Arrhythmias of the European Society of Cardiology: The Sicilian gambit: a new approach to the classification of antiarrhythmic drugs based on their actions on arrhythmogenic mechanisms, *Circulation* 1991;84:1831. Copyright © 1991, American Heart Association.

RESOURCES

Heart Failure

American Heart Association: *Heart disease and stroke statistics—2017 update*, Dallas, TX, 2017, The Association.

The Digitalis Investigation Group: The effect of digoxin on mortality and morbidity in patients with heart failure. *N Engl J Med* 336(8):525–533, 1997.

Kim J, Jacobs DR, Jr, Luepker RV, et al: Prognostic value of a novel classification scheme for heart failure: the Minnesota Heart Failure Criteria. *Am J Epidemiol* 164(2):184–193, 2006.

Mann D, Zipes D, Libby P, Bonow R, editors: *Braunwald's heart disease: a textbook of cardiovascular medicine*, ed 10, Philadelphia, 2015, Elsevier.

Yancy CW, et al; Writing Committee Members; ACC/AHA Task Force Members: 2016 ACC/AHA/HFSA Focused Update on New Pharmacological Therapy for Heart Failure: An Update of the 2013

ACCF/AHA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Failure Society of America. *J Card Fail* 22(9):659–669, 2016.

Dysrhythmias

Fernandez-Falgueras A, Sarquella-Brugada G, Brugada J, et al: Cardiac channelopathies and sudden death: recent clinical and genetic advances. *Biology (Basel)* 6(1):2017. doi:10.3390/biology6010007.

Miller JM, Zipes DP: Therapy for cardiac arrhythmias. In Mann D, Zipes D, Libby P, Bonow R, editors: *Braunwald's heart disease: a textbook of cardiovascular medicine*, ed 10, Philadelphia, 2015, Elsevier, pp 685–720.

Rubart M, Zipes DP: Genesis of cardiac arrhythmias: electrophysiological considerations. In Mann D, Zipes D, Libby P, Bonow R, editors: *Braunwald's heart disease: a textbook of cardiovascular medicine*, ed 10, Philadelphia, 2015, Elsevier, pp 629–661.

20

Shock

Benjamin J. Miller

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- What is the pathogenesis of shock?
- What are the common cellular and tissue responses to shock of any cause?
- How does the body try to compensate for insufficient cardiac output during shock states?
- How do the complications of shock affect body functioning?
- What are the common causes of cardiogenic, hypovolemic, obstructive, and distributive shock?
- How do clinical and hemodynamic findings differ among types of shock?
- What is the role of the immune system in septic shock and the progressive stage of other types of shock?
- How is shock managed?
- Why does shock have high mortality?

CHAPTER OUTLINE

Pathogenesis of Shock, 434
Impaired Tissue Oxygenation, 435
Compensatory Mechanisms and Stages of Shock, 437
Types of Shock, 439
Cardiogenic Shock, 439
Etiology and Pathogenesis, 439
Clinical Manifestations, 440
Treatment, 440
Pharmacotherapy, 441
Mechanical Assist Devices, 441
Obstructive Shock, 441
Etiology and Pathogenesis, 441
Clinical Manifestations, 441
Treatment, 441
Hypovolemic Shock, 442
Etiology and Pathogenesis, 442
Classification, 442
Clinical Manifestations, 442
Treatment, 443
Distributive Shock, 443
Anaphylactic Shock, 443
Neurogenic Shock, 444
Septic Shock, 444
Assessment and Hemodynamic Monitoring, 447
Cardiac Output, 447
Arterial Oxygen Content, 447
Distribution of Blood Flow, 448
Hemodynamic Monitoring, 448
Complications of Shock, 449
Acute Respiratory Distress Syndrome, 449
Disseminated Intravascular Coagulation, 449
Acute Renal Failure, 449
Multiple Organ Dysfunction Syndrome, 449

Shock is a life-threatening condition characterized by insufficient delivery of oxygenated blood to the microcirculation, resulting in tissue hypoxia and cellular dysfunction. In 1895 John Collins Warren described shock as a momentary pause in the act of death. In spite of advances in the understanding and management of shock, it still has a high rate of mortality. This chapter presents an overview of circulatory shock, including the major causes, cellular and systemic pathogenesis, clinical manifestations, and general therapeutic management.

PATHOGENESIS OF SHOCK

Shock is characterized by an imbalance between oxygen supply and oxygen requirements at the cellular level. When the cell does not have adequate amounts of oxygen and nutrients, it is unable to meet its metabolic demands. Cellular hypoxia results in impaired cellular function and may progress to irreversible organ damage and death. The causes of circulatory shock classically are divided into four general types:

cardiogenic, obstructive, hypovolemic, and distributive. Each of these types is associated with a number of primary causes (Box 20.1). Cardiogenic shock results from heart disorders that cause inadequate cardiac output despite sufficient vascular volume. Obstructive shock develops when circulatory blockage disrupts cardiac output, such as a large pulmonary embolus or cardiac tamponade. Because the causes of obstructive shock are associated with failure of the heart to pump sufficiently, some sources include obstructive shock within the category of cardiogenic shock. Hypovolemic shock is associated with loss of blood volume as a result of hemorrhage or excessive loss of extracellular fluids, such as through vomiting, diarrhea, or excessive diuresis. Distributive shock is characterized by a greatly expanded vascular space because of inappropriate vasodilation without actual loss of vascular volume. Vasodilation leads to hypotension and altered perfusion of tissues. Anaphylactic, neurogenic, and septic are forms of distributive shock. Each type of shock has certain unique features (Table 20.1), but all are associated with impaired tissue oxygenation that can progress to refractory shock and organ failure.

BOX 20.1 Etiology of Circulatory Shock

Cardiogenic Shock
Myocardial infarction
Cardiomyopathy
Valvular heart disease
Ventricular rupture
Congenital heart defects
Papillary muscle rupture

Obstructive Shock
Pulmonary embolism
Cardiac tamponade
Tension pneumothorax
Dissecting aortic aneurysm

Hypovolemic Shock
Acute hemorrhage
Dehydration from vomiting, diarrhea
Overuse of diuretics
Burns
Pancreatitis

Distributive Shock
Anaphylaxis
Neurotrauma
Spinal cord trauma
Spinal anesthesia
Sepsis

Impaired Tissue Oxygenation

The common denominator of all forms of shock is impaired oxygen utilization by cells, which disrupts function and, if ongoing or severe, may lead to cell death, organ dysfunction, and stimulation of inflammatory reactions. Recent discoveries about the contribution of inflammatory reactions in the pathogenesis of shock have provided new insight into this complicated syndrome.

The reason for impaired oxygen utilization by cells differs with the various types of shock, but the outcomes are similar. A continuous supply of oxygen is needed by cells to allow sufficient production of energy in the form of adenosine triphosphate (ATP). Inadequate oxygen availability at the cellular level quickly impairs aerobic metabolism of glucose, fatty acids, and amino acids and causes the cells to rely on the relatively inefficient processes of glycolysis to produce cellular ATP. (A review of ATP synthesis can be found in Chapter 3.) Glycolysis is the enzymatic process of converting glucose to pyruvate, with the net production of two ATP molecules per glucose molecule. If oxygen were available, pyruvate would normally enter the mitochondria and proceed through the citric acid cycle. In the absence of cellular oxygen, the citric acid cycle is inhibited, and pyruvate accumulates in the cytoplasm. Pyruvate accumulation would quickly inhibit further glycolysis and shut down ATP production entirely if not for the conversion of pyruvate to *lactate.* Lactate diffuses from the cell and into the extracellular fluid, and accumulation of lactate in the bloodstream (more than 5 to 6 mmol/L) is considered a sign of significant tissue hypoxia.

An inadequate supply of cellular ATP inhibits energy-requiring cellular functions, including maintenance of ion concentrations across the plasma membrane. Because of their steep electrochemical gradients, extracellular sodium and calcium ions tend to leak into the cell. ATP-dependent pumps in the cell membrane are needed to continuously pump these ions back out. Failure of ion pumps leads to sodium and water accumulation in the cell (hydropic swelling) and an excess of intracellular free calcium. Intracellular calcium ions trigger a cascade of cellular events that further impair energy production and plasma membrane integrity. Cell death from oxygen deprivation takes from minutes to several hours, depending on the rate of cellular metabolic activity. However, even a short period of oxygen deprivation often sets in motion a complex cascade of events that lead to further cell damage (Fig. 20.1). Two important aspects of this cascade are (1) formation of oxygen free radicals, and (2) induction of inflammatory cytokines.

Ischemic cells may produce oxygen free radicals when oxygen supplies are restored. This process has been called *reperfusion injury.* Reactive oxygen molecules include superoxide (O_2–), peroxide (H_2O_2), hydroxyl radicals (OH–), and singlet oxygen (O). These molecules are unstable and will attack membrane structures, denature proteins, and damage DNA. Another source of oxygen free radicals is immune cells, particularly neutrophils, which are recruited to the area of tissue injury.

TABLE 20.1 Comparison of Clinical Findings in Different Types of Shock

Parameter	Cardiogenic	Obstructive	Hypovolemic	Distributive
Hypotension	Yes	Yes	Yes	Yes
Systemic vascular resistance	High	High	High	Low
Cardiac output	Low	Low	Low	High
Cardiac preload	High	High	Low	Normal to low
Venous oxygen saturation	Low	Low	Low	High
Urine output	Low	Low	Low	Low
Skin temperature	Cool	Cool	Cool	Warm

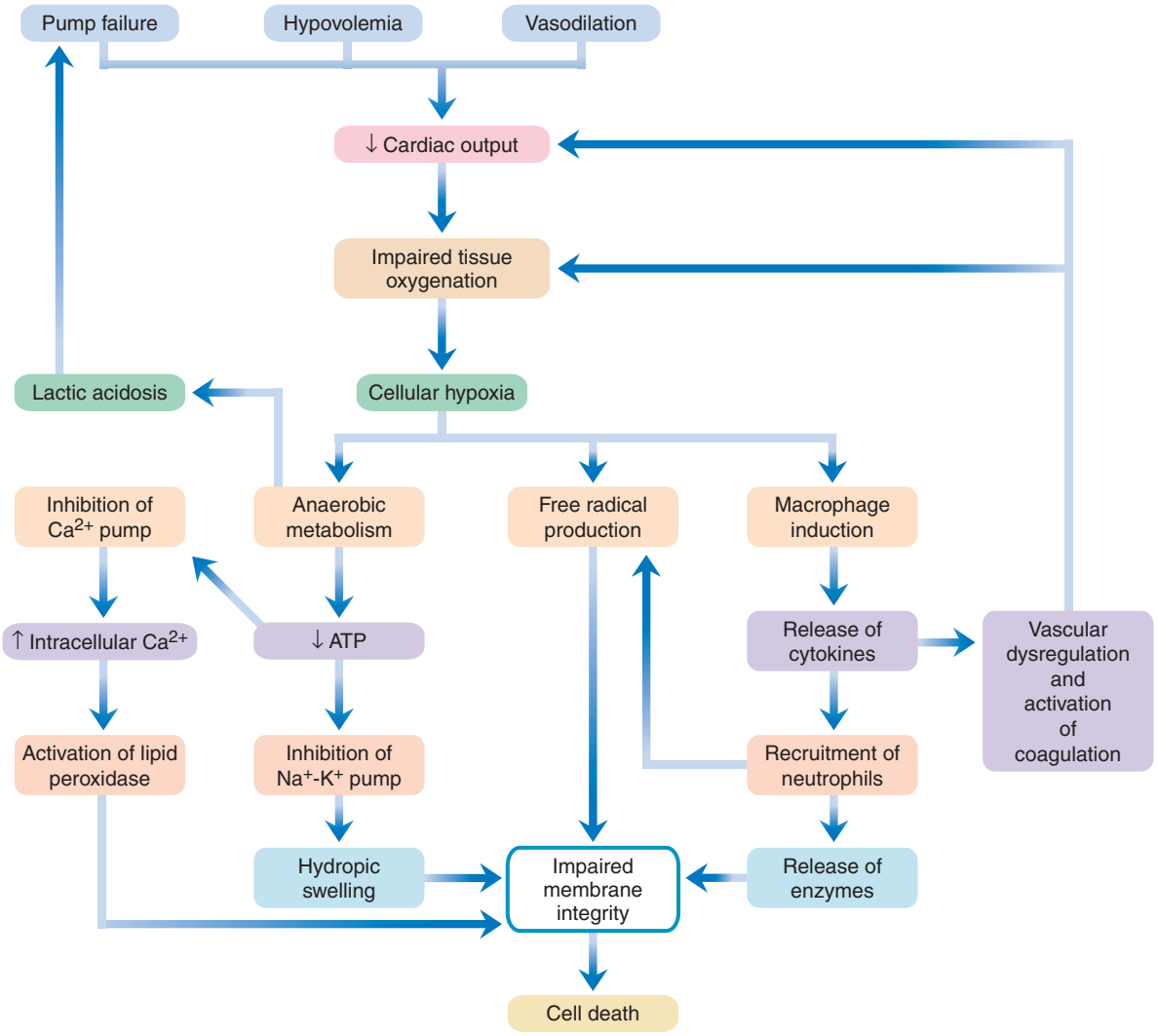


FIG 20.1 Shock is a complex process involving cellular hypoxia, free radical formation, and systemic inflammation. All forms of shock are associated with impaired tissue oxygenation, which triggers a cascade of events leading to tissue injury and death. *ATP,* Adenosine triphosphate.

Thus cellular injury may continue and progress long after the initial hypoxic insult has been resolved.

The role that immune cytokines play in shock has been studied extensively in septic shock. The roles of these cytokines are thought to be similar in the late stages of other types of shock as well. Macrophages and tissue cells are stimulated to release inflammatory cytokines in response to hypoxic tissue injury and, in the case of septic shock, in response to endotoxin or other microorganism antigens. The levels of tumor necrosis factor-α (TNF-α) and interleukin-1 (IL-1) cytokines in particular have been shown to increase in the bloodstream of patients with septic shock, and these cytokines are thought to be important mediators of vascular failure and progressive organ damage. Numerous other immune cytokines and neurohormonal mediators have been implicated in the pathogenesis of shock (Table 20.2). These mediators represent potential therapeutic targets for a disorder that is notoriously difficult to manage effectively.

A hallmark of shock is failure of the microcirculation to appropriately autoregulate blood flow. Normal tissues are able to match blood flow with metabolic needs across a wide range of blood pressures. This property ensures that blood flow is evenly distributed to tissues according to metabolic needs. In shock, autoregulation fails, leading to an abnormal distribution of blood flow. Some capillary beds receive inadequate flow and become progressively more hypoxic, whereas other vascular routes are excessively dilated and receive too much flow. Advances in technology have allowed direct visualization of capillary flow through tissue beds and confirmed the presence of perfusion abnormalities in the microcirculation. These abnormalities differ depending on the primary cause of shock. In cases of reduced cardiac output resulting from hypovolemic or cardiogenic shock, there is a homogeneous reduction in blood flow through a given tissue's arterioles, capillaries, and venules that is directly related to the severity of cardiac output reduction. In distributive shock states, such as septic shock, the degree of microcirculatory flow is heterogeneous within a tissue, with some capillaries being closed and others open. Often there is high flow rate through some of the venules, and the degree of microcirculatory dysfunction is poorly correlated with systemic hemodynamics. Overall, this imbalance leads to a so-called *oxygen debt* in the tissues. It has been suggested that the overall oxygen debt can be estimated clinically by the serum lactate level and degree of metabolic acidosis, both of which imply a switch to anaerobic metabolism by oxygen-deprived tissues.

In septic shock, immune cytokines are believed to be at the root of the microcirculatory maldistribution problem. TNF-α, IL-1, and other

TABLE 20.2 Immune Cytokines and Neurohormones Associated With Circulatory Shock

Mediator	Associated Dysfunction
IL-1α	Inflammation, vasodilation, vascular leakiness
IL-1β	Inflammation, vasodilation, vascular leakiness
IL-6	Fever, increased acute phase protein
TNF-α	Inflammation, neutrophil activation
IL-10	Antiinflammatory, may suppress shock
TGF-β	Fibrosis, pulmonary edema
PAF	Platelet activation, chemotaxis
PAI-1	Increased clotting, thrombosis
Substance P	Proinflammatory
Chemokines	Neutrophil recruitment and binding to vessel endothelium
Nitric oxide	Vasodilator
C5a	Chemotactic
Protein C	Inhibits thrombus formation
Vasopressin	Improves vascular tone and responsiveness to NE/E
Cortisol	Antiinflammatory, improves vascular response to NE/E
Endothelin	Vasoconstriction
Adrenomedullin	Vasodilation
Norepinephrine	Vasoconstriction
Epinephrine	Positive inotropic, bronchodilation,
Leukotrienes	Inflammation, bronchospasm
Histamine	Increased vascular permeability, edema
Heparin	Inhibits action of histamine
Angiotensin II	Vasoconstriction
Heat-shock proteins	Protect protein structure and function, inhibit apoptosis

E, Epinephrine; *IL,* interleukin; *NE,* norepinephrine; *PAF,* platelet-activating factor; *PAI,* plasminogen activator inhibitor; *TGF,* transforming growth factor; *TNF,* tumor necrosis factor.

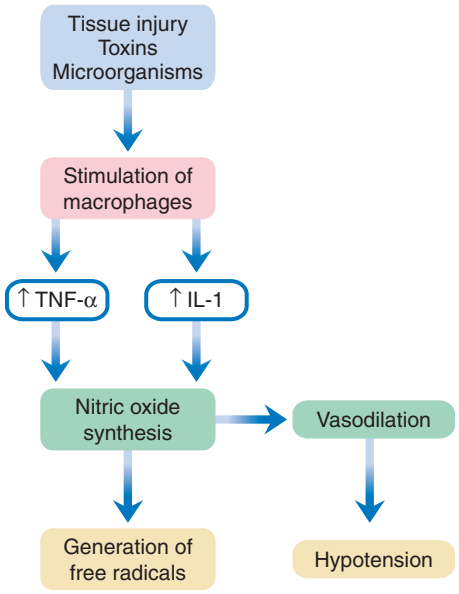


FIG 20.2 Excess production of nitric oxide is an important mechanism of vascular failure in shock. The tumor necrosis factor-α and interleukin-1 cytokines are promoters of inducible nitric oxide synthase. These cytokines are released from macrophages that have been activated by tissue injury or toxins. *IL-1,* Interleukin-1; *TNF-α,* tumor necrosis factor-α.

inflammatory mediators induce vascular cells to produce excessive amounts of the vasodilator nitric oxide. Nitric oxide in normal quantity is thought to be protective for tissues during shock, whereas excessive production is detrimental. Nitric oxide is produced in endothelial cells and vascular smooth muscle by two enzymes: nitric oxide synthase (NOS) and inducible nitric oxide synthase (iNOS). TNF-α and IL-1 increase the activity of iNOS and thereby cause excessive production of nitric oxide (Fig. 20.2). Efforts to minimize the microcirculatory oxygen debt in early shock and to quickly restore adequate microcirculatory blood flow distribution, as evidenced by normalized serum lactate concentration and acid–base balance, are an important focus of therapy for all types of shock.

Compensatory Mechanisms and Stages of Shock

A number of compensatory responses are set in motion to restore tissue perfusion and oxygenation in the early stage of shock. Historically, these responses to shock have been divided into three clinical stages: compensated shock, progressive shock, and refractory shock. Although these stages may be useful for determining prognosis and the likelihood of the patient's recovering, shock is viewed as a continuum in which compensatory mechanisms become progressively less effective as function of the microcirculation becomes increasingly impaired.

Insufficient cardiac output and decreased effective tissue perfusion are early defects in all types of shock. Insufficient cardiac output may be a consequence of an ineffective cardiac pump (cardiogenic, obstructive) or insufficient blood volume to fill the vascular space (hypovolemic, distributive). A number of compensatory mechanisms are triggered in response to inadequate cardiac output in an attempt to restore adequate perfusion pressure (Fig. 20.3). Baroreceptors located in the aorta and carotid arteries quickly sense the decrease in pressure and transmit signals to the vasomotor center in the brainstem medulla. Stimulation of the sympathetic nervous system (SNS) results in increased cardiac output and vascular resistance. Because blood pressure is determined by the product of cardiac output and vascular resistance—an increase in one or both of these factors will help restore blood pressure. The SNS increases cardiac output through several mechanisms. The adrenal medulla is stimulated to release increased amounts of the catecholamines epinephrine and norepinephrine (NE), which circulate to the heart and stimulate β_1 receptors. The β_1 receptors respond by increasing the heart rate and force of contraction in an attempt to increase cardiac output. The SNS also enhances venous return to the heart by constricting systemic arterioles and venules. Arterial vasoconstriction reduces flow through the capillary bed, which causes hydrostatic pressure in the capillaries to fall. Fluid reabsorption from interstitial spaces helps increase blood volume and improve preload. Blood vessels in the skin, kidneys, and gastrointestinal tract constrict and shunt blood to the heart and brain.

The SNS stimulates cells in the kidney to release renin, which triggers the renin–angiotensin–aldosterone system (RAAS). Renin is also secreted from the kidneys in response to decreased blood flow and pressure in the afferent arterioles. Renin triggers the formation of angiotensin II, which is a potent vasoconstrictor and also stimulates kidney nephrons to conserve sodium and water. Conservation of volume by the kidney is further enhanced by aldosterone, which is secreted from the adrenal cortex in response to angiotensin II. Reabsorption of fluid from the kidney helps increase blood volume and enhances venous return to the heart. Another hormone, antidiuretic hormone (vasopressin), is secreted from the posterior pituitary in response to reduced blood volume. Antidiuretic hormone stimulates the kidney tubules to reabsorb water and improves the vascular response to catecholamines. In shock, urine output may fall to zero as the kidneys attempt to conserve fluid to

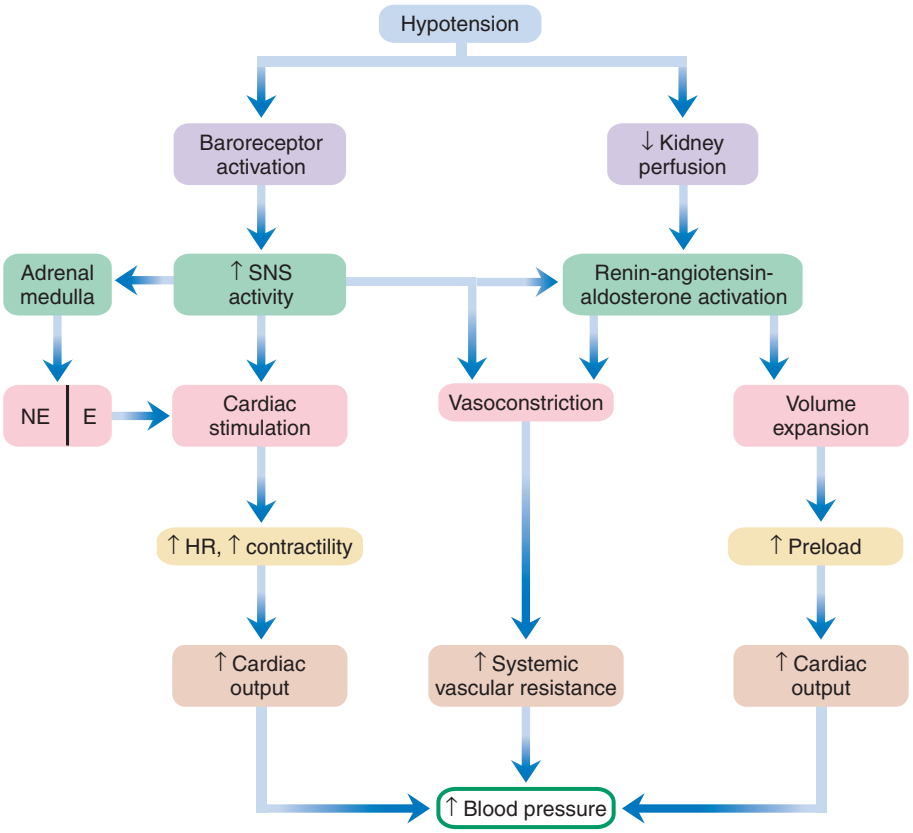


FIG 20.3 Compensatory mechanisms are triggered in shock to help maintain arterial blood pressure despite a fall in cardiac output. *E,* Epinephrine; *HR,* heart rate; *NE,* norepinephrine; *SNS,* sympathetic nervous system.

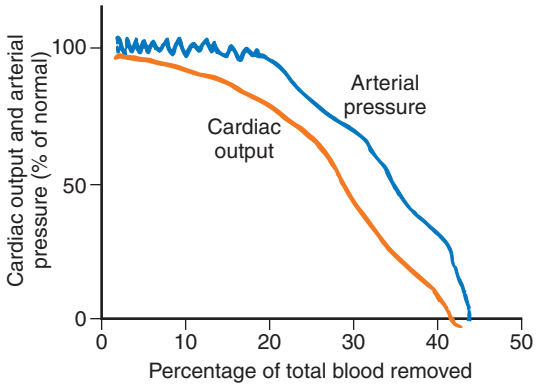


FIG 20.4 In early stages of hypovolemia, blood pressure is stable even though cardiac output is falling. When volume losses equal about 25% of the total blood volume, blood pressure falls precipitously. (Redrawn from Hall JE: *Guyton and Hall textbook of medical physiology,* ed 13, Philadelphia, 2014, Elsevier, p 294.)

maintain blood volume and cardiac output. Unfortunately, the kidney tubules often sustain damage because of the low-flow state, which may result in the complication of acute renal failure.

These compensatory mechanisms work well in the early stage of hypovolemic shock and may maintain blood pressure within the normal range until the volume of blood loss becomes too great (Fig. 20.4). In other forms of shock, compensatory mechanisms are less effective in restoring cardiac output. In cardiogenic shock, the compensatory responses may worsen the already high preload and impose a greater workload on the failing heart. In distributive shock, the vasculature is not responsive to SNS signals to constrict. Blood pools in the peripheral tissues, which makes it difficult for the heart to maintain cardiac output despite sympathetic stimulation to increase the heart rate and contractility.

The early, compensated stage of shock may be difficult to detect clinically (Fig. 20.5). A high index of suspicion is needed in patients with heart failure, trauma, blood loss, and severe infection. In addition, the following clinical findings may be present:

- A narrow pulse pressure, with or without hypotension
- Tachycardia greater than 100 beats/min
- Fast and deep respirations
- Decreased urinary output
- Increased urine-specific gravity
- Cool, clammy skin
- Altered mentation
- Dilated pupils

At some point, which is highly variable and differs among individuals depending on age, comorbidities, and specific etiology, the compensatory mechanisms can no longer sustain adequate perfusion to tissues, and cells begin to suffer significant hypoxic injury. This condition is sometimes called the *progressive stage of shock.* Active therapeutic intervention is required at this stage or the patient will probably not survive. As previously described, reduced delivery of oxygen to tissues results in hypoxic injury, free radical damage, and stimulation of the inflammatory response. Lactic acidosis may occur during the progressive stage of shock. In addition to being a marker of anaerobic metabolism, lactate can alter the acid–base balance of the blood and create metabolic acidosis. Metabolic acidosis places a greater burden on the respiratory and renal systems and may contribute to further dysfunction. Metabolic acidosis can affect electrolyte balance and contribute to cardiac dysrhythmias and conduction disturbances. In addition, myocardial-depressant factors

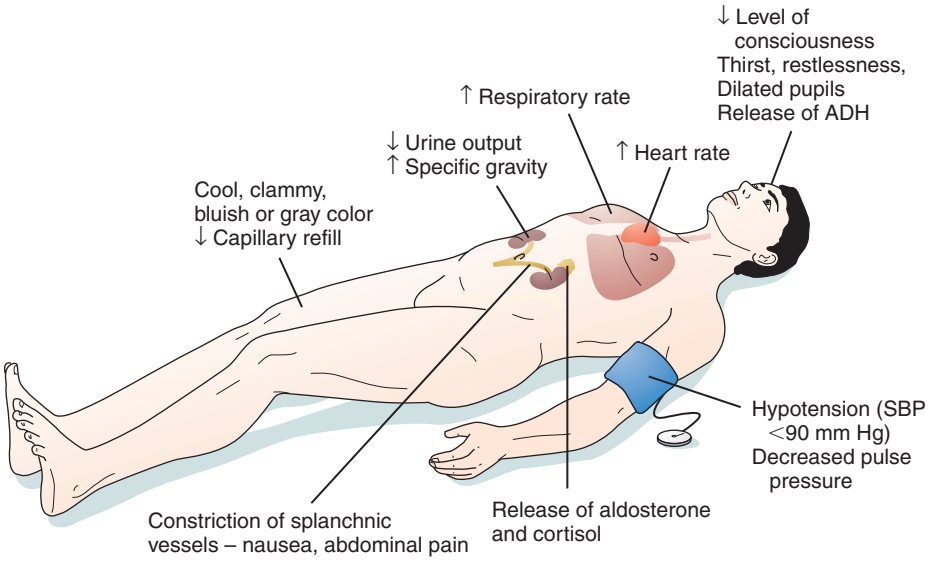


FIG 20.5 Classical manifestations of shock. *ADH,* Antidiuretic hormone; *SBP,* systolic blood pressure.

are released that impair myocardial contractility. These factors contribute to reduced cardiac output and a progressive cycle of worsening tissue hypoxia.

As shock continues to progress, the vascular system begins to fail. Arterioles become unresponsive to catecholamines, and previously constricted vascular beds begin to dilate. Widespread dilation and low cardiac output combine to produce severe hypotension. The low blood pressure is not sufficient for organ perfusion. At this stage the effects of shock produce more shock processes. Tissue damage often activates the clotting cascade, which contributes to sluggish blood flow, vascular thrombosis, and more severe tissue ischemia. Release of inflammatory mediators, along with vascular occlusion, may precipitate organ failure. The kidney, liver, and lung are particularly susceptible. At some point, the stage of refractory shock occurs, and the patient becomes unresponsive to therapeutic interventions.

The progressive stage of shock is characterized by the following clinical manifestations:

- Low blood pressure, usually lower than 90 mm Hg
- Narrow pulse pressure
- Tachycardia
- Acute renal failure (e.g., oliguria, increased levels of blood urea nitrogen and serum creatinine)
- Decreased level of consciousness
- Increased respiratory rates
- Metabolic and respiratory acidosis with hypoxemia

KEY POINTS

- Shock represents a diverse group of life-threatening circulatory conditions. The common factor among all types of shock is hypoperfusion and impaired cellular oxygen utilization. Inadequate cellular oxygenation may result from decreased cardiac output, maldistribution of blood flow, or reduced blood oxygen content.
- During the compensatory stage of shock, homeostatic mechanisms are sufficient to maintain adequate tissue perfusion despite a reduction in cardiac output. Manifestations of sympathetic nervous system (SNS) activation are an elevated heart rate, increased myocardial stimulation, bronchodilation, vasoconstriction, cool clammy skin, dilated pupils, and decreased urine output. Blood pressure is maintained even though cardiac output has fallen.
- During the progressive stage of shock, compensatory mechanisms begin to fail, and hypotension and progressive tissue hypoxia result. Shift of cells to anaerobic metabolism results in lactate production and metabolic acidosis. A lack of cellular adenosine triphosphate (ATP) production leads to cellular swelling, dysfunction, and death. Generation of oxygen free radicals, release of inflammatory cytokines, and activation of the clotting cascade lead to further cellular and organ dysfunction.

TYPES OF SHOCK

In addition to the general pathophysiology of shock just described, each type of shock has special features that affect prevention, diagnosis, and treatment. These features are briefly reviewed here, and the reader is referred to the specific chapters that describe the primary disorders that predispose to shock, including cardiogenic (Chapters 18 and 19), hypovolemic (Chapter 24), anaphylactic (Chapter 10), neurogenic (Chapter 44), and septic (Chapter 8).

Cardiogenic Shock

Etiology and Pathogenesis

Cardiogenic shock occurs primarily as a result of severe dysfunction of the left, right, or both ventricles that results in inadequate cardiac pumping. The most common cause of cardiogenic shock is myocardial infarction resulting in a significant dysfunction or loss (greater than 40%) of left ventricular myocardium. Other causes of cardiogenic shock include right ventricular myocardial infarction, end-stage cardiomyopathy, papillary muscle dysfunction, free wall rupture, and congenital heart defects.

The low cardiac output state is associated with a high left ventricular diastolic filling pressure (preload), a finding that differentiates cardiogenic from hypovolemic forms of shock (Fig. 20.6). High left ventricular preload leads to movement of fluid from the pulmonary vascular beds into the pulmonary interstitial space, which initially results in interstitial pulmonary edema and later in alveolar pulmonary edema.

The SNS is stimulated as a compensatory mechanism to increase cardiac output. The result is an increase in heart rate and systemic vascular resistance. High systemic vascular resistance increases the workload on the heart. Activation of the RAAS results in further increases

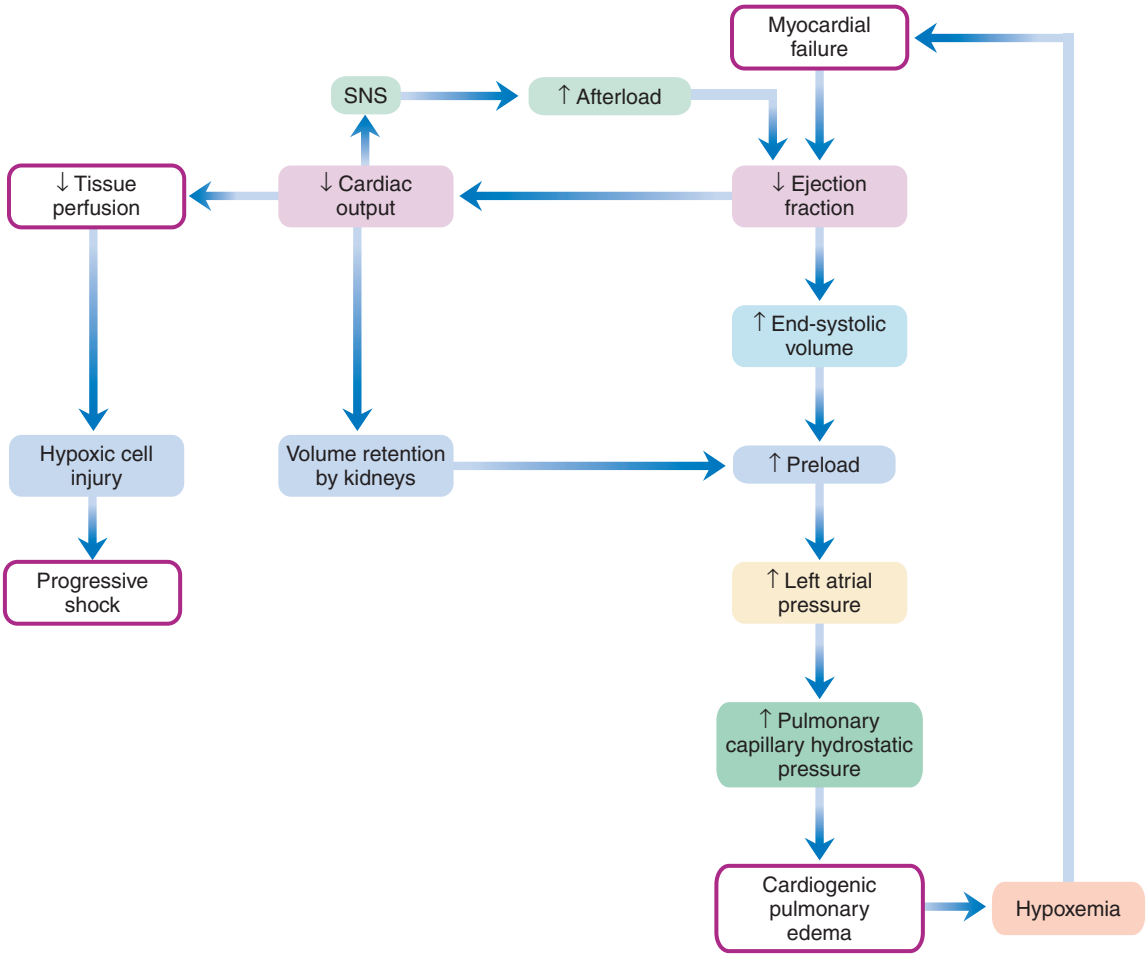


FIG 20.6 Cardiogenic shock results in decreased tissue perfusion and cardiogenic pulmonary edema because of reduced myocardial function, usually associated with left ventricular failure. *SNS,* Sympathetic nervous system.

in resistance and preload. The net effect of the activation of compensatory mechanisms is to increase myocardial workload and oxygen demand. Consequently, the compensatory responses can precipitate further cardiac damage and cause a progressive decline in cardiac output.

Clinical Manifestations

Sympathetic nervous stimulation increases the heart rate and vascular resistance, which maintain blood pressure even though cardiac output has decreased. As compensatory mechanisms fail, systolic blood pressure falls, and diastolic pressure increases (as a result of the sympathetic stimulation), thus narrowing the pulse pressure. Heart rates exceed 100 beats/min. Peripheral vasoconstriction occurs and produces cool, clammy skin. Auscultation of the lungs reveals coarse crackles resulting from pulmonary edema. An S_3 summation gallop may be audible over the left apex as a result of increased preload in the left ventricle.

Frequent assessments of cardiac output and cardiac index are helpful in the clinical treatment of a patient in cardiogenic shock. A pulmonary artery catheter may be inserted to measure cardiac index and left heart filling pressures (e.g., pulmonary capillary occlusion pressure). Pulmonary artery pressures are increased, with the pulmonary capillary occlusion pressure typically being greater than 15 mm Hg (normal, less than 12 mm Hg). When pulmonary capillary occlusion pressure acutely increases, pulmonary congestion may develop because fluid shifts from the capillary into the interstitial and alveolar spaces. Patients with chronic congestive heart failure may not develop pulmonary edema until pulmonary occlusion pressures exceed 20 to 40 mm Hg. Arterial blood gas values initially demonstrate a respiratory alkalosis secondary to hyperventilation. As pulmonary edema progresses, respiratory acidosis with hypoxemia may occur. Hypoxemia further impairs myocardial function.

Determination of mixed venous oxygen levels in blood samples obtained from the pulmonary artery catheter is helpful in assessing the adequacy of cardiac output. Decreased tissue oxygen delivery because of low cardiac output increases the degree of oxygen extraction. Mixed venous blood samples show decreased venous oxygen saturation (SvO_2) with a decreased cardiac output. An increase in mixed venous oxygenation would be expected with improved cardiac output.

Treatment

Cardiogenic shock is difficult to manage because the underlying myocardial damage is often not reversible. Prevention of cardiogenic shock through measures to limit infarct size during acute myocardial ischemia is desirable. Early efforts to restore coronary perfusion are associated with a decrease in the incidence of cardiogenic shock after myocardial infarction. (A discussion of reperfusion therapy can be found in Chapter 18.)

The goal of treatment for cardiogenic shock is to decrease myocardial oxygen demands, increase myocardial oxygen delivery, and increase cardiac output. It is difficult to achieve these goals because interventions to increase cardiac output tend to increase myocardial oxygen demands.

Pharmacotherapy

Positive inotropic drugs are frequently used in the management of cardiogenic shock to increase contractility. Positive inotropes include β-adrenergic agonists such as NE, dobutamine, and dopamine and phosphodiesterase inhibitors that prevent the degradation of cyclic adenosine monophosphate. These drugs have the ability to increase contractility, increase cardiac output, and increase tissue perfusion; however, these drugs increase myocardial oxygen demand. NE is the natural neurotransmitter of the sympathetic nerves and mimics SNS activation by increasing heart rate, contractility, and vascular resistance. Dobutamine increases contractility by stimulating β receptors. However, unlike NE or dopamine, dobutamine has minimal α-receptor activity. The major effect of dobutamine is on contractility rather than heart rate. Dobutamine may contribute to a decrease in vascular resistance and must be used with caution in hypotensive patients. Vasodilators may be used to decrease the workload of the heart by decreasing left ventricular afterload and preload. Examples of commonly used vasodilators include nitroprusside and nitroglycerin.

Mechanical Assist Devices

Cardiogenic shock is sometimes managed by mechanical assist devices. For temporary management, intraaortic balloon counterpulsation may be indicated. A catheter with a balloon at the distal segment is inserted through the femoral artery and positioned in the aorta just distal to the left subclavian artery. The balloon is connected to a console that triggers the balloon to inflate in diastole and deflate in systole. The effect of balloon inflation during diastole is to increase perfusion pressure of the coronary arteries. Sudden deflation of the balloon just before ventricular systole creates a vacuum effect in the aorta that reduces left ventricular afterload. A reduction in afterload decreases left ventricular workload and increases stroke volume. Balloon counterpulsation restricts mobility and is associated with a number of vascular complications. Long-term management of patients with low cardiac output can be achieved with mechanical pumps that take over the function of the ventricle or ventricles (ventricular assist devices [VADs]). VADs are commonly used in patients waiting for a heart transplant. In some cases, the temporary decrease in cardiac workload afforded by the VAD is associated with significant improvement in cardiac structure and function, and the device can be removed.

KEY POINTS

- Cardiogenic shock is usually a result of severe ventricular dysfunction associated with myocardial infarction. Other causes include cardiomyopathy, ventricular rupture, and congenital heart defects.
- Diagnostic features of cardiogenic shock include decreased cardiac output as a result of left ventricular dysfunction, along with elevated left ventricular end-diastolic pressure, S_3 heart sounds, and pulmonary edema. Sympathetic activation leads to an increased heart rate, vasoconstriction, and a narrow pulse pressure.
- Low cardiac output leads to reduced oxygen delivery to tissues. Tissues extract a greater percentage of oxygen from the delivered blood, which leads to reduced Svo_2.
- Therapy is aimed at improving cardiac output and myocardial oxygen delivery while reducing cardiac workload. Pharmacologic treatment often includes the use of inotropic agents, afterload-reducing agents (e.g., vasodilators), and preload-reducing agents such as nitrates and diuretics. Intraaortic balloon counterpulsation may be used to reduce afterload and improve coronary artery perfusion. Ventricular assist devices (VADs) may be used for longer-term circulatory support, whereas heart transplantation provides definitive treatment.

Obstructive Shock

Etiology and Pathogenesis

Obstructive shock develops when the heart is prevented from pumping because of a mechanical obstruction to blood flow. Impaired ventricular filling leads to reduced cardiac output and signs and symptoms similar to cardiogenic shock, and is considered by many to be a form of cardiogenic shock. Causes of mechanical obstruction include pulmonary embolism, cardiac tamponade, and tension pneumothorax. These mechanical changes cause a significant increase in the right ventricle (RV) afterload resulting in RV enlargement and left ventricle (LV) compression. With the decrease in LV size and stroke volume, the cardiac output declines. Increased RV preload aggravates the low cardiac output with resultant hypotension. This cycle perpetuates rapidly, resulting in death if not treated. Prompt relief of the obstruction is necessary to restore cardiac output and prevent cardiovascular collapse.

Clinical Manifestations

Obstructive shock is usually characterized by manifestations of right-sided heart failure. Depending on the location of the obstruction, elevated pressures in the cardiac chambers may be evident.

Pulmonary embolism results in elevated right-sided heart pressures, but left-sided pressures remain normal to low. Pulmonary emboli are usually generated in the veins of the lower extremities in patients with immobility, trauma, or hypercoagulable states. Pulmonary embolism is manifested as sudden, severe dyspnea and deteriorating arterial blood gas values. A perfusion scan of the lung may demonstrate an area of reduced blood flow. Pulmonary emboli are not generally detectable by chest radiographs.

Cardiac tamponade, which results from an accumulation of fluid in the pericardial sac, causes elevation of pressures on both the right side and the left side of the heart. Despite the elevated pressure, preload in the heart chambers is low, as is stroke volume. The elevated pressure is due to external compression of the heart chambers. Risks for the development of cardiac tamponade include pericarditis, blunt trauma to the chest, and cardiac surgical procedures. In pericarditis, a pericardial friction rub can sometimes be heard and may help with the diagnosis.

Tension pneumothorax results in shifting and compression of mediastinal structures, including the heart, which compromises left ventricular filling. Accumulation of air in the pleural space may occur because of trauma or spontaneous rupture of lung parenchyma. A tension pneumothorax develops when the air in the pleural space begins to exert a positive pressure on lung and mediastinal structures. A deviated trachea and decreased or absent breath sounds may occur. Arterial blood gas values can deteriorate rapidly. Tension pneumothorax is detectable by chest radiography.

Treatment

Management of obstructive shock is aimed at identifying and removing the offending obstruction. Compensatory mechanisms are generally ineffective in obstructive shock, and the patient's condition may deteriorate rapidly.

KEY POINTS

- Obstructive shock results from mechanical obstructions that prevent effective cardiac filling and stroke volume.
- Pulmonary embolism, cardiac tamponade, and tension pneumothorax are common causes of obstructive shock.
- Prompt management of the underlying obstruction is necessary to prevent cardiovascular collapse.

Hypovolemic Shock

Etiology and Pathogenesis

Hypovolemic shock results when circulating blood volume is inadequate to perfuse tissues. The pathogenesis of early-stage hypovolemic shock is straightforward: decreased intravascular volume leads to a decrease in venous return, which causes a decrease in cardiac output (Fig. 20.7). The decrease in cardiac output results in decreased tissue perfusion and decreased oxygen delivery.

Circulatory volume deficits may be the result of internal or external losses. Internal losses can result from internal hemorrhage, fracture of long bones, or leakage of fluid into the interstitial spaces. External losses can result from external hemorrhage, burns, severe vomiting and diarrhea, or diuresis. External hemorrhage is the most common cause of hypovolemic shock. Other forms of shock (septic and anaphylactic) have characteristics similar to hypovolemic shock. Despite having an absolute loss of intravascular volume, increased capillary permeability results in intravascular volume moving into the interstitial space, resulting in a relative hypovolemia

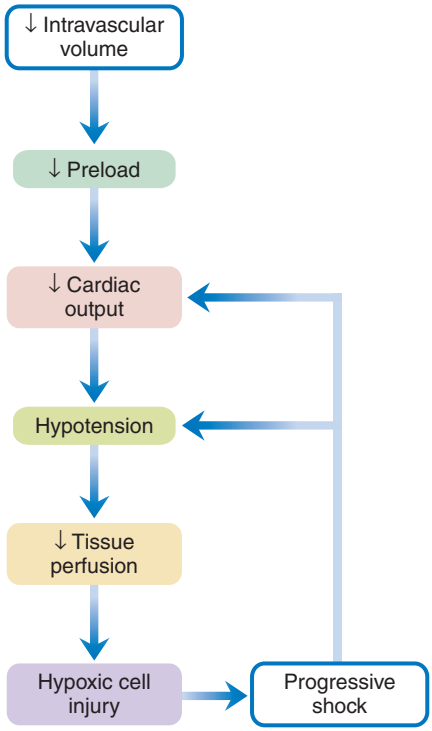


FIG 20.7 Pathogenesis of hypovolemic shock.

Classification

The American College of Surgeons stratifies hemorrhagic shock into four classes according to the degree of blood volume lost (Table 20.3). These values are based on a 70-kg adult and are guidelines that may not apply to all patients with hemorrhage, depending on etiology, rate of blood loss, and comorbidities.

The initial stage hemorrhage (Class I) occurs with blood loss up to 750 mL, or 15% of total blood volume. Compensatory mechanisms maintain cardiac output, and the patient's vital signs remain within the normal range. Class II compensated hemorrhage is categorized as blood loss between 750 and 1500 mL (15% to 30% of total blood volume). The patient becomes anxious and restless. Blood pressure remains normal when the patient is supine but decreases upon standing (orthostatic hypotension). The heart rate is between 100 and 120 beats/min. The respiratory rate is normal to mildly increased. Urine output is between 20 and 30 mL/hr. The capillary refill time may be prolonged. (The capillary blanch test is performed by depressing a patient's fingernail and observing how long after release the skin color takes to return to normal. Normal capillary refill times are less than 2 seconds.)

Class III hemorrhage (progressive stage) is blood loss between 30% and 40% of total blood volume (1500 to 2000 mL). The patient is anxious and confused. Blood pressure is decreased with a narrow pulse pressure. The heart rate is greater than 120 beats/min. Respiratory rates are between 30 and 40 respirations/min. Urine output is 5 to 20 mL/hr. The capillary refill test is prolonged.

Severe Class IV hemorrhage (refractory stage) occurs when more than 40% of total blood volume is lost (2000 mL or more). The patient is lethargic and has severe hypotension with a narrow pulse pressure. The heart rate usually exceeds 140 beats/min, and the respiratory rate is markedly increased. Urine output is negligible. The capillary refill test is prolonged.

The clinical features of other forms of hypovolemic shock are similar to those of hemorrhagic shock, although the volume loss has usually occurred more gradually.

Clinical Manifestations

The cardiac output and cardiac index are found to be decreased. Pulmonary artery pressures and pulmonary capillary wedge pressures are decreased because of the decreased preload. The finding of a low preload distinguishes hypovolemic shock from cardiogenic shock. In cardiogenic shock, preload is high and cardiac output is low. In hypovolemic shock, preload and cardiac output are both low.

Systemic vascular resistance is increased as a result of sympathetic activation. This increase is a compensatory mechanism in hypovolemia

TABLE 20.3 Classification of Hemorrhagic Shock

	Class I (Initial Stage)	Class II (Compensated Stage)	Class III (Progressive Stage)	Class IV (Refractory Stage)
Blood loss (mL)	≤750 (15%)	750–1500 (15%–30%)	1500–2000 (30%–40%)	>2000 (>40%)
Clinical features	Minimal tachycardia Normal or increased pulse pressure	Slight tachycardia Tachypnea Decreased pulse pressure Cool, clammy skin Delayed capillary refill Slight anxiety	Tachycardia Tachypnea Decreased systolic blood pressure Oliguria Changes in mental status such as confusion and agitation	Marked tachycardia Decreased systolic blood pressure Narrowed pulse pressure Markedly decreased urinary output Loss of consciousness Cold, pale skin

to maintain perfusion pressure. SvO_2 may be decreased because of decreased oxygen delivery and increased oxygen extraction.

Treatment

The first intervention for hemorrhagic shock is to control the source of blood loss. Second, volume losses are replaced with appropriate fluids to normalize blood pressure, cardiac output, and perfusion of the microvasculature. In severe, uncontrolled hemorrhage, efforts to increase blood pressure should be postponed until the hemorrhage is under control. Otherwise, the increased blood pressure may worsen the hemorrhage. In all types of hypovolemic shock, fluid replacement is the primary therapy. The three main types of fluid therapy agents are colloids, crystalloids, and blood products. There continues to be controversy about which type of fluid is the most appropriate for resuscitation of hypovolemic shock.

Colloids are solutions that increase the serum colloid osmotic pressure within the vascular compartment. Increased colloid pressure pulls fluid from the interstitium into the vascular space. Examples of colloid solutions are normal human serum albumin, dextran, and hetastarch. Colloids generally are not recommended for hypovolemic shock unless the patient has significant interstitial edema.

Crystalloids are solutions that contain electrolytes. Isotonic solutions such as lactated Ringer solution or normal saline solution are commonly used crystalloid solutions. Isotonic fluids are preferred over hypotonic solutions because isotonic solutions remain in the extracellular space and are more effective in increasing blood volume. Isotonic crystalloid fluids are preferred for volume resuscitation in hypovolemic shock that is not associated with severe anemia.

When significant anemia accompanies hypovolemia, *blood products* may be the treatment of choice. In hemorrhagic shock, whole blood or packed red blood cells (RBCs) with normal saline may be given to replace blood volume loss. Significant anemia may contribute to tissue hypoxia simply because of reduced oxygen-carrying capacity; however, recent research suggests that RBCs may also regulate capillary blood flow by releasing vasodilating substances, including ATP, into the microcirculation under conditions of hypoxia. Significant reduction in RBC flow through capillaries has been suggested as a mechanism for further impairing microcirculatory function.

In general, pharmacologic agents are not indicated for hypovolemia. Restoration of blood volume is essential. However, in some cases of shock, blood pressure remains low despite large amounts of fluid replacement, so vasoconstrictor agents may be used to support blood pressure.

KEY POINTS

- Hypovolemic shock results from inadequate circulating blood volume precipitated by hemorrhage, burns, dehydration, or leakage of fluid into interstitial spaces.
- The classic features of hypovolemic shock are the result of low cardiac output and low intracardiac pressures. Manifestations are due primarily to sympathetic nervous system (SNS) activation: elevated heart rate, vasoconstriction, and increased myocardial contractility.
- The severity of symptoms of hemorrhagic shock correlates with the amount of blood loss; however, there is significant variation in the clinical presentation of hypovolemic and hemorrhagic shock, and careful assessment is required to diagnose it in the early stages.
- Therapy for hypovolemic shock is aimed at fluid replacement and control of the source of volume loss. Colloids, isotonic crystalloids, and blood products may be used as replacement fluids.

Distributive Shock

Distributive shock is characterized by an abnormally expanded vascular space caused by excessive vasodilation. Vasodilation results in peripheral pooling of blood in the venous capacitance vessels and creates a relative hypovolemia. Preload and stroke volume are insufficient to maintain perfusion of the brain and tissues. Anaphylactic, neurogenic, and septic are the types of distributive shock. All are characterized by vasodilation and profound hypotension, but the cause and pathogenesis of each type differ significantly.

Anaphylactic Shock

Etiology and pathogenesis. Type I anaphylactic reactions involve an antigen/immunoglobulin E (IgE) antibody reaction on the surface of mast cells and basophils. IgE antibodies attach to receptor sites on these cells, where they await activation by specific antigens. Exposure to that antigen causes receptors on mast cells and basophils to cross-link and become activated. A host of vasoactive chemicals are released, including histamines, leukotrienes, bradykinins, and prostaglandins. These substances result in bronchoconstriction, peripheral vasodilation, and increased capillary permeability. (A detailed discussion of type I anaphylaxis can be found in Chapter 10.) In some cases, mast cell degranulation is triggered by a mechanism that does not involve IgE. These reactions may be called *anaphylactoid* (Box 20.2).

Most type I anaphylactic reactions are mild and do not result in shock. Even in more severe anaphylaxis, prompt treatment can prevent the shock syndrome. Shock occurs when peripheral dilation is massive

BOX 20.2 Common Triggers of Anaphylaxis

Anaphylactic (IgE-Dependent)

Foods
- Peanuts
- Tree nuts
- Crustaceans (crab, shrimp)

Medications
- β-Lactam antibiotics
- Other antibiotics
- Aspirin and other nonsteroidal antiinflammatory drugs

Venoms
- Bee sting
- Snake bite

Animal proteins
- Cat
- Dog
- Horse

Anaphylactoid (IgE-Independent)

Radiocontrast media
Opioids
Muscle relaxants
Temperature
- Cold
- Heat

Transfusion reactions
- IgG
- IgM

Unidentified Triggers

Idiopathic

and a type of hypovolemic shock is precipitated. In this case blood volume may be normal, but the sudden enlargement of the vascular space causes blood to pool in the periphery. In some cases, there is also significant leakage of fluid from the bloodstream into the interstitial spaces. Both of these conditions cause cardiac preload to drop, followed by a decrease in cardiac output.

Clinical manifestations. The onset of symptoms is usually within 2 to 30 minutes of exposure to the antigen; however, symptoms may not develop for several hours. The clinical presentation can vary widely in severity depending on the stimulus and rapidity of therapy. Initially the patient appears very anxious, with an increased heart rate and respiratory rate. Hypotension, urticaria (hives), pruritus (itching), and angioedema then develop. Often the patient has a sense of impending doom. Bronchoconstriction causes wheezing and cyanosis, and laryngeal edema results in hoarseness and stridor.

Prevention and treatment. Prevention of anaphylactic shock is achieved by avoidance of precipitating allergens. Anaphylactic shock is most frequently associated with antibiotic therapy, in particular β-lactams. Other common causes include other types of drugs, peanuts and tree nuts, insect stings, and snake bites (see Box 20.2). More than one third of cases are of unknown cause. Anaphylaxis is not a reportable disease, and the incidence is unknown; however, the estimated risk of occurrence is 1% to 3% per person in the United States.

Initial therapy for anaphylactic shock is directed to removing the inciting antigen if possible. Airway management and circulatory support are critical. Tracheal intubation and assisted ventilation may be needed. Bronchodilators can be used to manage bronchospasm. Epinephrine is helpful in stabilizing mast cells to prevent further release of inflammatory mediators and increasing blood pressure. Intravenous fluid therapy is used to increase intravascular volume and fill the enlarged vascular space. Increased preload will enhance cardiac output. A vasopressor may be given in an attempt to constrict the arterioles and raise blood pressure. Steroids may be given for their antiinflammatory effects, but their onset of action tends to be slow. Antihistamines may be administered to block histamine receptors, although their effectiveness is reduced once symptoms are present and the inflammatory mediators have been released. New therapeutics are being developed to target endotoxin and proinflammatory cytokines. Response to therapy for anaphylactic shock is usually rapid with a good outcome if instituted early; however, approximately 1% of anaphylactic episodes are fatal, and the majority of these are associated with antibiotics and nut allergies.

Neurogenic Shock

Neurogenic shock is often transitory. It may result from depression of the vasomotor center in the medulla or from interruption of sympathetic nerve fibers in the spinal cord. Causes of neurogenic shock include brain trauma that results in depression of the vasomotor center, spinal cord injury, high spinal anesthesia, and drug overdose.

Interruption of the neural pathway for the baroreceptor reflex results in loss of sympathetic tone in the vasculature. Profound peripheral vasodilation of both arterioles and veins occurs and leads to peripheral pooling of blood and hypotension. Decreased venous return to the heart results in decreased cardiac output and hypotension. Body position greatly influences the development of neurogenic shock. When the body is horizontal, venous return may be adequate, and cardiac output and blood pressure are sufficient. However, when an upright position is assumed, peripheral pooling from gravitational effects causes a severe drop in cardiac output and blood pressure. Syncope and fainting will follow unless measures are taken to redistribute the blood. Elevation of the legs, slow position changes, and the use of pressure stockings on the legs may help prevent peripheral pooling. Vasoconstricting drugs and fluid expansion may sometimes be used to increase blood pressure if mechanical measures are ineffective. The outcome of neurogenic shock depends in large part on the severity of neurologic injury. Shock resulting from head trauma has a poor prognosis, whereas acute neurogenic shock associated with spinal cord injuries may resolve as spinal cord reflexes return in the weeks following the injury.

Septic Shock

Septic shock is a common cause of death in intensive care units in the United States, and the incidence continues to increase. The latest annual prevalence estimates are that between 0.4 and 1.0/1000 of the population will develop severe sepsis. Large numbers of immunocompromised individuals in the population and extensive use of invasive technology contribute to the high rates of septic shock. The mortality associated with septic shock averages between 30% and 50% in various clinical trial registries. High mortality is in part because of the underlying diseases that often accompany sepsis, such as trauma, peritonitis, cancer, and immunodeficiency diseases.

Etiology. Sepsis results from an inappropriate host response to the presence of pathogens. Noninfectious causes of systemic inflammatory response include pancreatitis, tissue ischemia, trauma, and surgical tissue injury. Infectious causes are the direct result of bacteria, fungi, and viruses. Most cases of bacteremia do not result in shock, and the body's defense systems effectively destroy the bacteria, preventing widespread dissemination of the infection. Immunocompromised individuals are prone to disseminated infections called *bacteremia* (the term *septicemia* is still in clinical use, but is discouraged because of imprecise meaning). When the body's response to infection or other insults results in systemic signs and symptoms of widespread inflammation, the term *systemic inflammatory response syndrome* (SIRS) is applied. Septic shock is a severe systemic inflammatory reaction to infection that results in abnormal vasodilation, hypotension, and tissue hypoxia attributable to the maldistribution of blood flow. In 2016 the European Society of Intensive Care Medicine and the Society of Critical Care Medicine published new definitions of sepsis and the related clinical criteria (Table 20.4).

Patients at high risk for septic shock include the very young and the elderly. Patients in these age groups are less likely to be able to destroy invading microorganisms. Patients who are debilitated, malnourished, or immunocompromised by acquired immunodeficiency syndrome or chemotherapy or have chronic health problems are also at increased risk. Medical interventions that predispose a patient to septic shock include the use of invasive lines, catheters, and procedures; surgery; and immunosuppressive therapy.

Pathogenesis. Septic shock commonly is associated with gram-negative infections. Gram-negative bacteria include *Escherichia coli, Klebsiella pneumoniae, Enterobacter aerogenes, Serratia marcescens, Pseudomonas aeruginosa,* and *Proteus* species. Gram-positive organisms *(Staphylococcus aureus, Staphylococcus epidermidis, Streptococcus pneumoniae)* and fungi (*Candida* species) are also important causes of septic shock. A frequent portal of entry is the genitourinary tract. Other entry sites include the gastrointestinal tract, the respiratory tract, and the skin.

Gram-negative bacteria have within their cell walls a lipopolysaccharide or endotoxin. The cell wall is composed of an O antigen side chain, an R core, and an inner lipid A, the toxic component of the endotoxin. Endotoxins are released into the blood during bacterial cell lysis and initiate a chain of pathophysiologic events. Macrophages are stimulated by endotoxin to release inflammatory cytokines, including TNF-α and IL-1. As previously described, TNF-α and IL-1 are thought to be major factors in the pathogenesis of septic shock because they stimulate release of more immune cytokines and the overproduction of nitric oxide.

TABLE 20.4 Definitions Related to Sepsis

Category	Old Definition
Bacteremia (fungemia)	Presence of viable bacteria (fungi) in bloodstream
Infection	Inflammatory response to invasion of normally sterile host tissue by microorganisms
Systemic inflammatory response syndrome (SIRS)	Systemic inflammatory response to a variety of clinical insults that can be infectious or noninfectious; response is manifested by two or more of the following conditions: T >38° C (100.4° F) or <36° C (96.8° F); HR >90 beats/min; RR >20 breaths/min or $Paco_2$ <32 mm Hg; WBC >12,000 cells/mm^3, <4000 cells/mm^3, or >10% immature (band) forms; positive fluid balance (>20 mL/kg over 24 hr); hyperglycemia; plasma C-reactive protein/procalcitonin >2 SD above normal value; arterial hypotension; cardiac index >3.5 L/min; arterial hypoxemia; acute oliguria; creatinine increase >0.5 mg/dL; coagulation abnormalities; ileus; platelets <100,000/μL; bilirubin >4 mg/dL; hyperlactatemia; decreased capillary refill
Sepsis	SIRS secondary to infection
Severe sepsis	Sepsis associated with one or more organ dysfunctions, hypoperfusion, or hypotension; hypoperfusion and perfusion abnormalities may include, but are not limited to, lactic acidosis, oliguria, or acute alteration in mental status
Septic shock	Sepsis with persistent hypotension despite fluid resuscitation, along with presence of perfusion abnormalities; patients taking inotropic or vasopressor agents may not be hypotensive at the time perfusion abnormalities are measured
Refractory septic shock	Persistent septic shock requiring dopamine >15 mcg/kg/min to maintain mean arterial blood pressure
Multiple organ dysfunction syndrome (MODS)	Presence of altered organ function requiring intervention to maintain homeostasis
Category*	**Additional Definitions***
Sepsis	Life-threatening organ dysfunction caused by a deregulated host response to infection
Septic shock	Severe sepsis requiring vasopressor therapy to increase mean arterial pressure to ≥65 mm Hg and a lactate level >2 mmol/L despite adequate fluid resuscitation

HR, Heart rate; *$PaCO_2$,* arterial carbon dioxide tension; *RR,* respiratory rate; *SD,* standard deviation; *T,* temperature; *WBC,* white blood cell count. Adapted from Dipiro JT et al, editors: *Pharmacotherapy: a pathophysiologic approach,* ed 8, New York, 2011, McGraw-Hill, p 2042.
*Adapted from Gotts JE, Matthay MA: Sepsis: pathophysiology and clinical management. *BMJ* 35(3):i1585, 2016.

Macrophage cytokines activate neutrophils and platelets, which release many toxic mediators such as platelet-activating factor, oxygen free radicals, and proteolytic enzymes. Activation of the arachidonic acid cascade in neutrophils and platelets results in prostaglandin, leukotriene, thromboxane, and prostacyclin release, all of which have profound effects on vascular smooth muscle. Increased levels of thromboxanes A_2 and B_2 produce pulmonary vasoconstriction, mediate bronchoconstriction, and act as potent platelet aggregators. Prostacyclin is a potent vasodilator and may contribute to the development of hypotension.

A number of other inflammatory cascades are activated in septic shock. The complement system is activated with release of C5a and C3a, and can produce microemboli and endothelial cell destruction. Histamine, a potent vasodilator, is released by mast cells. Histamine also increases capillary permeability, which enhances edema formation. The coagulation system is activated and may enhance the development of thrombi. The kinin system is activated and bradykinin is released, which results in vasodilation and increased capillary permeability. All these immune responses are normal reactions to microbial invasion and are necessary for eradicating infections. In overabundance, however, these mechanisms constitute a systemic inflammatory response that can result in shock. The major components of the complex pathophysiologic processes of septic shock are illustrated in Fig. 20.8.

Septic shock is associated with profound peripheral vasodilation. Systemic vascular resistance is decreased, and despite the increased cardiac output, blood pressure falls. The veins also dilate, and intravascular pooling occurs in the venous capacitance system. Because of maldistribution of blood flow, some portions of the tissue are underperfused and some are overperfused. Excessive flow to areas of lower metabolic demand limits oxygen extraction, which contributes to a common finding of lower overall oxygen consumption. In the initial stage of septic shock, a relative hypovolemia is present because of the increased size of the vascular compartment. Fluid administration to increase preload to a central venous pressure between 8 and 12 mm Hg is advocated at this stage, even though cardiac output may already be quite high. Cardiac output between 8 and 12 L/min is common in early septic shock. Even this level of cardiac output may be inadequate to perfuse the expanded vascular bed.

The generalized inflammatory response triggered in septic shock affects capillary permeability. Increased capillary permeability results in fluid movement out of the vascular beds into the interstitial space. Generalized soft tissue edema occurs and can interfere with tissue oxygenation and organ function.

Clinical manifestations. In contrast to other forms of shock, the clinical manifestation of early septic shock is a hyperdynamic state characterized by high cardiac output and warm extremities (see Table 20.1).

In the hyperdynamic stage of septic shock, blood pressure falls because of the decreased systemic vascular resistance and decreased venous return. Diastolic pressure declines because of a lack of sympathetic tone, and a widened pulse pressure results. The heart rate and stroke volume increase, and cardiac output is higher than normal, but the patient remains hypotensive. If uncorrected, the myocardium can experience global ischemia, leading to systolic dysfunction resulting in a mixed presentation of both septic and cardiogenic shock.

The patient is usually febrile and may have associated chills. In contrast to cardiogenic and hypovolemic shock, in which the peripheral circulation is reduced and extremities are cool and constricted, the skin is pink and warm to the touch in sepsis as a result of peripheral vasodilation. The patient's level of consciousness may be altered as a result of cerebral ischemia. In septic shock, Svo_2 levels may be higher

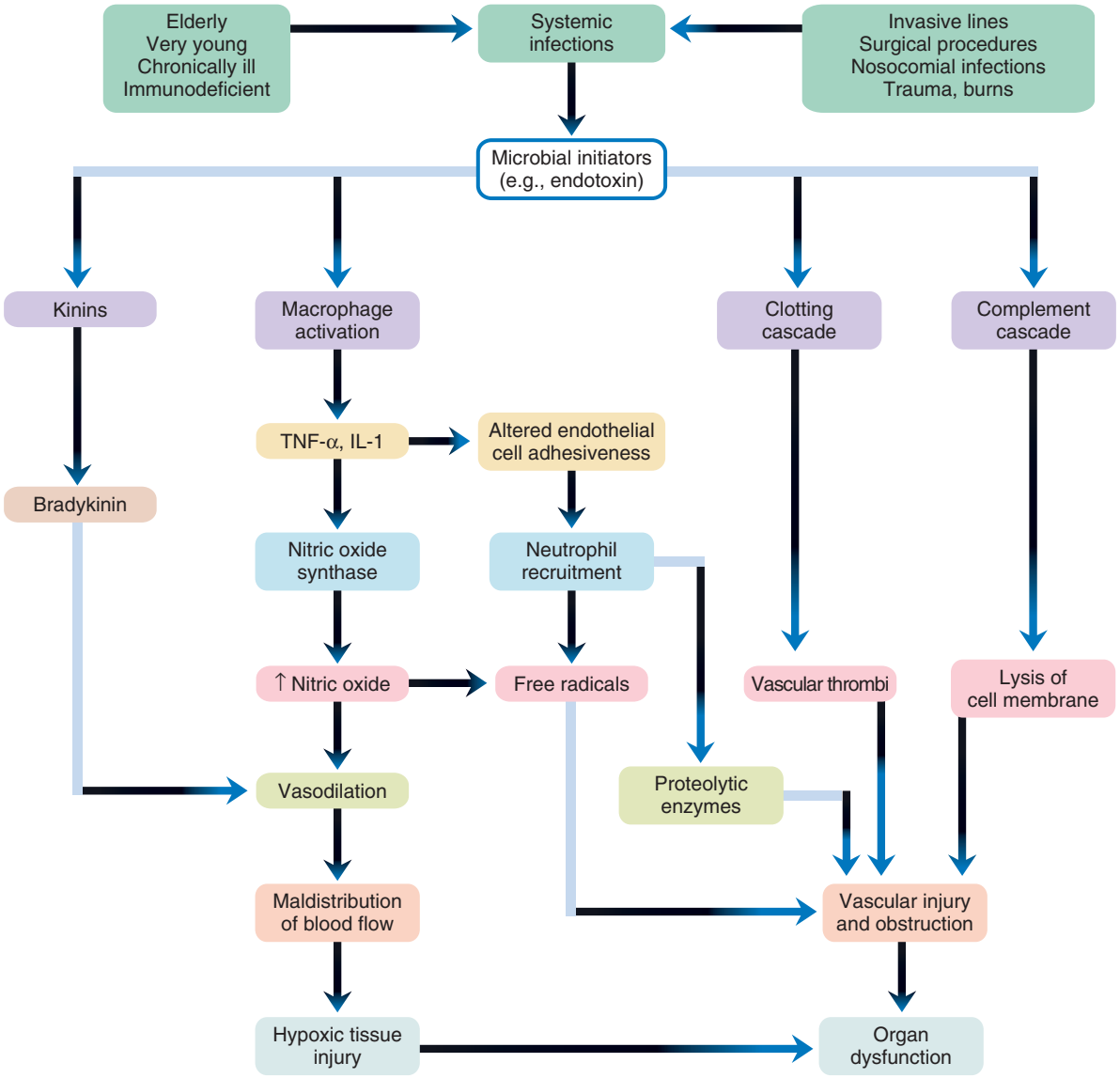


FIG 20.8 Pathophysiologic process of septic shock. Septic shock is characterized by immune-mediated mechanisms of cellular injury and organ dysfunction. *IL-1,* Interleukin-1; *TNF-α,* tumor necrosis factor-α.

than normal because of the maldistribution of blood flow. Abnormal vasodilation causes greater flow through areas with low metabolic activity. Oxygen consumption by tissues is decreased because metabolically active tissues do not receive enough flow. Lactic acidosis may be present because of tissue hypoxemia.

In the progressive stage of septic shock, some patients deteriorate to a hypodynamic phase. The hypodynamic phase is characterized by decreased cardiac output and the development of organ ischemia. The pulse pressure narrows, and the skin becomes cool and clammy. Profound hypotension unresponsive to catecholamines generally occurs. Arterial blood gas analysis reveals a metabolic and respiratory acidosis with hypoxemia. Myocardial depression either from ischemia or from toxins acting as myocardial depressants contributes to a decreasing cardiac output, deteriorating tissue perfusion, and refractory shock.

Treatment. The primary treatment in early septic shock is buffered isotonic fluid administration at 30 mL/kg to restore adequate ventricular preload. If fluid administration does not restore hemodynamic stability, vasopressor therapy may be indicated to increase the mean arterial pressure (MAP) to more than 65 mm Hg. In patients where crystalloid solutions and vasopressors are inadequate, inotropic treatment may be indicated to increase cardiac output and oxygen delivery to tissues. Vascular unresponsiveness to these agents may improve with administration of glucocorticoids. In some cases the normal adrenal production of cortisol may be insufficient, and glucocorticoid replacement is helpful. High-dose and high-potency glucocorticoids are not recommended because of their immunosuppressive activity.

Appropriate broad-spectrum antibiotic therapy is started as soon as septic shock is suspected and after quickly obtaining appropriate cultures. Positive blood cultures can be used to narrow the antibiotic regimen to cover the specific microbes; however, blood cultures are negative in about 50% of cases despite the probable presence of systemic microorganisms. Eradication of the inciting organism reduces the stimulus perpetuating SIRS. Shock itself may propagate sepsis by impairing circulation to the intestinal wall and allowing resident microorganisms to traverse from the colon to the bloodstream. Antibiotic selection for septic shock must be modified as new infective organisms are detected.

Because the inflammatory response is believed to be a critical aspect of septic shock, numerous agents designed to inhibit various components

of SIRS have been investigated. Unfortunately, nearly all of these agents have failed to provide significant benefit. Some subgroups of septic patients may be helped by these agents, but overall the trials have been disappointing. Activated protein C is currently recommended for those with severe sepsis and a high-risk score for mortality. Protein C has both antiinflammatory and antithrombotic actions, which may be helpful in sepsis, but the complication of bleeding is common. Therapy for septic shock is complicated and requires intensive monitoring and treatment of complications. These aspects are discussed in the next sections of this chapter.

KEY POINTS

- Anaphylactic, neurogenic, and septic shock are types of distributive shock characterized by excessive vasodilation and peripheral pooling of blood. Cardiac output is inadequate because of reduced preload.
- Anaphylactic shock is a result of excessive mast cell degranulation in response to antigen. Mast cell degranulation usually is mediated by IgE antibodies. Release of vasodilatory mediators, such as histamine, into the circulation by mast cells results in severe hypotension. Urticaria, bronchoconstriction, stridor, wheezing, and itching are usually present. Treatment includes maintenance of airway patency and the use of epinephrine, antihistamines, vasopressors, and fluids to restore blood pressure.
- Neurogenic shock results from loss of sympathetic activation of arteriolar smooth muscle. The usual causes include medullary depression from brain injury, drug overdose, or lesions of sympathetic nerve fibers, such as spinal cord injury.
- Septic shock results from a severe systemic inflammatory response to infection. Gram-negative bacteria, gram-positive bacteria, and fungal infections are common causes of septic shock. In gram-negative shock, endotoxins in bacterial cell walls stimulate massive immune system activation. Septic shock from any organism is characterized by release of large numbers of immune mediators (e.g., cytokines) resulting in widespread inflammation. The clotting cascade, complement system, and kinin system are activated as part of the immune response.
- Widespread inflammation leads to profound peripheral vasodilation with hypotension, maldistribution of blood flow with cellular hypoxia, and increased capillary permeability with edema formation.
- Initially, septic shock is characterized by abnormally high cardiac output resulting from immune-mediated vasodilation and sympathetic activation of the heart. The patient is usually febrile, pink, and warm. Even though cardiac output is high, cellular hypoxia is present because of maldistribution of blood flow. Reduced cellular oxygen utilization is manifested as a high Svo_2.
- Therapy for septic shock is aimed at improving the distribution of blood flow and managing infection with antibiotics. Administration of fluid and drugs to increase cardiac and vascular performance is done to improve the distribution of blood flow.

ASSESSMENT AND HEMODYNAMIC MONITORING

Astute assessment and appropriate hemodynamic monitoring are essential for the prevention, detection, and management of shock. Although new methods of direct visualization of capillary flow are in development, it is not clinically feasible to directly measure the adequacy of cellular oxygenation in all of the body tissues. A number of indirect measures and clinical signs and symptoms are used to help indicate when tissue hypoxia is probably occurring. Clinical signs of shock represent tissue hypoperfusion and may include alteration of mentation, cool diaphoretic skin, and a reduction in urinary output (less than 30 mL/hr). These same parameters are used to tailor therapy and assess outcomes of that therapy.

Most hospitalized patients experiencing shock will have monitoring devices in place to facilitate assessment of cardiac output, blood pressure, preload, vascular resistance, arterial oxygen content, and venous oxygen content. In addition, frequent measurement of serum lactate concentration, acid–base status, and urine output can be used to indirectly assess the severity of tissue hypoperfusion and hypoxemia.

An understanding of hemodynamic principles and monitoring techniques is helpful to the discussion of shock states. A more thorough discussion can be found in Chapters 15 and 17, and only the main points are reviewed here. The most important factors determining adequate tissue oxygenation are cardiac output, arterial oxygen content, and distribution of blood flow.

Cardiac Output

When the cardiac index falls below 2.2 L/min/m^2, the potential for inadequate tissue perfusion is high. The cardiac index is the cardiac output divided by the body surface area. Cardiac output is the product of heart rate and stroke volume. Stroke volume is the amount of blood ejected by the ventricle with each heartbeat. Stroke volume is influenced by three major factors: preload, contractility, and afterload. *Preload* is the amount of blood in the ventricle at the end of diastole. In patients with low preload, a significant improvement in cardiac output often can be achieved by administering blood or intravenous fluids.

Afterload is the aortic impedance that the left ventricle must overcome to eject blood during systole. The major factors determining aortic impedance are the patency of the aortic valve and the resistance in the systemic vascular system. As resistance to left ventricular ejection increases, stroke volume decreases. Conversely, as resistance falls, stroke volume increases. In patients with high afterload, vasodilating agents may be useful in reducing the workload on the left ventricle. Care must be taken not to lower arterial blood pressure excessively, which would reduce perfusion.

Contractility is the inherent state of activation of cardiac muscle fibers. Contractility depends on the amount of free calcium ions available in the cardiac muscle cells after each electrical impulse. Contractility is influenced by sympathetic and parasympathetic nervous system neurotransmitters and other hormones and drugs. Contractility also depends on the amount of muscle mass and is influenced by myocardial ischemia and necrosis, in addition to myocardial-depressant factors that may be released by tissues in shock. Contractility can be increased by measures that increase myocardial perfusion and oxygenation, as well as by use of positive inotropic agents.

Most of the therapeutic interventions aimed at increasing cardiac output also increase myocardial workload and myocardial oxygen consumption. Especially in cardiogenic shock, these parameters must be carefully monitored and manipulated to avoid further cardiac compromise.

Arterial Oxygen Content

Oxygen delivery ($\dot{D}o_2$) can be determined by multiplying cardiac output and arterial oxygen content (Cao_2). Cao_2 is the sum of dissolved oxygen and oxygen bound to hemoglobin. Adequate gas exchange in the pulmonary capillaries is necessary to fully saturate hemoglobin with oxygen. Impaired ventilation may result in reduced Cao_2 and impair $\dot{D}o_2$. Mechanical ventilation and supplemental oxygen administration may be used to improve arterial oxygen saturation. For patients with low levels of hemoglobin, blood transfusion may significantly improve $\dot{D}o_2$.

A concept closely related to $\dot{D}o_2$ is oxygen consumption ($\dot{V}o_2$). Whereas $\dot{D}o_2$ is a measure of the oxygen delivered to the tissues each minute, $\dot{V}o_2$ is the amount of oxygen actually used by the tissues per minute. In a normal physiologic state, only about 25% of the oxygen delivered is taken up by tissues, which leaves about 75% of the oxygen

to return to the heart in venous blood. When $\dot{D}o_2$ falls because of low cardiac output, tissues extract a greater percentage of the oxygen delivered such that the amount returning in venous blood is lower. When distribution of blood flow and tissue extraction of oxygen are impaired, as in septic shock, oxygen consumption falls and the amount of oxygen returning in venous blood will be higher than it should be. The amount of oxygen returning to the heart in venous blood can be measured by a special catheter in the pulmonary artery that detects venous oxygen saturation (Svo_2). Outcomes of therapy to improve $\dot{D}o_2$ can be assessed by monitoring Svo_2. In cardiogenic shock, for example, one would expect to see Svo_2 increase from a low value back toward 75% as cardiac output improves. In septic shock, one would expect to see Svo_2 decrease from a high value back toward 75% as distribution of blood flow to metabolically active tissues improves.

Distribution of Blood Flow

In addition to adequate cardiac output and arterial oxygen content, optimal blood circulation is imperative to prevent cellular hypoxia and secondary injury. During states of reduced circulating volume a number of compensatory and neurohormonal pathways are activated to maintain tissue perfusion (as described earlier). As prolonged states of cellular hypoperfusion continue, the severity of shock progresses from decompensated to irreversible shock and death.

Hemodynamic Monitoring

Sophisticated monitoring equipment is available to assess the hemodynamic status of patients in shock. Normal values are listed in Table 20.5. A flow-directed, pulmonary artery catheter can be inserted through the jugular or subclavian vein to allow measurement of intracardiac pressures, cardiac output, and Svo_2 (Fig. 20.9). A catheter lumen in the right atrium allows measurement of right atrial pressure. Right atrial pressure is used to indicate right ventricular end-diastolic volume or preload. The primary value of monitoring right atrial pressure is in the management of blood volume. A low right atrial pressure is associated with a low preload and may indicate a need for extracellular volume replacement to enhance cardiac output. Conversely, a high right atrial pressure may indicate a need for extracellular volume reduction to decrease cardiac workload and congestive symptoms.

TABLE 20.5 Hemodynamic Parameters

Parameter	Formula	Normal Values
Central venous pressure (CVP) (Right atrial pressure)	Direct measurement	2–6 mm Hg
Left ventricular end-diastolic pressure (LVEDP)	Direct measurement	5–12 mm Hg
Mean arterial pressure (MAP)	= (SBP – DBP / 3) + DBP	70–105 mm Hg
Pulm art press systolic (PAS)	Direct measurement	15–25 mm Hg
Pulm art press diastolic (PAD)	Direct measurement	8–15 mm Hg
Pulm art press mean (PAPM)	Direct measurement	10–20 mm Hg
Pulm art occlusion pressure (PAOP)	Direct measurement	6–12 mm Hg

Adapted from Parrillo, Dellinger, eds: *Critical care medicine: principles of diagnosis and management in the adult*, ed 4, Philadelphia, 2014, Elsevier Saunders.

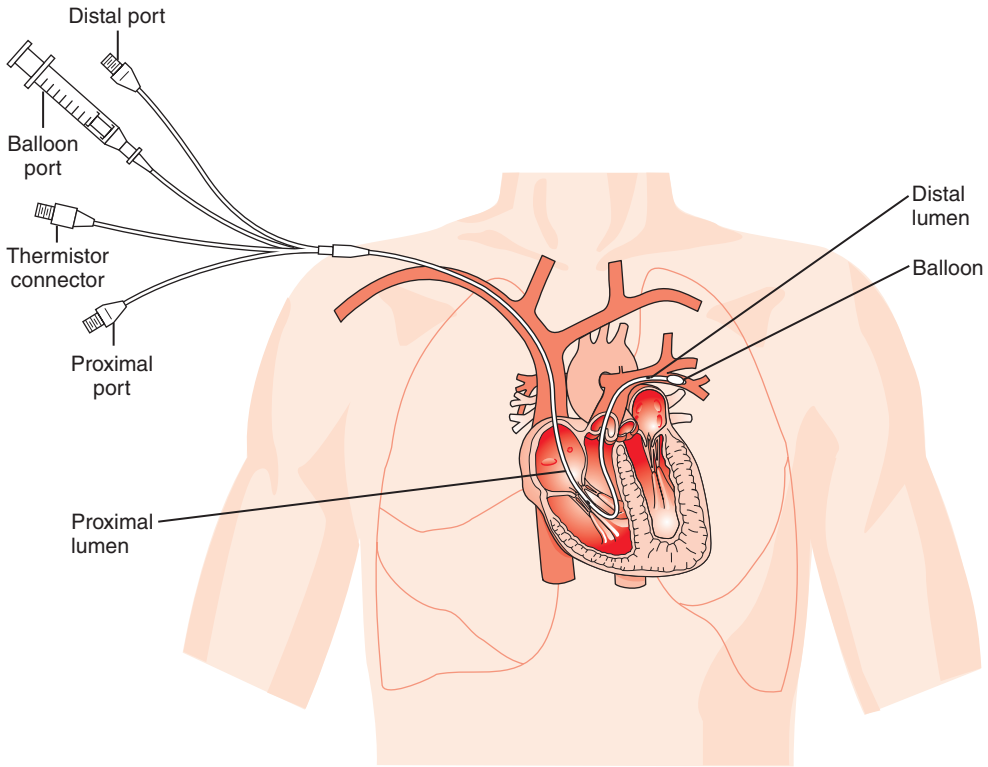


FIG 20.9 Properly positioned pulmonary artery catheter showing the proximal port in the right atrium and the distal port in the pulmonary artery. Cardiac output determinations can be made by injecting hypothermic solution into the proximal port and measuring the degree of warm-up near the distal port. A balloon at the end of the catheter can be intermittently inflated to measure pulmonary capillary occlusion pressure. When the balloon is inflated, it will float into a small artery and wedge there. Then the distal port measures the pressure in the capillary, which is a direct reflection of left atrial pressure.

Another catheter lumen located in the pulmonary artery allows measurement of pulmonary artery pressure. Measurement of pulmonary artery pressure is helpful in assessing pulmonary complications of shock. An increase in pulmonary artery pressure may occur in progressive shock as the lungs react to inflammatory mediators and become edematous. In the absence of lung disease, pulmonary artery diastolic pressure reflects left atrial pressure. Assessment of left atrial pressure is important because it indicates left ventricular preload—an important determinant of cardiac output. A more accurate assessment of left atrial pressure can be obtained by using a small balloon at the tip of the catheter to obtain a pulmonary capillary occlusion pressure. When the balloon is inflated, the catheter tip floats into a small pulmonary artery and wedges itself there. The balloon blocks the arterial pressure events behind it and allows measurement of pressure in the capillary. Pulmonary capillary occlusion pressure is a direct reflection of left atrial pressure. Low left atrial pressure indicates reduced left ventricular preload and may signify the need for extracellular volume replacement (Table 20.5).

KEY POINTS

- Hemodynamic monitoring during shock states is helpful for assessing cardiac output, volume status, oxygen delivery, and oxygen consumption. The pressures usually monitored include right atrial pressure, pulmonary artery pressure, and left atrial pressure.
- Hemodynamic monitoring is used to guide management of cardiac preload, afterload, and contractility to optimize cardiac output, while minimizing cardiac workload.
- Normally, about 25% of the oxygen in arterial blood is extracted by the tissues, so the mixed venous oxygen saturation (Svo_2) is approximately 75%. Low cardiac output may result in greater oxygen extraction and lower Svo_2; maldistribution of flow, as occurs in septic shock, may result in less oxygen extraction and higher Svo_2.

COMPLICATIONS OF SHOCK

The pathologic process of the shock state and the effects on other organs may precipitate life-threatening complications. In severe shock of any cause, particularly in septic shock, a generalized inflammatory reaction may occur and is thought to contribute to the organ damage associated with shock states. Complications associated with shock include acute respiratory distress syndrome (ARDS), disseminated intravascular coagulation (DIC), acute renal failure, and multiple organ dysfunction syndrome (MODS). Damage to organ systems may be ongoing even after the initial precipitating event has been addressed. Inflammatory cytokines are thought to mediate this organ damage by altering metabolism, recruiting neutrophils, initiating the coagulation cascade, and altering capillary permeability. The complexities of this syndrome are being slowly unraveled as the mechanisms of immune signaling are better understood.

Acute Respiratory Distress Syndrome

ARDS, a form of respiratory failure, is most commonly associated with septic shock. ARDS is characterized by the development of refractory hypoxemia, decreased pulmonary compliance, and radiologic evidence of pulmonary edema associated with normal cardiac preload (noncardiogenic pulmonary edema). The mortality in patients with shock that is complicated by ARDS ranges from 34% to 64%. The primary cause of death in ARDS patients is multiple organ failure, not severe hypoxemia.

The lungs are a common target of immune-mediated damage in all types of shock. Tissue ischemia, even in areas distant from the lungs, leads to neutrophil migration to pulmonary capillaries. Neutrophils release destructive proteolytic enzymes, produce oxygen free radicals, and secrete inflammatory chemicals that make pulmonary capillaries leaky. A protein-rich inflammatory exudate leaks into the interstitial spaces and alveoli of the lung, where it interferes with pulmonary gas exchange. Inflammation may also damage type II pneumocytes, that normally produce surfactant. Surfactant deficiency alters alveolar surface tension and causes smaller alveoli to collapse. The effort to breathe is very great in patients with ARDS because of pulmonary edema and alveolar collapse (atelectasis). Further discussion of ARDS and its clinical manifestations can be found in Chapter 23.

Disseminated Intravascular Coagulation

DIC is a serious complication of septic shock characterized by abnormal clot formation in the microvasculature throughout the body. DIC is thought to result from immune activation of the clotting cascade. Obstruction of blood flow by small clots in the microcirculation leads to ischemic tissue damage. In addition, widespread clot formation consumes platelets and clotting factors, which leaves the patient at risk for serious bleeding. Laboratory assessment of the platelet count and clotting function is helpful in detecting and monitoring DIC. The platelet count and fibrinogen levels are typically low, whereas levels of fibrin degradation products (e.g., D-dimer) are elevated. Measures of the intrinsic and extrinsic clotting cascades demonstrate an elevated partial thromboplastin time and prothrombin time.

The clinical features of DIC are variable, depending on the location and severity of vascular thrombi. Vascular obstruction may be manifested as acute ischemia of the fingers and toes, with pain, pallor, and poor capillary refill. Obstruction of the kidney, liver, spleen, and lung by clots may result in signs and symptoms of organ failure. Patients may demonstrate various degrees of bleeding. Intravenous lines and catheters may begin to ooze around insertion sites. Previously stable incision lines may begin to bleed, and hematuria and hemoptysis may be present. Spontaneous intracranial hemorrhage is a particularly disastrous complication of DIC. Further discussion of DIC can be found in Chapter 14.

Acute Renal Failure

In shock, the kidneys undergo prolonged periods of hypoperfusion. Vasoconstriction of the afferent arterioles causes decreased glomerular blood flow, decreased glomerular hydrostatic pressure, and decreased glomerular filtration rates. Hypoxic cellular damage occurs after 15 to 20 minutes of acute ischemia and results in necrosis of tubular epithelial cells. Acute tubular necrosis (ATN) is associated with decreased urinary excretion of waste products such as creatinine and urea. Rapidly increasing blood urea nitrogen and serum creatinine concentrations are indicative of ATN.

Urine output quickly falls toward zero, and the kidneys do not respond to fluids or diuretics. Renal tubular epithelial cell casts in the urine indicate sloughing of tubular cells. ATN is potentially reversible, although renal function must generally be supported for a time with dialysis. Recovery of tubular function begins 1 to 2 weeks after the initial injury and may take up to 1 year to be completed. Further discussion of ATN can be found in Chapter 28.

Multiple Organ Dysfunction Syndrome

When organ dysfunction develops in two or more systems, the term *MODS* may be applied. When the patient sustains multiple organ injury from a primary insult such as trauma, the term *primary MODS* is used. *Secondary MODS* is associated with SIRS and usually develops days to weeks after the primary insult. Sepsis and septic shock are the most common causes of secondary MODS. Mortality from MODS differs depending on the number of organs affected; involvement of two organ

systems carries a 54% mortality, and involvement of five organs carries a 100% mortality. In MODS, the body is unable to maintain homeostasis, and intensive intervention is necessary to maintain life.

As with other manifestations of septic shock, MODS is thought to be initiated by immune mechanisms that are overactive and destructive. Immune cytokines affect endothelium throughout the body and cause recruitment of neutrophils and activation of inflammation in vascular beds. Ongoing inflammation leads to tissue destruction and organ dysfunction. Inflammatory cytokines and stress hormones stimulate an increased body metabolism, which places a greater demand on already dysfunctional organs.

KEY POINTS

- Shock states result in reduced or inadequate cellular oxygen consumption and may affect all organs and systems in the body. Complications of shock can be viewed as inflammatory in nature. Inflammation is triggered by hypoxic injury to cells, by antigen, or by endotoxin. Excessive or inappropriate immune system responses lead to leaking capillaries; damage from proteolytic enzymes; and systemic activation of the clotting, complement, and kinin systems.
- Respiratory failure and kidney failure are commonly associated with shock. Inappropriate activation of the clotting cascade may result in disseminated intravascular coagulation (DIC). Multiple organ dysfunction syndrome (MODS) may occur with widespread cellular hypoxia and necrosis.

SUMMARY

Shock is a life-threatening syndrome associated with high mortality. Early identification of patients at risk and initiation of therapeutic measures may decrease the development of shock syndrome. Four major categories of circulatory shock have been described: cardiogenic, obstructive, hypovolemic, and distributive. Although each type of shock has specific characteristics, all are associated with a deficiency of cellular oxygen consumption. Tissue ischemia leads to hypoxic cellular dysfunction and death, generation of oxygen free radicals, and stimulation of a systemic inflammatory response. In late-stage shock and in septic shock, ongoing systemic inflammation leads to progressive organ dysfunction and can precipitate a number of shock complications, including ARDS, DIC, ATN, and MODS. Ongoing research into effective ways to improve microcirculatory function and intervene in the inflammatory cascade is needed to improve outcomes.

RESOURCES

Pathophysiology of Shock

American College of Surgeons: *Advanced trauma life support course for physicians*, Chicago, 2008, Author.

Andrades ME, Movina A, Spasic S, Spasojevic I: Bench-to-bedside review: sepsis—from the redox point of view. *Crit Care* 15(5):230, 2011.

Deleston D, Opal SM: Future perspectives on regulating pro- and anti-inflammatory responses in sepsis. *Contrib Microbiol* 17:137–156, 2011.

Devlin JW, Matzke GR: Acid-base disorders. In Dipiro JT, et al, editors: *Pharmacotherapy: a pathophysiologic approach*, ed 8, New York, 2011, McGraw-Hill, pp 923–942.

Elbers PW, Ince C: Bench-to-bedside review: mechanisms of critical illness—classifying microcirculatory flow abnormalities in distributive shock. *Crit Care* 10(4):221–228, 2006.

Ellis CG, Jagger J, Sharpe M: The microcirculation as a functional system. *Crit Care* 9(Suppl 4):S3–S8, 2005.

Rixen D, Siegel JH: Bench-to-bedside review: oxygen debt and its metabolic correlates as quantifiers of the severity of hemorrhagic and post-traumatic shock. *Crit Care* 9(5):441–453, 2005.

Warren JC: *Surgical pathology and therapeutics*, Philadelphia, 1895, Saunders.

Cardiogenic and Hypovolemic Shock

Erstad BL: Hypovolemic shock. In Dipiro JT, et al, editors: *Pharmacotherapy: a pathophysiologic approach*, ed 8, New York, 2011, McGraw-Hill, pp 421–436.

Hochman JS, et al: Early revascularization in acute myocardial infarction complicated by cardiogenic shock. SHOCK Investigators. Should we emergently revascularize occluded coronaries for cardiogenic shock? *N Engl J Med* 341(9):625–634, 1999.

MacLaren R, Rudis MI, Dasta JF: Use of vasopressors and inotropes in the pharmacotherapy of shock. In Dipiro JT, et al, editors: *Pharmacotherapy: a pathophysiologic approach*, ed 8, New York, 2011, McGraw-Hill, pp 399–420.

Shoemaker WC, et al: Resuscitation from severe hemorrhage. *Crit Care Med* 24(Suppl):S12–S23, 1996.

Topalian S, Ginsberg F, Parrillo JE: Cardiogenic shock. *Crit Care Med* 36(Suppl 1):S66–S74, 2008.

Trost JC, Hillis LD: Intra-aortic balloon counterpulsation. *Am J Cardiol* 97:1391–1398, 2006.

Septic Shock

Dellinger RP: Cardiovascular management of septic shock. *Crit Care Med* 31(3):946–955, 2003.

Dellinger RP, et al: Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock. *Crit Care Med* 36(1): 296–327, 2008, 2008.

Del Sorbo L, Slutsky AS: Acute respiratory distress syndrome and multiple organ failure. *Curr Opin Crit Care* 17(1):1–6, 2011.

Dhainaut JF: International integrated database for the evaluation of severe sepsis (INDEPTH): clinical evaluation committee report on the safety of drotrecogin alfa (activated) therapy. *Curr Med Res Opin* 24(4):1187–1197, 2008.

Gotts JE, Matthay MA: Sepsis: pathophysiology and clinical management. *BMJ* 353:i1585, 2016.

Kakihana Y, Ito T, Nakahara M, et al: Sepsis-induced myocardial dysfunction: pathophysiology and management. *J Intensive Care* 4(22):2016. doi:10.1186/s40560-016-0148-1.

Kang-Birken SL, Killgore-Smith K: Severe sepsis and septic shock. In Dipiro JT, et al, editors: *Pharmacotherapy: a pathophysiologic approach*, ed 8, New York, 2011, McGraw-Hill, pp 2041–2054.

Levy MM, Dellinger RP, Townsend SR, et al: The Surviving Sepsis Campaign: results of an international guideline-based performance improvement program targeting severe sepsis. *Intensive Care Med* 36:222–231, 2010.

Simons FE: Anaphylaxis. *J Allergy Clin Immunol* 125:S161, 2010.

Singer ME, Deutschman CS, Seymour CW, et al: The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* 315(8):762–774, 2016.

Trzeciak S, Dellinger R, Parrillo J: Septic shock. In Parrillo J, Dellinger R, editors: *Critical care medicine: principles of diagnosis and management in the adult*, ed 4, Philadelphia, 2014, Elsevier Saunders, pp 365–378.

21

Respiratory Function and Alterations in Gas Exchange

Lorna L. Schumann

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- How do the structures involved in gas exchange in the lungs differ from conducting structures?
- What factors determine the work of breathing?
- How are alveolar ventilation and oxygenation estimated and assessed?
- What factors affect the distribution of ventilation and perfusion in the lungs?
- How are oxygen and carbon dioxide transported in the circulation?
- What pathophysiologic factors might alter ventilation-perfusion matching in the lungs?
- How can the outcomes of ventilation-perfusion mismatching be assessed clinically?
- What pulmonary function test abnormalities are characteristic of obstructive pulmonary disorders?
- What are the risk factors and complications of pulmonary venous thromboembolism and hypertension?
- What are the various types of pulmonary malignancies?

CHAPTER OUTLINE

Functional Anatomy, 452
Development of the Pulmonary System, 452
Upper Airway Structures, 452
Lower Airway Structures, 453
Pulmonary Circulation, 457
Age-Related Variations, 457
Ventilation, 460
Lung Volumes and Capacities, 460
Dead Space, 460
Minute Ventilation, 460
Alveolar Ventilation/Oxygenation, 461
Mechanics of Breathing, 461
Airway Resistance, 461
Lung Compliance, 462
Distribution of Ventilation, 462
Neurologic Control of Ventilation, 462
Pulmonary Blood Flow, 465
Pulmonary Vasculature, 465
Distribution of Blood Flow, 465
Ventilation–Perfusion Ratios, 465
Hypoxic Vasoconstriction, 466
Diffusion and Transport of Respiratory Gases, 466
Barriers to Diffusion, 466
Oxygen Transport, 467
Carbon Dioxide Transport, 467
Alterations in Pulmonary Function, 468
Hypoventilation and Hyperventilation, 468
Hypoxemia and Hypoxia, 468
Acute Respiratory Failure, 469
Etiology, 469
Clinical Manifestations, 470
Diagnosis, 470
Treatment, 470
Diagnostic Tests, 471
Pulmonary Function Testing, 471
Bronchial Provocation Tests, 472
Alterations in Pulmonary Vasculature, 472
Pulmonary Hypertension, 472
Etiology, 472
Pathogenesis, 472
Clinical Manifestations, 472
Diagnosis, 473
Treatment, 473

Pulmonary Venous Thromboembolism, 473
Etiology, 473
Pathogenesis, 474
Clinical Manifestations, 474
Diagnosis, 474
Treatment, 475

Pulmonary Malignancies, 475
Etiology, 475
Pathogenesis, 475
Clinical Manifestations, 476
Diagnosis, 476
Treatment, 476

The primary function of the lungs is gas exchange. Oxygen is transported to the body tissues, and carbon dioxide, a waste product, is transported out of the body. The exchange of these gases takes place at the alveolar-capillary membrane. For effective gas exchange to occur, the processes of ventilation, perfusion, and diffusion must occur simultaneously at the alveolar-capillary interface. Problems with any of these three processes can result in hypoxemia (low arterial oxygen concentration) or hypercarbia (high arterial carbon dioxide concentration). An understanding of the anatomy and physiology of pulmonary gas exchange is necessary for learning about the pathophysiologic processes that follow.

FUNCTIONAL ANATOMY

Development of the Pulmonary System

Initially, the laryngotracheal diverticulum includes the esophagus and the trachea as a single tube. Then longitudinal ridges begin to develop along the tube and form a septum (wall), which separates the esophagus from the trachea. Failure of this septum to develop leads to a tracheoesophageal fistula (abnormal opening), leaving a communication between the esophagus and the trachea. This abnormality occurs about once in every 2500 births. Approximately 90% of the cases of esophageal atresia (blind pouch) are of the type seen in Fig. 21.1, *A*. Parts *B* through *D* show variations of tracheoesophageal fistulas.

As the laryngotracheal tube continues to elongate, the lung bud divides into two bronchial buds, which become the bronchi and the right and left lungs. The right bronchus becomes larger than the left. The right mainstem bronchus is normally more vertical than the left because it is the main continuation of the laryngotracheal tube and branches off the trachea at a 20-degree angle. The left bronchus branches off the trachea at an angle of 40 to 60 degrees. This normal anatomic development increases the chances that an inhaled foreign body will lodge in the right mainstem bronchus rather than the left.

Fetal lung development can be divided into the following four periods:

1. During the pseudoglandular period (5 to 17 weeks) the bronchial divisions are differentiated, and the major elements of lung tissue are present except for those involved in gas exchange: the respiratory bronchioles and alveoli.
2. During the canalicular period (16 to 25 weeks) the bronchi and bronchioles enlarge and vascularization of lung tissue takes place. At the end of this period, respiration is possible because of the development of respiratory bronchioles and primitive alveoli. Alveoli are grapelike sacs in which gas exchange occurs. Type II pneumocytes (epithelial cells that are on the internal surface of alveoli) begin to secrete surfactant at the end of this period. Surfactant is a phospholipid essential for maintaining alveolar patency.
3. During the terminal sac period (24 weeks to birth), terminal air sacs become thinner, preparing the lung tissue for gas exchange. Proliferation of pulmonary capillaries is also prominent during this period. Infants born prematurely in the early weeks of this period (25 to 28 weeks) are susceptible to the development of respiratory distress syndrome because of the immaturity of the pulmonary structures.
4. The alveolar period (late fetal life to 8 years) is the final period of lung development when alveolar ducts form from terminal sacs and alveoli mature by increasing in size and number. Approximately one eighth to one sixth of the adult number of alveoli are present at birth. During this growth period, there is a lack of structural collateral pathways necessary for maintaining open airways. This may make the individual more susceptible to atelectasis (incomplete expansion) and obstruction. Lung damage during this period may cause permanent defects in lung development.

Upper Airway Structures

The respiratory system can be divided into two major anatomic areas: the upper airway and the lower airway. The upper airway consists of the nasopharyngeal cavity (nasopharynx, oropharynx, laryngopharynx) (Fig. 21.2). The lower airway contains the larynx, trachea, bronchi, bronchopulmonary segments, terminal bronchioles, and the acinus (the alveolar region supplied by one terminal bronchiole, which includes numerous alveoli) (see Fig. 21.2).

The nasal cavity conducts gases to and from the lungs and filters, warms, and humidifies the air. It is a rigid box composed of two-thirds cartilage and one-third bone, which prevents collapse during movement of air. The convoluted turbinates (cone-shaped bones) of the nasal cavity are highly vascular, and their blood flow forms an efficient heat exchanger. Evaporation of water from the turbinate surface and from the mucus secreted by mucosal glands raises the water vapor of the inspired air to normal saturation. Therefore air is warmed to body temperature and humidified.

Air is filtered by the large hairs (vibrissae) of the nasal cavity mucus and cilia that line the nasal cavity. The cilia sweep foreign particles trapped by mucus into the nasopharynx, where they are swallowed or expectorated. An electron micrograph of the tracheobronchial lining is shown in Fig. 21.3. Pseudostratified ciliated columnar epithelium lines the trachea and bronchi. Goblet cells and mucus-producing glands are contained in this area and are responsible for synthesizing approximately 100 mL/day in the adult, more with disease. The composition of mucus is 95% water, with the remaining 5% consisting of mucopolysaccharides, mucoproteins, and lipids. Maintenance of water content and fluid balance is important to the mobilization of secretions. A child has more mucus-producing glands and therefore produces more mucus than an adult. Consequently, in an ill child the overproduction of mucus in combination with small airway size may precipitate tracheobronchial obstruction.

Cilia (Fig. 21.4) beat in a sweeping motion like oars rowing a boat at approximately 1000 to 1500 strokes per minute. Mucociliary transport (movement of mucus and trapped particles upward) is a primary defense mechanism of the tracheobronchial tree. Inhaled particles, bacteria, and macrophages are removed from the respiratory tract by ciliary clearance and the cough reflex. Ciliary function is impaired by smoking, alcohol ingestion, hypothermia, hyperthermia, cold air, low humidity, starvation, anesthetics, corticosteroids, noxious gases, the common cold, and increased mucus production.

The four paranasal sinuses are air-containing spaces adjacent to the nasal passages that provide speech resonance and increase the surface

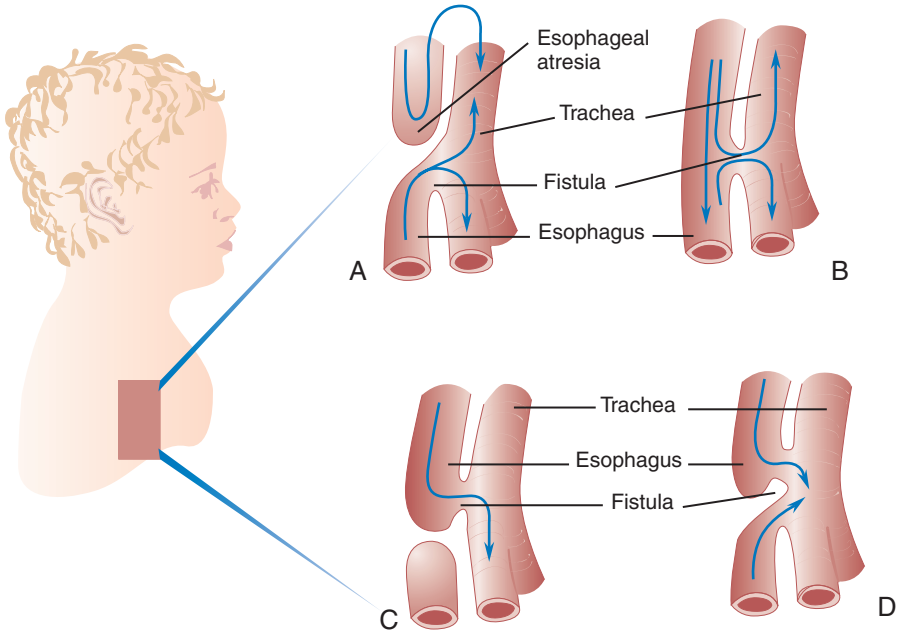


FIG 21.1 Four primary types of tracheoesophageal fistulas. **A,** The most common type, with complete atresia (blind pouch) of the esophagus. **B,** A common opening between the trachea and esophagus. **C,** An opening from an esophageal pouch into the trachea. **D,** A double opening from two unconnected ends of the esophagus. *Arrows* indicate flow of fluid from the esophagus to the trachea.

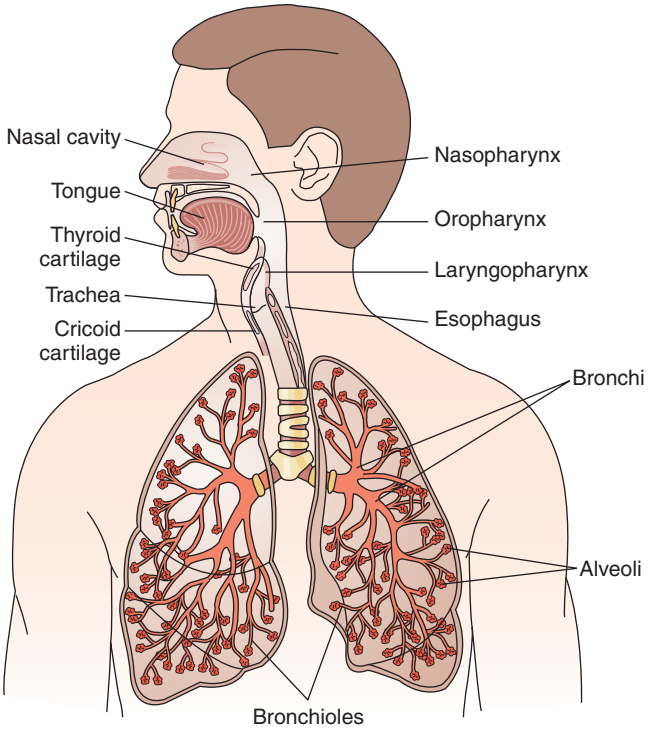


FIG 21.2 Sagittal view diagram of the nasopharyngeal cavity (nasal cavity, nasopharynx, oropharynx, and laryngopharynx) and the respiratory passages beginning at the trachea and ending at the alveoli.

area for heat and water vapor exchange. The sinuses are swept clean by mucociliary action when the communicating passages that connect them with the nasal passages remain open.

The eustachian tube between the middle ear and the posterior nasopharynx maintains the air in the middle ear at atmospheric pressure. To prevent secretions or food from entering the middle ear during swallowing, the pharyngeal muscles close the eustachian tube briefly. The nasal end of the eustachian tube is surrounded by flexible cartilage arranged in a spiral configuration. The muscles surrounding the eustachian cartilage close the opening by pulling the cartilage tighter. Because the tube is shorter in children, the potential for otitis media (infection of the middle ear) is increased.

Lower Airway Structures

After air passes through the nasal cavity or oral cavity into the pharynx, it moves into the larynx and finally into the tracheobronchial tree. The acinus (Fig. 21.5) is located at the end of the tracheobronchial tree and is composed of bronchioles, alveolar ducts, and alveoli.

The larynx is the transition area between the upper and lower airways. Anatomically it is considered part of the lower airway, but functionally it is similar to the upper airway. The larynx contains the epiglottis, vocal cords, and cartilages. The anatomic arrangement of the larynx functions to prevent aspiration during swallowing and to assist in phonation and coughing. Each vocal cord is attached anteriorly to the thyroid cartilage and posteriorly to the arytenoid cartilage. Vibration of the cords leads to phonation. Food is prevented from entering the trachea during swallowing by closure of the epiglottis. If food or fluid should bypass the epiglottis and enter the tracheobronchial tree, the cough reflex is initiated. The majority of cough receptors lie at the carina. A cough reflex is produced when the epiglottis and vocal cords close tightly against air entrapped in the lungs. Occasionally, individuals cough hard enough to break a rib. When the expiratory muscles contract forcefully against the closed epiglottis and vocal cords, a pressure of approximately 100 mm Hg is created. When the cords and epiglottis suddenly open, the high-pressure buildup is allowed to escape. This reflex rapidly removes foreign matter from the tracheobronchial tree.

The major cartilages of the larynx are the thyroid, cricoid, and arytenoid. The thyroid cartilage is a large, shield-shaped cartilage often referred to as the *Adam's apple.* Immediately below the thyroid cartilage is the site for emergency opening (cricothyroidotomy) of the tracheal

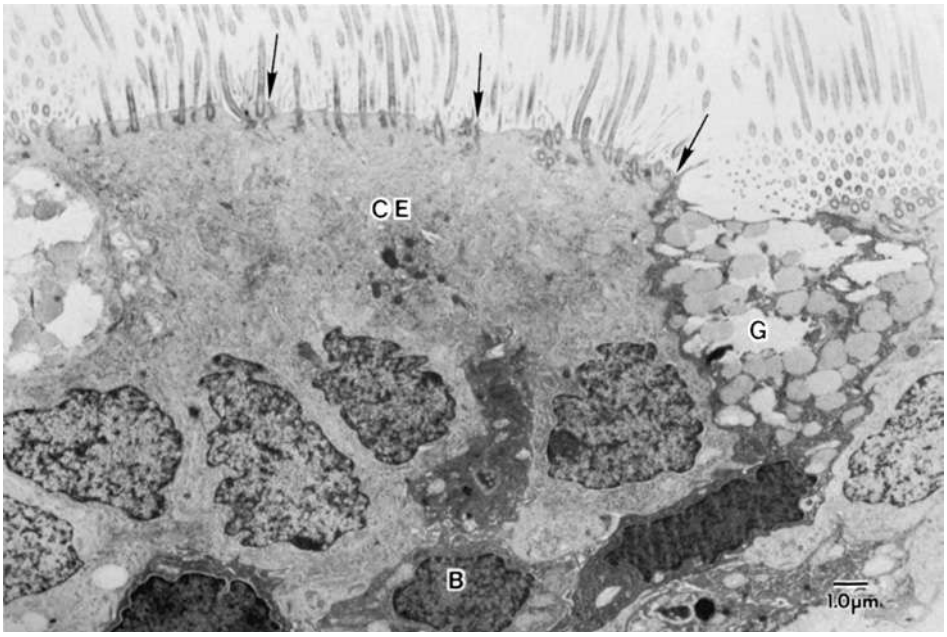


FIG 21.3 Cells composing the bronchial epithelium are ciliated epithelial cells *(CE),* goblet cells *(G),* and basal cells *(B).* Goblet cells have abundant mucus granules in the cytoplasm, and their apical surface is devoid of cilia. Basal cells, as their name indicates, are located along the abluminal portion of the lining epithelium, adjacent to the basal lamina. The *arrows* at the apical surface of the airway cells indicate the location of junctional complexes between contiguous epithelial cells. (Human lung surgical specimen, transmission electron microscopy.) (From Murray JF, Nadel JA: *Textbook of respiratory medicine,* ed 4, Philadelphia, 2005, Elsevier.)

passageway. The cricoid cartilage lies below the thyroid cartilage and is the narrowest point in the airway of a child. It is the only complete tracheal ring, and because of its narrowness in the small child's airway, an endotracheal tube cuff is not necessary for required intubation of the airway.

The trachea, bronchi, and bronchioles make up the conducting airways that allow passage of gases to and from the gas exchange units (alveoli). These conducting airways comprise a proportionately larger amount of the total airway system in the infant and child than in the adult. The trachea (Fig. 21.6) contains incomplete cartilaginous rings; it is approximately 11 to 13 cm long and lies between the cricoid cartilage and the carina (ridge located at the lower end of the trachea). Individual variations in tracheal shape include U, circular, D, C, triangular, and elliptical (Fig. 21.7). Of 111 adult tracheas studied, the incidence of shapes in order of frequency was 48.6% C, 27% U, 12.6% D, 8.2% elliptical, 1.8% circular, and 1.8% triangular. These tracheal variations may affect ventilation of patients who have endotracheal tubes in their airways and require mechanical ventilation.

The trachea divides into two mainstem (primary) bronchi, which contain cartilage and smooth muscle. Viewing the body anteriorly, the carina is located at the angle of Louis, between the sternum and manubrium at the second intercostal space.

The small size of the conducting airway in the infant and child makes even a small decrease in the size of the lumen from an obstruction critical to airway conduction. Primary bronchi further divide into five (secondary) lobar branches, three to the right lung (upper, middle, and lower lobes) and two to the left lung (upper, which includes the lingual [area around the heart] and lower lobes). Each lobar branch enters a lobe of the lung and further divides into bronchopulmonary segments (10 segments in the right lung, 9 segments in the left lung) (Fig. 21.8). Each bronchopulmonary segment is composed of 50 or more terminal bronchioles (conducting airways), which branch into respiratory bronchioles, where gas exchange begins.

Terminal bronchioles, which include the conducting airways, further subdivide into two or more respiratory bronchioles in which gas exchange begins. The respiratory bronchioles divide into two or more alveolar ducts, which in turn supply several alveoli.

Nervous system control of the bronchi and bronchioles is mediated by the autonomic nervous system. Stimulation of the parasympathetic nervous system via the vagus nerve leads to constriction (by means of acetylcholine receptors) of bronchial smooth muscle. Stimulation of the sympathetic nervous system leads to relaxation of bronchial smooth muscle. Sympathetic stimulation is mediated by β_2-adrenergic receptors, which are under the control of circulating catecholamines. (See the discussion in the "Neurologic Control of Ventilation" section later in this chapter for additional information.)

The lung is fully developed by the eighth year of life. The large alveolar surface area in conjunction with pulmonary surfactant, a phospholipid produced by type II alveolar cells, lowers surface tension and facilitates gas exchange. Two other types of cells are found: type I alveolar cells (type I pneumocytes), which are the epithelial structural cells of the alveoli, and alveolar macrophages, which act as a defense mechanism by phagocytizing particles in the alveoli. Alveolar macrophages can be damaged by cigarette smoking and by inhalation of silica (SiO_2).

Adult lungs contain approximately 300 million alveoli, and the newborn lung contains one eighth to one sixth the adult number. An elderly person may also have a reduction in the number of alveoli as part of the normal aging process, but many elderly people retain the same number of alveoli they had as a younger adult.

Gas exchange occurs in the alveolar units (see Fig. 21.5) where oxygen and carbon dioxide transfer across the alveolar-capillary membrane.

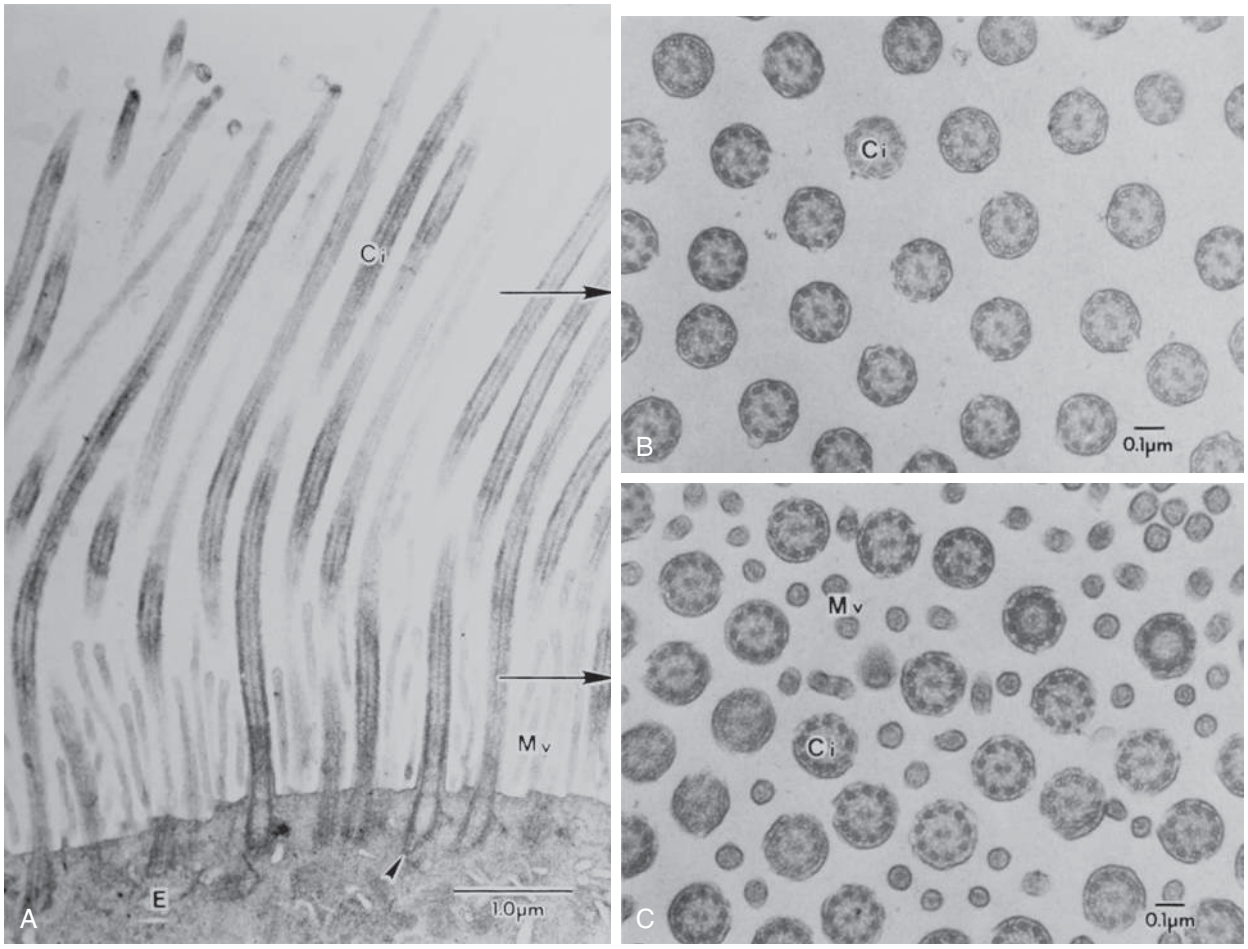


FIG 21.4 A, Electron micrograph shows the ultrastructural characteristics of cilia *(Ci)* on airway epithelial cells *(E).* Each cilium has a long, slender shaft that ends with a conical tip. The base of the cilium is anchored in the cell's apical cytoplasm by a curved and tapered basal foot (modified centriole; *arrowhead*). Also extending from the apical surface of ciliated airway epithelial cells are microvilli *(Mv).* The two horizontal arrows in panel A represent the cross-sectional planes illustrated in panels B and C. **B** and **C,** Airway cilia have the classic microtubular arrangement of motile cilia, namely, nine peripheral doublets and two central singlets. Microvilli are randomly distributed among the cilia. (Human lung surgical specimen, transmission electron microscopy.) (From Murray JF, Nadel JA: *Textbook of respiratory medicine,* ed 4, Philadelphia, 2005, Elsevier.)

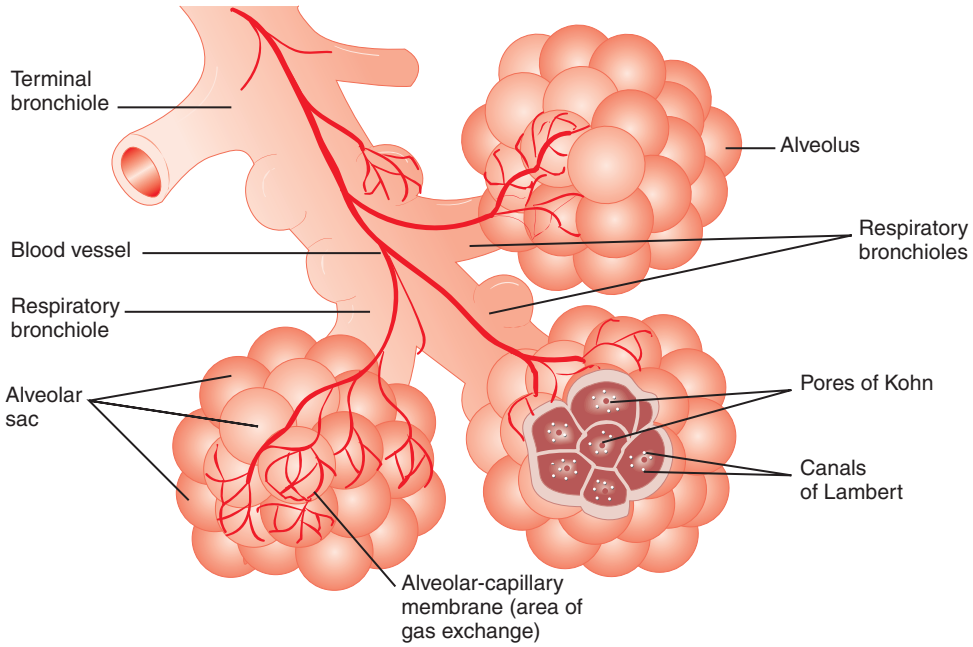


FIG 21.5 A portion of the lower respiratory tract, including a terminal bronchiole, respiratory bronchioles, and alveoli, where interchange of O_2 and CO_2 occurs between the thin walls of the alveoli and the capillary membrane.

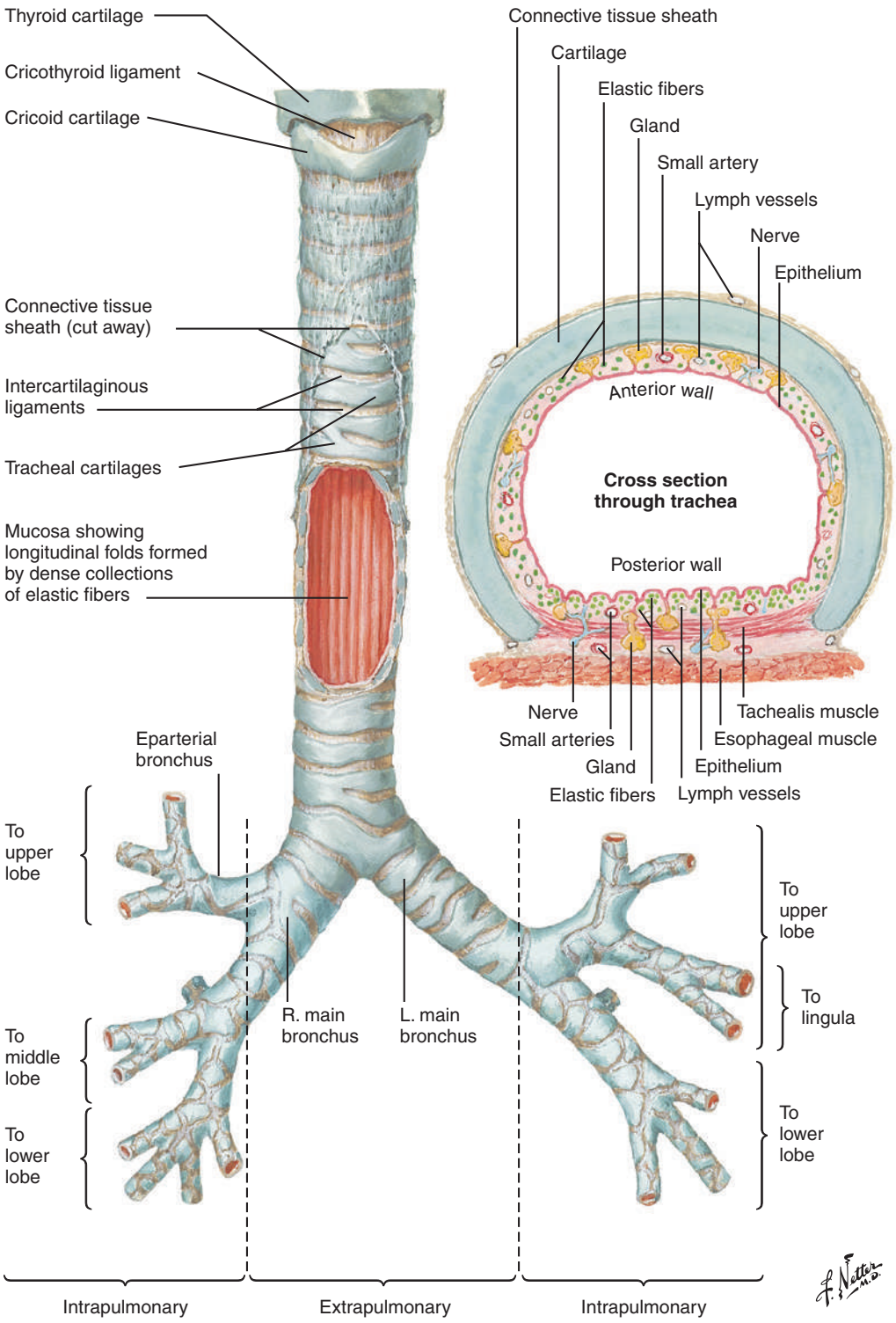


FIG 21.6 Anterior diagram of the trachea and major bronchi. *L,* Left; *R,* right. (Netter illustration from www.netterimages.com. Copyright © Elsevier, Inc. All rights reserved.)

The partial pressures of gases in the alveoli are termed PAo_2 for oxygen and $PAco_2$ for carbon dioxide. The partial pressures of gases in the blood are termed Pao_2 for oxygen and $Paco_2$ for carbon dioxide. Collateral alveolar ventilation can also occur through holes in the alveolar walls, called the *pores of Kohn* or *canals of Lambert.* A small child has less collateral ventilation because of fewer pores of Kohn. The alveolar membrane is thicker in the neonate and reaches the adult thinness of 0.5 mm by the age of 8 years. This thinner membrane may allow increased transfer of O_2. The healthy older adult has very thin-walled, enlarged air sacs and fewer capillaries than a younger adult. Respiratory system changes associated with normal aging are described in "Geriatric Considerations: Changes in the Respiratory System."

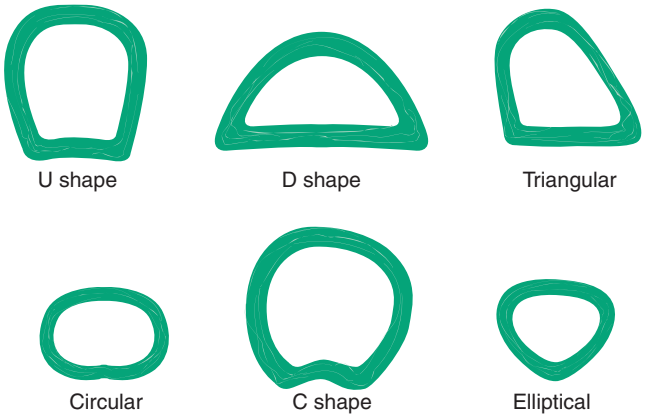


FIG 21.7 Examples of variation in tracheal shape.

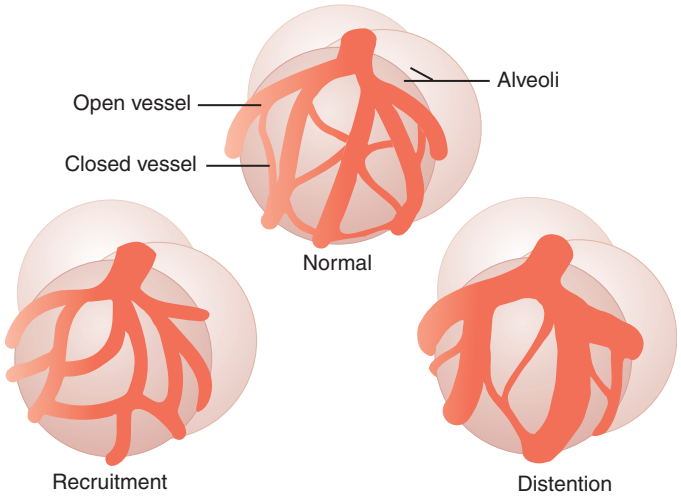


FIG 21.9 Two mechanisms for lowering pulmonary vascular resistance in capillary vessels. Recruitment allows for opening of previously closed capillaries. Distention allows for widening of capillary vessels.

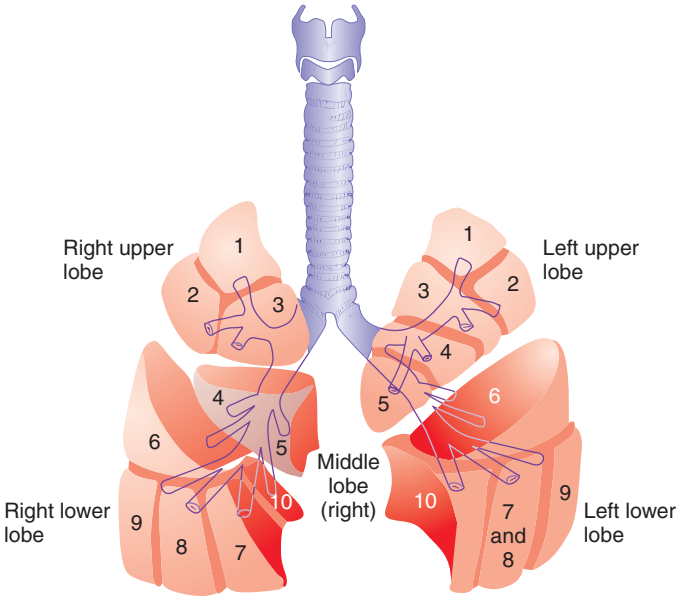


FIG 21.8 Bronchopulmonary segments of the human lung. Right and left upper lobes: *1,* apical segment; *2,* posterior segment; *3,* anterior segment. Left upper lobe: *4,* superior segment; *5,* inferior segment. Middle lobe (right): *4,* lateral segment; *5,* medial segment. Right and left lower lobes: *6,* superior (apical) segment; *7,* medial basal segment; *8,* anterior basal segment (on left, *7* and *8* combine to form the anteromedial basal segment); *9,* lateral basal segment. (*10,* posterior basal segment visible in medial view; not shown here.)

Pulmonary Circulation

Blood supply to the lungs comes from two sources: the bronchial artery system, which supplies a small amount of oxygenated blood to the pleura and lung tissues, and the pulmonary artery system, which provides a vast capillary network for O_2 and CO_2 exchange. The capillary networks of the neonate, young child, and elderly person are less than those in the average healthy adult. Oxygen-depleted (unoxygenated) blood leaves the right ventricle by way of the pulmonary artery trunk, which branches into the right and left pulmonary arteries. The pulmonary arteries further divide into smaller arteries and arterioles that feed into the capillary network where gas exchange occurs from the alveolar-capillary membrane. Pulmonary artery blood is unoxygenated, and blood in the pulmonary veins is oxygenated. The opposite is true in the rest of the body, where the arterial blood is oxygenated and the venous blood is unoxygenated.

The capillary network is a low-pressure system that can expand two to three times the normal size before a significant increase in pulmonary capillary pressures is detectable. The normal pulmonary arterial pressure in a healthy adult is about 22/8 to 25/8 mm Hg. The mean pulmonary arterial pressure is approximately 15 mm Hg. This compares with the high pressure of the systemic circulation, which is normally considered to be 120/80 mm Hg, with a mean arterial pressure of 96 mm Hg.

Under normal resting conditions, some pulmonary capillaries are closed and not perfused (filled with blood), and some pulmonary capillaries are open with no blood perfusing through them. The pulmonary circulation has two mechanisms for lowering pulmonary vascular resistance when vascular pressures are increased because of increased blood flow (Fig. 21.9). The first mechanism is recruitment, which allows opening of previously closed capillary vessels. The second mechanism is distention, which allows for increase in caliber of capillary vessels, allowing more blood to circulate through.

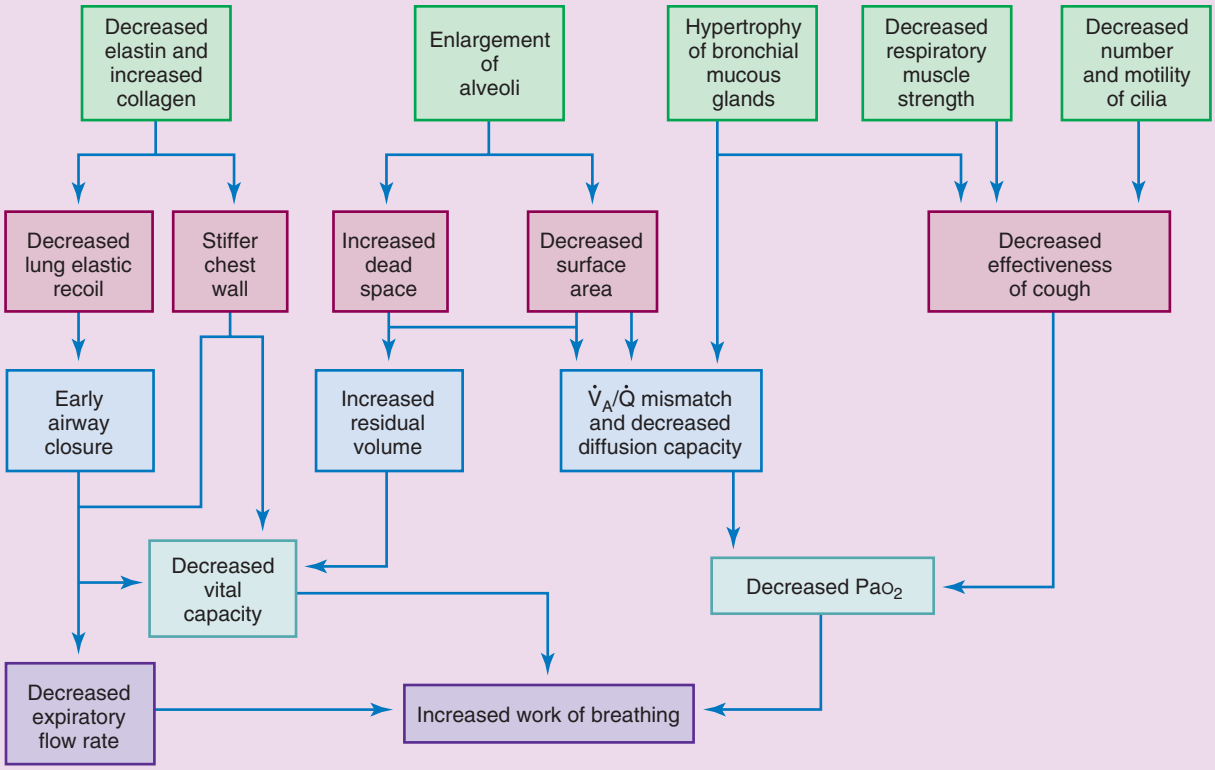
Another factor influencing pulmonary circulation is the fluid balance of the lung tissues. Fluid balance is regulated by the hydrostatic pressure, colloid osmotic pressure, and capillary permeability. When capillary hydrostatic pressure exceeds colloid osmotic pressure, fluid moves from the capillary to the interstitium. If the fluid shift is not controlled, the fluid volume will continue to increase until fluid is moved into the alveoli. Alveolar edema is more serious than interstitial edema (fluid in the interstitial space), because of its negative effects on gas exchange. Pulmonary interstitial and alveolar edema is common in disease processes such as congestive heart failure and infectious diseases of the lung. Other disease processes that also increase capillary permeability are acute respiratory distress syndrome (ARDS) and infant respiratory distress syndrome. (See Chapter 23 for further discussion.)

Age-Related Variations

Structural and physiologic variations occur at each end of the age continuum. A summary of anatomic and physiologic respiratory variations by age group is presented in Table 21.1. Pediatric considerations are shown in the following box.

GERIATRIC CONSIDERATIONS

Changes in the Respiratory System



With aging, the result of all pulmonary changes is an increase in the work of breathing. The lungs show a reduction in the amount of elastin and an increase in collagen concentration, leading to decreased elastic recoil and increased compliance (stiff chest wall and early airway closure). These changes lead to increased residual volume and early airway closure. The chest wall becomes stiffer or more rigid as a result of rib and cartilaginous calcification. The strength of the diaphragm, intercostal muscles, and accessory muscles declines. The stiff chest wall and diminished respiratory muscle strength cause other functional changes, including an increase in dead space and decreased expiratory flow rates and vital capacity.

There is a reduction in the number and motility of cilia, resulting in a decrease in respiratory clearance. There is an increase in and hypertrophy of bronchial mucous glands. The decreased respiratory muscle strength, increased mucus production, increased chest wall stiffness, and loss of cilia together reduce cough effectiveness.

Within the lungs, there is enlargement of alveoli and respiratory bronchioles with subsequent decreased surface area. The arterial blood flow through the pulmonary vessels decreases proportionally to changes in cardiac output. The loss of elastic recoil causes the enlarged respiratory bronchioles to collapse or close before the alveoli empty. Alveolar enlargement, along with reduced pulmonary artery blood flow and early airway closure, lowers diffusion capacity and the amount of gas exchange. It also increases air trapping and residual volume.

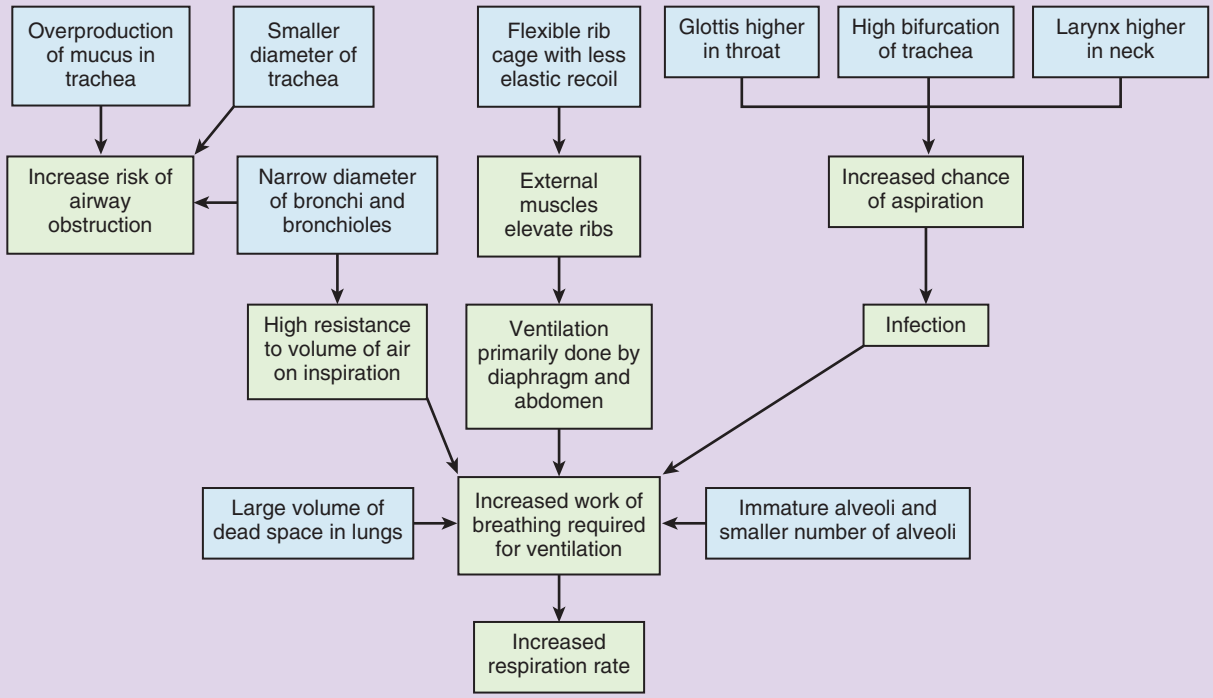
Because of chest wall stiffness and lung rigidity, apical ventilation increases in the elderly, whereas basilar ventilation decreases. Ventilation–perfusion mismatch occurs as a result of increasing apical ventilation with poor apical capillary blood flow. The result of these changes leads to reduced arterial oxygen pressure (Pa_{O_2}). Because of increased ventilation–perfusion mismatch, the Pa_{O_2} may decrease when the elderly individual reclines.

KEY POINTS

- The respiratory system development begins at about day 26 of gestation. Abnormal development of the septum during this time can lead to tracheo-esophageal fistula. At 25 weeks' gestation, the fetal lungs have developed sufficiently to allow respiration, although alveolar development and surfactant production are just beginning.
- The upper airway includes the nasopharynx, oropharynx, and laryngopharynx. The primary functions of the upper airway are to warm, filter, and humidify inspired air.
- The lower airway includes structures below the larynx: the trachea, bronchi, bronchioles, and alveoli. The larynx functions to prevent aspiration during swallowing and is the location of the vocal cords.
- The trachea, bronchi, and bronchioles serve as conducting passageways for air. They do not engage in gas exchange. Sympathetic influence on these airways causes relaxation (by means of β_2-adrenergic receptors), and parasympathetic influence causes constriction (by means of acetylcholine receptors).
- Exchange of respiratory gases occurs in the alveoli. The epithelial cells that comprise the alveoli are called *type I cells* (type I pneumocytes). Type II pneumocytes produce surfactant in the alveoli. The grapelike structure of the alveoli provides a huge surface area for gas exchange.
- The upper and lower airways are lined with cilia, which move rhythmically to transport mucus and trapped debris out of the respiratory tree. Ciliary function is impaired by a number of factors, including smoking, alcohol consumption, low humidity, and anesthesia.
- The lungs are perfused by two sources: bronchial arteries bring a small amount of oxygenated blood to nourish lung tissues; pulmonary arteries transport the entire cardiac output of the right ventricle to the alveoli for gas exchange.
- The lung has a large reserve capacity for gas exchange. At rest, some of the pulmonary capillaries are not perfused. During periods of high lung blood flow (such as high cardiac output during exercise), previously unperfused capillaries are recruited, and already perfused capillaries become distended.
- Filtration of fluid through pulmonary capillaries is influenced by hydrostatic pressure and colloid osmotic pressure in the same way as other capillaries. Excessive filtration can lead to pulmonary edema, which interferes with normal gas exchange.

PEDIATRIC CONSIDERATIONS

Changes in Respiratory System in Children



The respiratory system in children is very different from that of the adult, which makes the child susceptible to airway obstruction, aspiration, and infection. The trachea in the infant has more mucus-producing glands, which can create an overproduction of mucus in the infant. The trachea, bronchi, and bronchioles are also smaller, with a narrower diameter. The excess mucus and narrower respiratory structures increase the risk of airway obstruction in the child. The positioning of respiratory structures in the infant increases the chance of aspiration. The glottis is higher in the throat of an infant compared with a 5-year-old child (MacGregor, 2008). The trachea bifurcates at the third thoracic vertebra compared with the sixth in adults (MacGregor, 2008), and the larynx is located higher in the neck of the infant. Aspiration can lead to infection in the child and increase the work of breathing.

The infant has to work harder for ventilation of the lungs because of several factors. The narrower diameter of bronchi and bronchioles creates a higher resistance to volume of air on inspiration. The large volume of dead space in the lungs requires the infant to breathe faster to meet oxygen demands. Compared with an adult, the alveoli are smaller and immature, which decreases the area for gas exchange to occur in the lungs. The number of alveoli and the size increase as the child ages. The flexible ribcage is unable to support the lungs adequately because it has less elastic recoil. The intercostal muscles of the ribcage also work inefficiently. The external intercostal muscles elevate the ribs for inspiration, whereas the internal intercostal muscles cannot lift the chest wall and do not help with inspiration. The infant depends on the diaphragm and abdomen for ventilation to compensate for the lack of intercostal muscle strength. All of these factors increase the work of breathing required for ventilation, and the infant compensates by increasing his or her respiratory rate. By age 8 the lungs are fully developed, and the child's respiratory system begins to resemble the adult's respiratory system (MacGregor, 2008).

From MacGregor J: *Introduction to the anatomy and physiology of children: a guide for students of nursing, child care and health,* ed 2, New York, 2008, Routledge.

TABLE 21.1 Variations in Anatomy and Physiology of the Respiratory System by Age Group

	Young Newborns	Children	Adults	Elderly (Over 60 Years Old)
Anatomic dead space	Proportional to size	Proportional to size	≈150 mL	≈150–200 mL
Number of alveoli	12.5%–16.5% of adult number	Adult number by 8 years old	300,000/lung	≤300,000/lung
Thickness of alveolar membrane	Thicker than adult	Adult by 8 years old	<0.5 μm	Thinner than adult
Number of capillaries	Less than adult	Adult by 8 years old	Adult	Less than adult
Vital capacity	Proportionately less than adult	Proportional to size	4.8 L	Less than adult
Tidal volume	Proportional to size	Proportional to size	500 mL	Less than adult (30% less by age 80 years)
Compliance	More compliant than adult	Similar to adult	Static compliance (90–100 mL/cm H_2O)	Less compliant
Airway resistance	Greater than adult	Greater than adult	1.0–1.5 cm H_2O/L/sec	Adult level or less

Data from Fretwell ME: Aging changes in structure and function. In Carnevali DL, Patrick M, editors: *Nursing management for the elderly,* ed 3, Philadelphia, 1993, Lippincott.

VENTILATION

Lung Volumes and Capacities

Ventilation is the process of moving air into the lungs and distributing air within the lungs to gas exchange units (alveoli) for maintenance of oxygenation and removal of carbon dioxide (CO_2). Measures of ventilation (amount of air moved) include four lung volumes and four lung capacities. Fig. 21.10 schematically presents the various lung volumes and capacities; Table 21.2 defines each term and provides further details.

Lung volumes and capacities vary according to the individual's body size, age (decreased in the neonate, young child, and the elderly), and body position (supine versus upright). Testing of pulmonary function to measure these volumes and capacities is covered under the "Diagnostic Tests" section later in this chapter. Other measures important to ventilation are dead space, minute ventilation, and alveolar ventilation.

Dead Space

Dead space includes three dimensions: anatomic dead space, alveolar dead space, and physiologic dead space. Anatomic dead space includes the volume of gas (not used in gas exchange) in the conducting airways from the nose to the respiratory bronchioles. Generally, in adults this area is equal to 1 mL per pound of *ideal* body weight, or approximately 150 mL. In newborns and young children, the anatomic dead space is proportionately larger for their size. The anatomic dead space of elderly persons may increase slightly over that of healthy young adults because of the loss of alveolar sacs. Alveolar dead space is composed of ventilated, but unperfused, areas of the lung, and is often referred to as *wasted ventilation.* Physiologic dead space (functional dead space) is the sum of the anatomic dead space and alveolar dead space. Approximately one third of each breath occupies dead space.

Minute Ventilation

Minute ventilation is the product of tidal volume (milliliters of air inhaled with each breath) times respiratory rate per minute. For example, a person with a tidal volume of 500 mL who is breathing at a rate of 15 breaths/minute has a minute ventilation of 7500 mL (see Table 21.2 for typical volumes).

TABLE 21.2 Lung Volumes and Capacities

Term	Definition (Typical Volume)
Lung Volumes	
Tidal volume	A normal breath (≈500 mL) or amount of gas entering or leaving lung during normal breathing
Inspiratory reserve volume	Amount of gas a person is able to inspire above a normal breath (e.g., maximal deep breath, ≈3 L)
Expiratory reserve volume	Amount of gas expired beyond tidal volume (≈1.2 L)
Residual volume	Volume of gas left in lungs at end of a maximal expiration (≈1.2 L)
Lung Capacities	
Vital capacity	Total volume of gas that can be exhaled during maximal expiration (≈4.8 L). Includes inspiratory reserve volume, tidal volume, and expiratory reserve volume
Inspiratory capacity	Amount of gas that can be inspired from a resting expiration (≈3.5 L)
Functional residual capacity	Amount of gas left in lungs at end of a normal expiration (≈2.4 L): includes expiratory reserve volume and residual volume
Total lung capacity	Amount of gas contained in lungs at maximal inspiration (≈6.0 L)
Forced expiratory flow rate (FEF25, FEF50, FEF75)	Volume of air forcibly exhaled per unit time (liters per second or liters per minute) at 25%, 50%, and 75% of forced vital capacity (FVC)
Peak expiratory flow rate	Highest rate of flow sustained for 10 msec or more at which air can be expelled from lungs

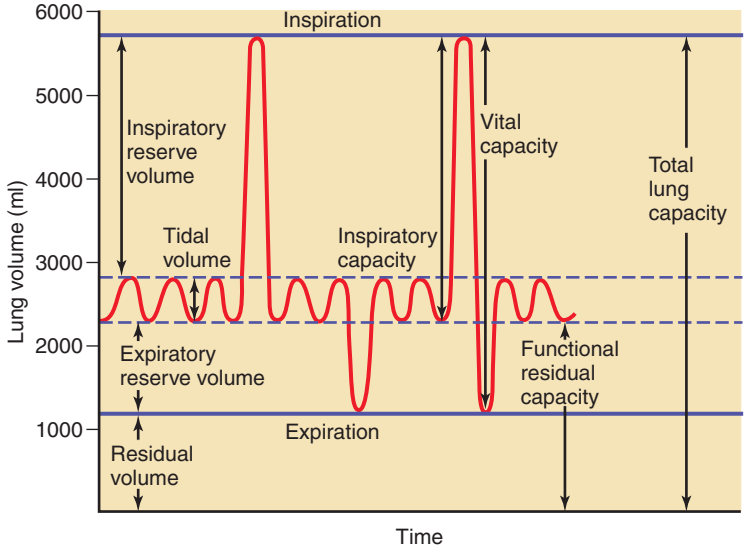


FIG 21.10 Schematic representation of the various lung volumes and capacities for a healthy adult (see also Table 21.2).

Alveolar Ventilation/Oxygenation

By comparison, alveolar ventilation ($\dot{V}_A$) equals the difference between tidal volume (V_T) and anatomic dead space volume (V_D) multiplied by the respiratory rate (RR) per minute.

$$\text{Alveolar ventilation } (\dot{V}_A) = (V_T - V_D) \times RR$$

Because alveolar ventilation is affected by both the anatomic dead space and the respiratory rate, slow deep breathing yields greater alveolar ventilation than does rapid shallow respiration. The patient breathing 25 times/minute at a V_T of 200 mL would have alveolar ventilation as follows:

$$(200\text{ mL} - 150\text{ mL [anatomic dead space]}) \times 25\text{ breaths/minute} = 1250\text{ mL}$$

A patient breathing 10 times/minute at a V_T of 600 mL would have an alveolar ventilation of

$$(600\text{ mL} - 150\text{ mL}) \times 10 = 4500\text{ mL}$$

The partial pressure of oxygen in the alveoli (P_{AO_2}) is the driving force to move O_2 into the blood and is estimated with the following equation:

$$P_{AO_2} = F_{IO_2}\,(P_B - 47) - (Paco_2 \div 0.8)$$

where F_{IO_2} is the fraction of inspired oxygen, P_B is the barometric pressure, 47 is the constant for water vapor pressure (mm Hg), 0.8 is the respiratory quotient, and $Paco_2$ is the laboratory measurement of arterial CO_2 pressure (mm Hg).

The value for P_{AO_2} is normally very close to that for Pao_2. The difference between alveolar and arterial oxygen tensions is called the *A*lveolar-*a*rterial *D*ifference in oxygen ($A - aDo_2$). A large $A - aDo_2$ value indicates poor matching of alveolar ventilation with alveolar blood flow ($\dot{V}_A/\dot{Q}$ matching).

For example, the calculation of $A - aDo_2$ for a person at sea level ($P_B = 760$ mm Hg) breathing room air ($F_{IO_2} = 0.21$) with $Pao_2 = 75$ mm Hg and $Paco_2 = 40$ mm Hg is

$$P_{AO_2} = 0.21(760 - 47) - (40/0.8) = 150 - 50 = 100$$

Therefore using the arterial blood gas value obtained for the Pao_2 and the calculated P_{AO_2} of 100, a difference of 25 mm Hg is determined:

$$A - aDo_2 = 100 - 75 = 25\text{ mm Hg}$$

This large of a difference indicates a significant problem with gas exchange.

In critical care settings, it is useful to calculate the $A - aDo_2$ value to monitor the efficacy of oxygen exchange across the lung. The normal $A - aDo_2$ gradient in a normal, healthy young adult is less than 10 mm Hg at room air, but it increases with age and increasing F_{IO_2}. A rising $A - aDo_2$ value indicates worsening lung function, even though hypoxemia (Pao_2 lower than 80 mm Hg at sea level) may not necessarily be present.

Hypoxemia that is primarily caused by hypoventilation suggests that the lung is normal, and treatment that increases ventilation will remedy the problem. This type of hypoxemia is characterized by a normal $A - aDo_2$ value.

A simple method of calculating expected P_{AO_2} is the "law of 5's." By multiplying the F_{IO_2} (%) by 5, the care provider has an estimate of what the oxygen level should be under normal, healthy conditions (e.g., $5 \times 21\%$ room air = 105).

Mechanics of Breathing

The mechanics of breathing include the concepts of airway resistance, lung compliance, and opposing lung forces (elastic recoil versus chest wall expansion) of the lung. These factors affect the overall performance of gas exchange and the work of breathing.

The lungs have a natural recoil tendency, whereas the chest wall favors the expanded state. During inspiration, the chest wall muscles (external intercostals) contract, elevating the ribs as the diaphragm moves downward. These two actions create a negative intrapleural pressure that causes the lung to expand. During expiration, the lung deflates passively because of the elastic recoil (elastic fibers in the lung tissue) and relaxation of the diaphragm. During heavy breathing, as seen with exercise, the elastic forces are not strong enough to cause the necessary rapid expiration, so abdominal muscles contract, pushing the abdominal contents upward, compressing the lungs. Fig. 21.11 shows the interaction of lung forces during inspiration and expiration. In the normal, healthy resting individual, expiration is accomplished almost entirely by relaxation of the diaphragm. At the end of a normal expiration, the alveoli still contain some air volume, known as the *functional residual capacity.* If the alveoli were allowed to empty completely, the high surface tension in the alveoli would make it more difficult to reinflate them and add significantly to the work of breathing. In the absence of surfactant, which reduces alveolar surface tension, the alveoli tend to collapse—a condition called *atelectasis.* Excessive surface tension can increase the work of breathing so much that mechanical ventilation may be required. This is often the case in ARDS and in infant respiratory distress syndrome (see Chapter 23).

Airway Resistance

Airway resistance is determined by the relationship between driving pressure and flow. It is influenced by airway radius and the pattern of gas flow. Resistance increases as the radius of the airway tube decreases. Resistance is calculated by the following formula:

$$\text{Resistance} = \text{driving pressure} \div \text{rate of airflow}$$

The radius of the airway decreases from the trachea to the terminal bronchioles. As mucus builds up in the airway, the passage is narrowed, and resistance to airflow increases. Other factors affecting airway resistance include stress, pulmonary conditioning, and age.

The trachea and bronchi contain cartilage and small amounts of muscle. The cartilage assists in maintaining airway passage stability, thus preventing airway collapse. The bronchioles and terminal bronchioles do not contain cartilage, but have increased amounts of smooth muscle that are innervated by the autonomic nervous system. Stimulation of cholinergic fibers leads to bronchoconstriction. Stimulation of the β_2-adrenergic receptors leads to bronchodilation. The bronchial muscles function to maintain an even distribution of ventilation. A circadian rhythm is associated with bronchial tone, with maximal bronchodilation occurring at about 6 PM and maximal bronchoconstriction occurring at 6 AM.

Airway resistance is also affected by the pattern of gas flow (Fig. 21.12). Air movement from the nasal cavity through the large bronchi occurs by turbulent flow, which creates friction and increases resistance. Bronchospasm in the smaller airways and high gas flow also create turbulent flow. Laminar flow occurs in the small airways of the lung and creates minimal resistance to airflow. Transitional flow (mixed

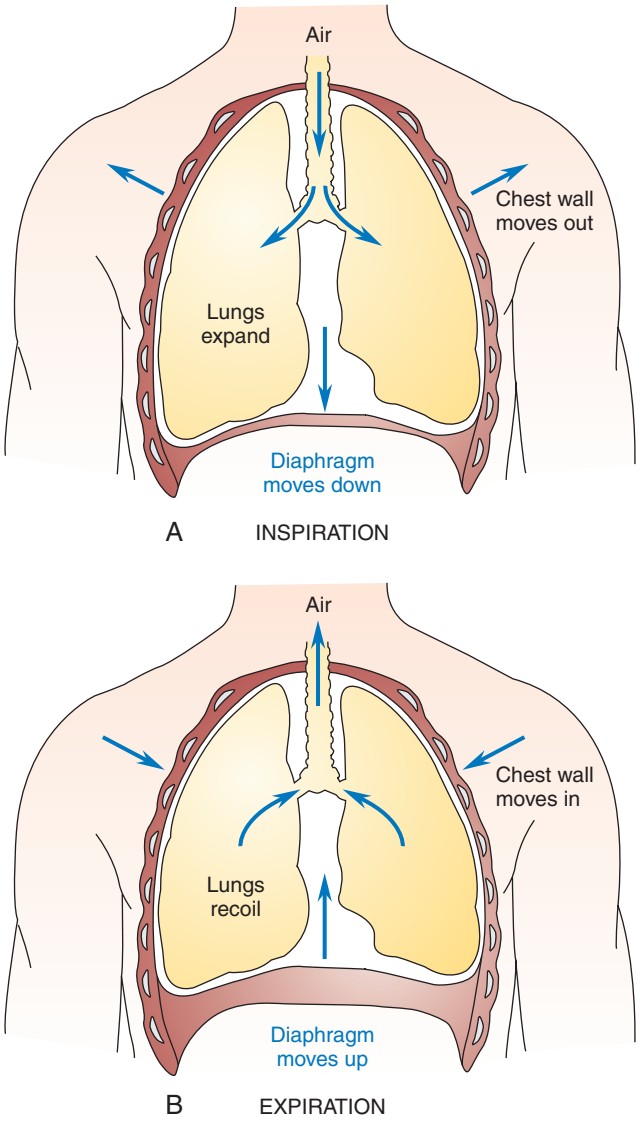


FIG 21.11 Lung forces during inspiration and expiration. **A,** During inspiration, the respiratory muscles contract, the chest wall expands, and air flows into the lungs. **B,** During expiration, the respiratory muscles relax, the lungs recoil, and air flows passively out of the lungs.

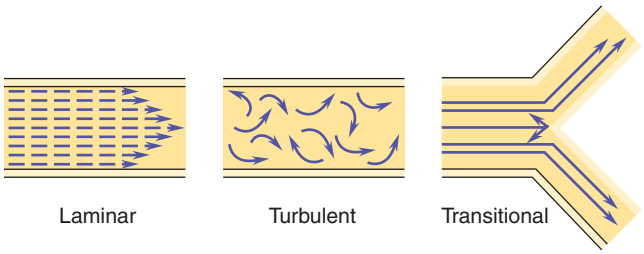


FIG 21.12 Patterns of gas flow.

pattern of flow) occurs in the larger airways, especially at bifurcations. The highest airway resistance is at the nose because of turbulent flow with high velocities of airflow. Airway resistance is even higher in the newborn than in the adult and continues to be greater than that of the adult up to the age of 5 years. Resistance changes very little in the elderly lung.

Lung Compliance

Lung compliance is another factor that influences the work of breathing. Compliance represents lung expandability and the ease of lung inflation. It is best illustrated by the effort required to blow up a new balloon compared with blowing up a balloon that has been inflated many times before. It is a measure of the relationship between pressure and volume. It is represented by the formula:

$$\text{Compliance} = \text{change in volume} \div \text{change in pressure}$$

Two factors associated with compliance are chest wall expandability and lung expandability. Lung compliance can be measured in the static (motionless) or dynamic state. Effective static compliance is determined by dividing the pressure required to deliver a volume of gas by the tidal volume as delivered by a ventilator. A more accurate measurement of compliance requires the insertion of an esophageal balloon. Normal static compliance in a healthy young adult would be 90 to 100 mL/cm H_2O.

Compliance provides an estimate of airway resistance and elasticity. Lung compliance is increased in neonates and children younger than 3.5 years, because of their chest wall flexibility. Lung compliance may decrease in the elderly because of increasing chest wall rigidity from calcification of costal cartilages, reduced mobility of ribs, and partial contraction of inspiratory muscles. Changes in the thoracic vertebrae and intervertebral disks also lead to decreased expansion of the chest wall in the elderly. Disease processes that make the lung stiffer and decrease respiratory function include pneumonia, pulmonary edema, atelectasis, ARDS, and pulmonary fibrosis. Other factors that decrease compliance by decreasing chest wall distensibility are obesity, abdominal distention, pregnancy, kyphoscoliosis, and abdominal surgery (attributable to decreased respiratory effort from surgical pain). Lung compliance may be increased by loss of the lung's elastic fibers that occurs with age and obstructive lung diseases. An abnormally high lung compliance, with loss of elastic recoil, increases the work of breathing by requiring greater effort to expel air from the lungs during exhalation.

Distribution of Ventilation

Distribution of ventilation is affected by body position. In the upright individual, the alveoli at the apices (top) of the lung are much larger than those at the base. Fig. 21.13 shows the variation in structural size of alveoli at the apex compared with that at the base. In the healthy upright individual, ventilation is greatest near the bottom of the lung and decreases toward the apices. The regional differences are less in a supine person. The greater lung expansion at the bases results from a greater compliance of the alveoli at the bases and the downward displacement of the diaphragm, which expands the lower lobes more than the upper lobes. When an individual is in the supine lateral position, ventilation is best in the dependent part of the lung fields, but the difference is not as great as that seen in the upright lung.

Neurologic Control of Ventilation

Respiration is influenced by a number of factors. These include neural control centers, chemoreceptors, lung receptors, proprioceptors, and pressure receptors. The factors that regulate respiration are reviewed in this section.

Neural control of the respiratory system is located in the medulla oblongata and the pons, which is commonly referred to as the *respiratory center.* Efferent nerve impulses travel from the brainstem by way of the phrenic nerve to the diaphragm to stimulate muscular contractions for inspiration.

The medullary respiratory center within the brainstem consists of two groups of widely dispersed neurons that function as a unit to regulate

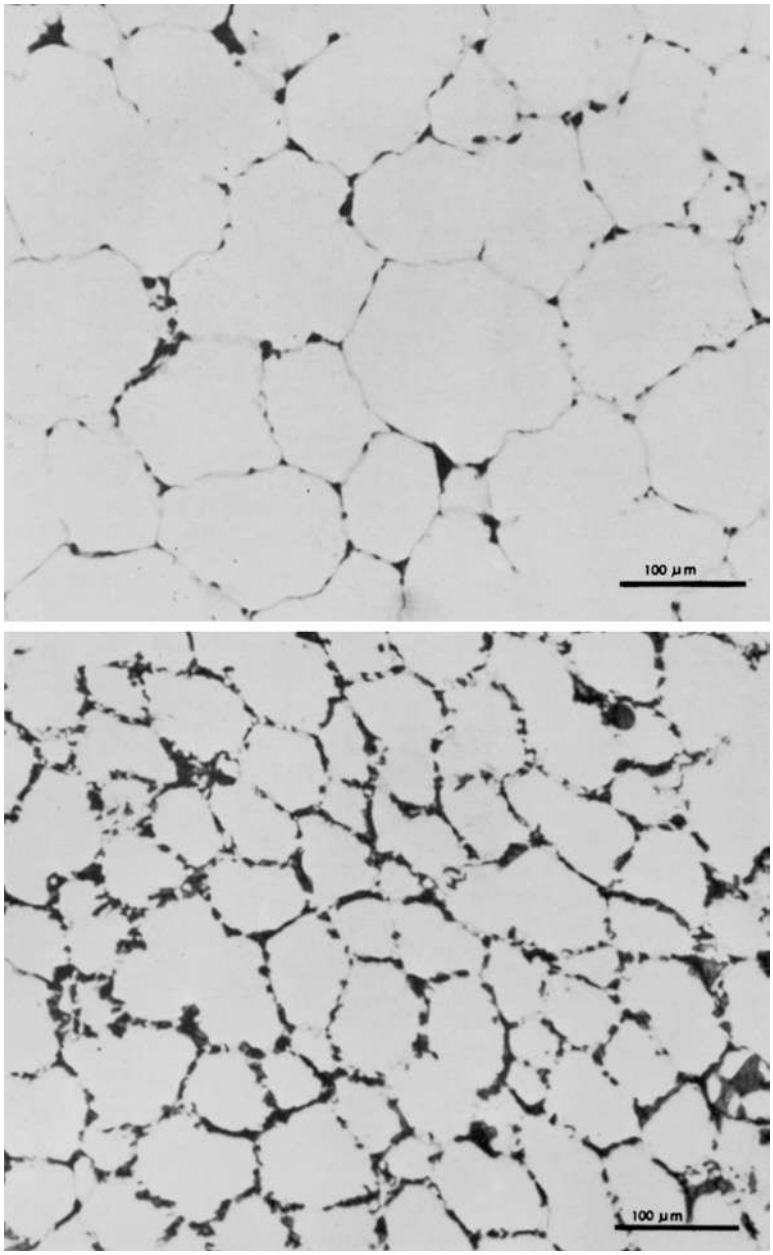


FIG 21.13 Sections of lung from the apex *(upper panel)* and 20 cm below the apex *(lower panel)* obtained from a greyhound dog lung (specimens frozen in a vertical position). The upper panel illustrates alveoli in the apex *(zone 1)* of the lung in the upright position: the air sacs are large, and blood flow is diminished. The lower panel represents the base of the lung zone with optimal ventilation and perfusion (×188). (From Murray JF: *The normal lung,* ed 2, Philadelphia, Saunders, p. 110. Courtesy Jon B. Glazier, MD.)

breathing. The dorsal respiratory group of neurons transmits impulses that stimulate inspiratory muscles (in the intercostals and diaphragm). The impulses are generated in increasing fashion, termed a *ramp signal.* Impulses begin slowly and increase steadily for about 2 seconds. Abrupt cessation of signals for 3 seconds allows for expiration, and then the cycle begins again. This system establishes the basic respiratory rhythm. Fig. 21.14 provides a schematic diagram of these interactive mechanisms on respiratory control.

The pneumotaxic center of the upper pons (see Fig. 21.14) appears to influence the rate of respiration and ends inspiration by inhibition of the inspiratory ramp. In addition, input from the spinal cord, cortex, and midbrain contributes to the normal smooth pattern of respiration. The apneustic center of the lower pons (demonstrated to exist in dogs) influences the pattern of respiration and may function to provide an extra driving force for the inspiratory neurons, thus prolonging inspiration.

Sensory inputs to the respiratory control center include central chemoreceptors, peripheral chemoreceptors, Hering–Breuer stretch receptors, proprioceptors, baroreceptors, and environmental sensations.

The central chemoreceptors within the medullary center respond to changes in CO_2 level and pH. A stimulus to breathe occurs when a small increase in arterial carbon dioxide tension ($Paco_2$) leads to stimulation of chemoreceptors. Alveolar ventilation can increase 10-fold with an acute rise in $Paco_2$.

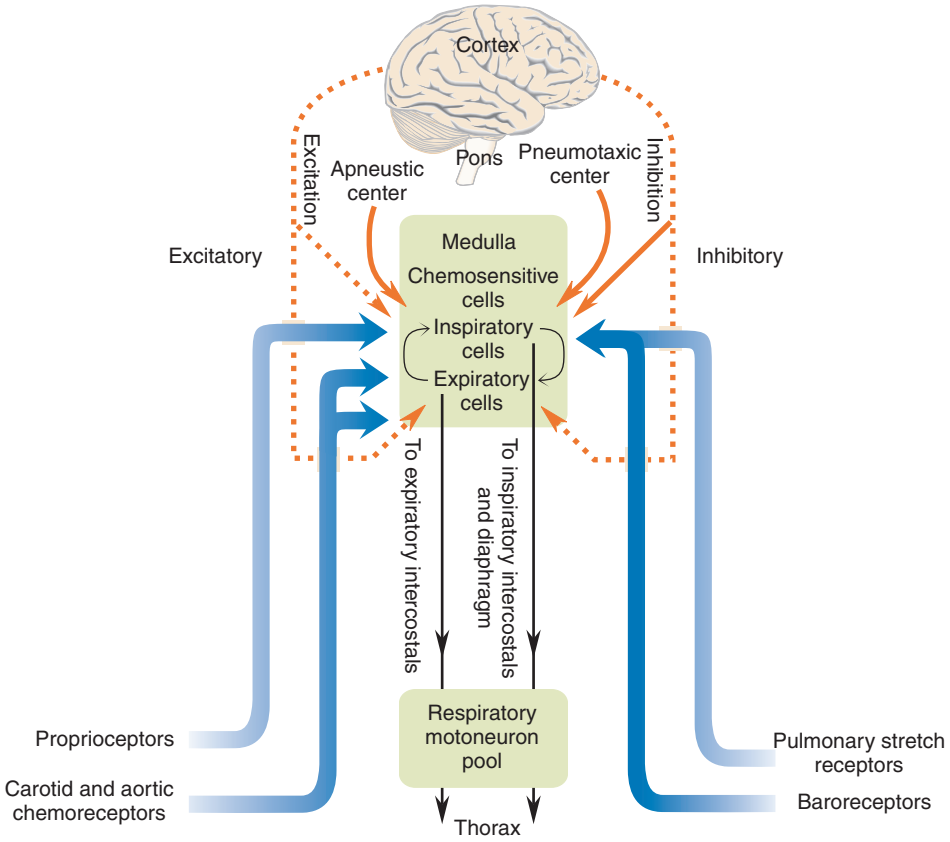


FIG 21.14 Interactive mechanisms influencing control of respiration.

The peripheral chemoreceptors located in the aortic arch and carotid bodies respond primarily to decreases in arterial O_2 concentration. Increases in the hydrogen ion concentration (decreased pH) or the $Paco_2$ also stimulate peripheral chemoreceptors; however, the response of peripheral chemoreceptors to a change in $Paco_2$ is minor compared with central chemoreceptors.

The Hering–Breuer reflex involves stretch receptors located in the alveolar septa, bronchi, and bronchioles. Inflation of the lung initiates the response that sends neuronal impulses up the vagus nerve to the medulla to cause inhibition of inspiration. Therefore the rate and duration of inspiration are affected. This reflex is primarily seen in neonates (less so in adults) and at high tidal volumes (greater than 1500 mL) and prevents overinflation of the lung.

Proprioceptors located in the muscles and tendons of movable joints respond to body movement (exercise). Body movement, such as with exercise, leads to stimulation of respiration (rate and depth) to maintain oxygen levels.

Baroreceptors located in the aortic arch and carotid arteries respond to changes in blood pressure. The aortic arch transmits impulses through the vagus nerve, and the carotid bodies transmit impulses through the glossopharyngeal nerve. An increase in arterial blood pressure leads to inhibition of respiration. A decrease in mean arterial blood pressure below 80 mm Hg leads to stimulation of respiration.

Environmental factors also influence respiration. Individuals demonstrate changes in respiration related to such factors as a cold shower, a pin prick, stress, or airway irritation from air pollution and smoking. Infection and fever also increase the respiratory rate. During normal breathing, energy expenditure is only 3% to 5% of total energy expenditure. During exercise and in patients with high airway resistance energy expenditure can increase to 50%.

KEY POINTS

- Approximately one third of each breath occupies areas of the lung that do not engage in gas exchange. Total (physiologic) dead space includes the anatomic dead space of the bronchial tree and the dead space of unperfused alveoli.
- Alveolar ventilation may be severely compromised in persons with small tidal volumes or increased dead space. When tidal volume is not significantly greater than dead space, increased respiratory rate is not effective in restoring alveolar minute ventilation.
- To move air into the lungs, the respiratory muscles generate a negative intrapleural pressure that causes air inflow owing to the pressure gradient between the atmospheric pressure at the mouth (zero pressure) and the alveolar pressure (negative pressure).
- Airways and tissues of the lung resist inflation. Resistance is provided by the airways, elastic fibers in the lung, and surface tension in the alveoli. The degree of resistance can be estimated by measuring overall lung compliance. A compliant lung requires minimal pressure to accomplish a large increase in volume; a noncompliant (stiff) lung requires the generation of high pressure to inflate the lung.
- Airway resistance is primarily determined by the diameter of the airways. Airway constriction greatly increases airway resistance. Parasympathetic stimulation of the airways results in constriction; sympathetic (β_2) stimulation results in dilation.
- Elastic fibers in the lung are stretched during inspiration, then recoil passively to achieve expiration. Destruction of elastic fibers increases lung compliance; excessive fiber production (fibrosis) decreases lung compliance.
- High surface tension in the alveoli causes the surfaces to adhere, making inflation more difficult. Surfactant functions to reduce surface tension. A lack of surfactant makes the lungs more difficult to inflate (decreased compliance).

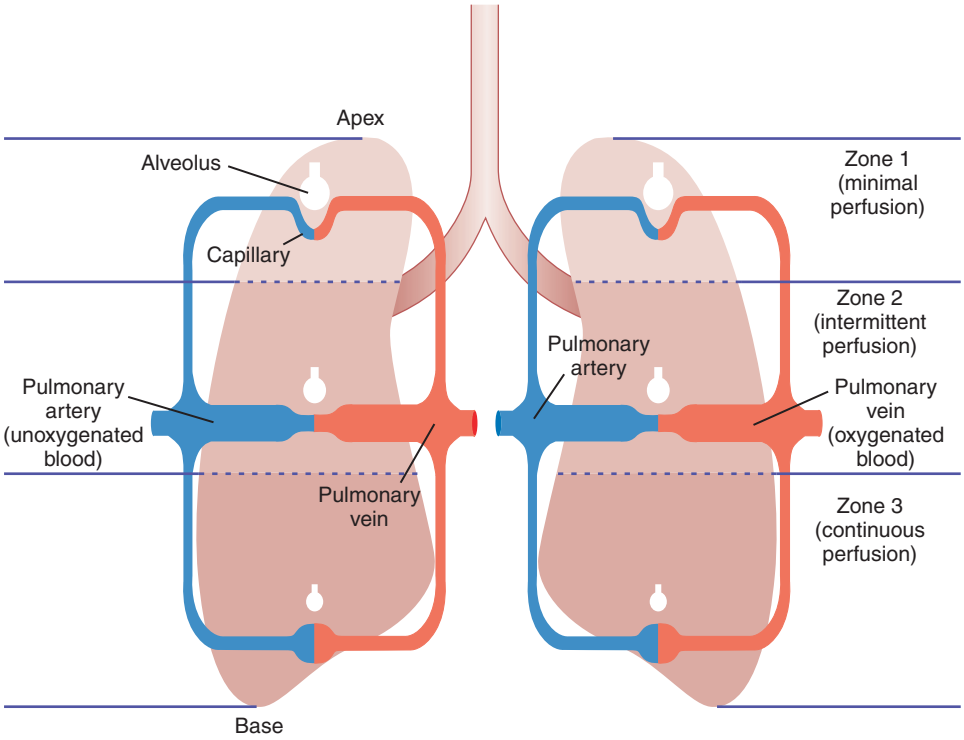


FIG 21.15 Schematic representation of the three lung zones in which different hemodynamic conditions govern blood flow (see text for discussion).

- The medulla oblongata and pons contain the neurons that integrate information regarding the ventilatory status of the body from chemoreceptors, proprioceptors, and stretch receptors. Respiratory neurons in the medulla initiate inspiration and establish the basic inspiratory-expiratory pattern. Pneumotactic center neurons in the pons primarily influence the rate and depth of respiration.
- Central chemoreceptors located within the medulla detect changes in pH and P_{CO_2}. Peripheral chemoreceptors are located in the aorta and carotid arteries and detect changes in arterial pH, P_{CO_2}, and P_{O_2}. An increase in P_{CO_2} or a decrease in pH or P_{O_2} stimulates ventilation.

PULMONARY BLOOD FLOW

Pulmonary Vasculature

Perfusion (blood flow) is the second process of respiration, the first being alveolar ventilation. The pulmonary circulation is a low-pressure system (25/8 mm Hg). The volume of blood in the lungs is about 450 mL, or 9% of the total amount of blood in the body. This volume can vary from one-half normal to twice normal. Unoxygenated blood from the right ventricle is pumped into the main pulmonary artery and then into its branches, which divide into capillary beds throughout lung tissue. The capillary beds surround the alveoli and allow for diffusion of O_2 and CO_2.

Distribution of Blood Flow

Distribution of blood flow (perfusion) is uneven and is affected by body position and exercise. When a person is upright, blood flow is decreased in the upper regions of the lungs (apices), compared with the lower regions (bases). When a person assumes the supine position, blood flow to the posterior dependent portion of the lung is higher than to the anterior lung, although the redistribution of blood flow is less dramatic than that seen in the upright lung.

The effect of gravity on the lung has led to the concept of lung zones. Fig. 21.15 depicts three lung zones. Zone 1 reflects blood flow in the apices of the lung. Blood flow is minimal at the apices because the enlarged alveolar sacs create an alveolar pressure that is higher than capillary pressure, leading to pulmonary capillary collapse.

Zone 2, the middle region of the lung, has a pulmonary arterial pressure greater than the pressure inside the alveoli during ventricular systole, but this may fall below alveolar pressure during diastole. Thus zone 2 is characterized by intermittent perfusion.

Zone 3 is continuously perfused throughout the entire cardiac cycle. Pulmonary arterial pressure is greater than pulmonary venous pressure, which in turn is greater than alveolar pressure. In this zone, capillary vessels are distended and vascular resistance is low.

Normally 2% of the cardiac output bypasses (right-to-left shunt) alveolar ventilation, creating a decrease in arterial oxygen pressure by 3 to 5 mm Hg. In bronchial anastomotic diseases, the amount of shunting may rise to 10% to 20%.

Ventilation–Perfusion Ratios

The discussion about distribution of ventilation and perfusion indicates that the best overall ventilation and perfusion occur in the dependent lung fields. A factor important to the concepts of ventilation and perfusion is the matching of an adequate volume of air in the alveoli to adequate pulmonary blood flow. In the ideal state, 4 L/min of alveolar ventilation is matched to 5 L/min of capillary blood flow in the lungs, creating a normal alveolar ventilation-to-perfusion ratio ($\dot{V}_A/\dot{Q}$) of 0.8 (Box 21.1). Two major physiologic factors that affect normal $\dot{V}_A/\dot{Q}$ ratio are right-to-left shunt and regional ventilation/perfusion changes. Other factors influencing the ratio are position changes, exercise, bed rest, and lung disease.

To review, in a normal person in the upright position, ventilation and perfusion are lower in the upper lung (apex) than the lower lung (base). In the apex, alveoli are large and receive limited blood flow, whereas in

BOX 21.1 Ventilation-Perfusion ($\dot{V}_A/\dot{Q}$)* Equations

Low $\dot{V}_A/\dot{Q}$ (underventilated):

$$\frac{\text{2L/min alveolar ventilation}}{\text{5L/min blood flow}}$$

Normal $\dot{V}_A/\dot{Q}$:

$$\frac{\text{4L min alveolar ventilation}}{\text{5L/min blood flow}}$$

High $\dot{V}_A/\dot{Q}$ (underperfused):

$$\frac{\text{4L min alveolar ventilation}}{\text{2L/min blood flow}}$$

*$\dot{V}_A/\dot{Q}$, Where $\dot{V}_A$ = alveolar ventilation and $\dot{Q}$ = blood flow.

the base, alveoli are smaller and allow for greater expansion of capillaries and thus more blood flow. In the apex, $\dot{V}_A/\dot{Q}$ is as much as 2.5 times the ideal value, causing a moderate degree of physiologic dead space. In the base, $\dot{V}_A/\dot{Q}$ is as low as 0.6 times the ideal value, representing lower volumes of air where the blood flow exceeds ventilation. During exercise, blood flow to the upper lung region increases dramatically, thus decreasing physiologic dead space. With bed rest, the dependent area of the lungs becomes the back region in the supine position so that blood flow is increased to that region and alveoli are smaller.

The three types of ventilation-perfusion imbalances are (1) high $\dot{V}_A/\dot{Q}$, (2) low $\dot{V}_A/\dot{Q}$, and (3) true shunt. High $\dot{V}_A/\dot{Q}$ is conceptually related to physiologic dead space and zone 1, in which the alveolar unit is ventilated but not perfused. High $\dot{V}_A/\dot{Q}$ units have a low P_{CO_2} and normal P_{AO_2} and can be viewed as a respiratory reserve, which can be used if perfusion is restored.

Low $\dot{V}_A/\dot{Q}$ is conceptually related to lower Pa_{O_2} (hypoxemia). Low $\dot{V}_A/\dot{Q}$ occurs regionally in areas where the airways are partially obstructed and airflow rates are low. Although an increase in total ventilation results in a decrease in alveolar CO_2 concentration, the increment in P_{AO_2} and O_2 content in end-capillary blood is minimal. Low $\dot{V}_A/\dot{Q}$ is responsive to treatment with oxygen because the airways are only partially obstructed, so it is possible for oxygen to enter the alveoli by diffusion.

True shunt, which is a right-to-left shunting of unoxygenated blood through the pulmonary circulation, contributes to lowering of Pa_{O_2}. Normally, in a healthy person, there is a small physiologic shunt of 2% of cardiac output because of bronchial, thebesian and other veins that bypass the alveoli. In patients with acute respiratory failure (ARF), physiologic shunt may rise to more than 50%. Although pulmonary shunt is similar to low $\dot{V}_A/\dot{Q}$ in affecting low arterial oxygen levels, true shunt is not responsive to oxygen therapy because the alveoli are collapsed or consolidated, and oxygen cannot gain entry into them. See the "Acute Respiratory Failure" section for details.

Hypoxic Vasoconstriction

Alveolar hypoxia leads to a normal compensatory hypoxic vasoconstriction of the pulmonary vessels passing through poorly ventilated portions of the lungs. Blood is diverted from areas of low alveolar oxygen concentration to areas of higher oxygen concentration as a means of compensatory adaptation. By diverting blood flow to areas of higher oxygen concentration, the negative effects on gas exchange are reduced. Low alveolar oxygen concentration leads to contraction of smooth muscle in the walls of the small pulmonary arterioles.

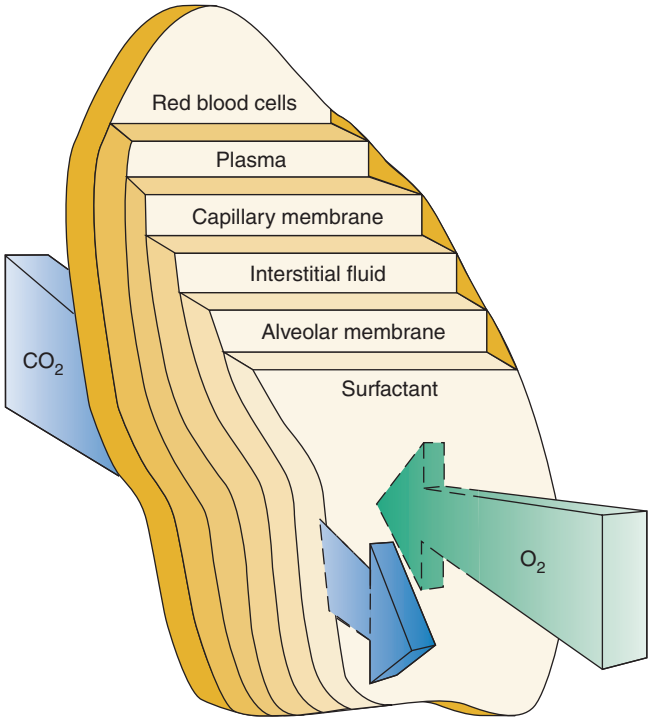


FIG 21.16 Schematic representing the six barriers through which O_2 and CO_2 must diffuse for gas exchange to occur.

KEY POINTS

- Distribution of blood flow is affected by gravity such that perfusion is greatest in dependent lung fields.
- Zones of the lung describe regional differences in perfusion. Zone 1 has no perfusion and is equivalent to dead space; zone 2 is intermittently perfused; zone 3 is continuously perfused throughout the cardiac cycle.
- Optimal alveolar-capillary gas exchange depends on matching of ventilation and perfusion at the alveolus. Abnormalities in $\dot{V}_A/\dot{Q}$ matching can result in inadequate oxygenation of the blood and insufficient CO_2 removal. Three types of regional $\dot{V}_A/\dot{Q}$ imbalance have been described: high $\dot{V}_A/\dot{Q}$ (dead space), low $\dot{V}_A/\dot{Q}$ (poor ventilation), and intrapulmonary true shunt (no ventilation).
- Vessels in lung areas that are poorly ventilated, and therefore hypoxic, will constrict to minimize imbalances by diverting blood to better ventilated areas. This is termed *hypoxic vasoconstriction*.

DIFFUSION AND TRANSPORT OF RESPIRATORY GASES

Barriers to Diffusion

Diffusion is the passive movement of gas from a high-concentration area to a low-concentration area. Diffusion occurs because of the random, kinetic motion of molecules through the respiratory membranes and fluids. The alveolar-capillary membrane, also known as the *respiratory membrane*, through which O_2 and CO_2 must diffuse, consists of six barriers (Fig. 21.16). The membrane averages about 0.6 micrometer in thickness. For O_2 to reach the hemoglobin molecule, it must pass through surfactant, the alveolar membrane, interstitial fluid, the capillary membrane, plasma, and the red blood cell (RBC) membrane. The rate of diffusion of a gas is proportional to the tissue area and the difference

in gas partial pressure between the two sides of alveoli, and inversely proportional to the tissue thickness through which the gas must move. Oxygen diffuses into the blood from the alveoli, and CO_2 diffuses out of the blood into the alveoli. Under normal conditions, O_2 and CO_2 move across the alveolar-capillary membrane in only 0.25 second. The RBC spends about 0.75 second within the pulmonary capillary system surrounding the alveoli, thus allowing an extra 0.50 second of exchange time. Even with mild disease processes, O_2 and CO_2 have adequate time for transfer. Oxygen concentration in the alveoli, as well as its partial pressure, is controlled by (1) the rate of absorption in the blood and (2) the rate of entry of new oxygen into the lungs by the ventilation process.

Under abnormal conditions, such as thickening of the alveolar-capillary membrane (pneumonia, pulmonary edema, and interstitial lung disease) and decreased available surface area (emphysema), the diffusion capacity of the lung tissue is impaired. Diffusion capacity may be further impaired by increased physical activity, because of the decreased time spent by the RBCs in the pulmonary capillary system. Thickening of the alveolar-capillary membrane also occurs with aging. However, barriers to diffusion are rarely a primary cause of abnormal Pao_2 or $Paco_2$; abnormal diffusion of gases is usually secondary to $\dot{V}_A/\dot{Q}$ mismatch.

CO_2 is more diffusible than O_2 because of its greater solubility. Factors that determine the ability and the speed of a gas to diffuse include the available surface area of alveoli and capillaries, the integrity of the capillary and alveolar membranes, the availability of hemoglobin to transport oxygen, the solubility of the gas, the diffusion coefficient of the gas, and the differences in partial pressure of the gases on each side of the alveolar membrane. For example, because CO_2 is 24 times more soluble than O_2, it diffuses 20 times more rapidly and requires a lower partial pressure for exchange.

The decreased diffusing capacity seen in the aged person is further compromised by a decrease in the number of pulmonary capillaries and decreased lung volume and capacities. The end result is a decreased Pao_2 and increased $\dot{V}_A/\dot{Q}$ mismatch. The Pao_2 value drops about 3 to 5 mm Hg for each decade after age 30 years. Therefore an 80-year-old individual could be expected to have a Pao_2 of 75 mm Hg. Diffusion is also decreased in the newborn because of the thickness of the alveolar membrane. In a healthy adult, the Pao_2 value would be 90 to 100 mm Hg (see Table 21.3 for variations in respiratory anatomy and physiology by age grouping).

Oxygen Transport

Oxygen is transported to the tissues by two mechanisms: (1) dissolved in plasma and (2) bound to the hemoglobin molecule. Only about 0.3 mL of O_2 per 100 mL is carried dissolved in the plasma. The remaining O_2 is transported on the hemoglobin molecule. A high concentration (partial pressure) of O_2 in the pulmonary capillaries causes O_2 to bind to the hemoglobin molecule. Heme is an iron–porphyrin compound that joins with the four polypeptide chains of the protein globin. Oxygen binds to iron in each of the four heme sites to form oxyhemoglobin. At the tissue level where the partial pressure of O_2 is low, O_2 is released from the hemoglobin molecule. Depending on tissue needs, 25% of the oxygen is normally unloaded at the tissues in a resting individual, which results in venous blood being 75% saturated with oxygen.

When hemoglobin is fully bound to O_2 it is nearly 100% saturated and yields a Pao_2 of 95 to 100 mm Hg. Increasing alveolar O_2 concentration above this level will have no further effect on increasing the amount of O_2 carried on the hemoglobin molecule (Fig. 21.17). Oxygen binds when there is a high affinity of hemoglobin for oxygen (at the lungs) and releases when the affinity is decreased at the tissue level to maintain adequate metabolic processes. When Pao_2 is less than 60 mm Hg, saturation of hemoglobin with oxygen (Sao_2) falls steeply (see Fig. 21.17). The oxyhemoglobin dissociation curve diagram shows the effects of increases and decreases in O_2 affinity at any Pao_2 level. Decreased O_2 affinity, also termed a *shift to the right,* aids in the release of O_2 from the hemoglobin molecule, thus facilitating movement of O_2 from the blood to the tissues. Factors that shift the curve to the right include acidosis, hyperthermia, increased $Paco_2$ value, and increased 2,3-bisphosphoglycerate (2,3-BPG) concentration, which is an end product of RBC metabolism. The availability of O_2 is also decreased by reduced cardiac output and anemia.

Increased O_2 affinity, termed a *shift to the left,* represents a tighter binding of O_2 to the hemoglobin molecule that helps loading of oxygen in the lungs; however, it may impair delivery to the tissues. Although an increased affinity for O_2 reflects a higher percentage of saturated hemoglobin, its ineffective release in the tissues may be profound. Factors that affect hemoglobin affinity and shift the curve to the left (increased affinity) include alkalosis, hypothermia, decreased $Paco_2$ value, and decreased 2,3-BPG concentration.

Cao_2 is the sum of dissolved oxygen in the plasma plus the oxygen carried on the hemoglobin (Hb) molecule. (See Chapter 13 for a more detailed discussion of oxygen carriage and transport.) Normal arterial blood oxygen content (Cao_2) is 20 mL of O_2 per 100 mL of blood (vol%) and can be calculated by the following formula:

$$CaO_2\ (\text{vol\%}) = [\text{Hb (g/dl)} \times 1.34\ (\text{ml of } O_2/\text{g of Hb}) \times Sao_2] + (Pao_2 \times 0.003)$$

Carbon Dioxide Transport

Carbon dioxide, a by-product of cellular metabolism, is transported in the blood in three ways: dissolved in plasma (5% to 10% of the total CO_2 transport), as bicarbonate (60% to 70%), and as carbamino compounds on the hemoglobin molecule (20% to 30%). The greatest

TABLE 21.3 Normal Arterial Blood Gas Values

Parameter	Adult*	Pregnancy	Newborn	COPD (Late Findings)
Pao_2 (mm Hg)	80–100	75–100	60–70	Decreased
$Paco_2$ (mm Hg)	34–45	30–37	35–45	Increased
pH	7.35–7.45	7.35–7.45	7.30–7.40	Decreased
HCO_3^- (mEq/L)	24–30	20–26	20–26	Increased
Base excess (mEq/L)	±2	—	—	—
O_2 saturation (%)	96–100	95–100	90–100	Decreased

*For elderly patients, Pao_2 can be estimated by the following formulas: 104 – (patient's age × 0.42) for patients lying supine; and 104 – (patient's age × 0.27) for patients sitting.
COPD, Chronic obstructive pulmonary disease.

bulk of CO_2 transport is in the bicarbonate form. In the presence of the RBC enzyme carbonic anhydrase, CO_2 combines with water to form carbonic acid, which in turn almost instantaneously breaks down into bicarbonate ions and hydrogen ions. The released hydrogen ions attach to the hemoglobin molecule, and the bicarbonate ion diffuses into the plasma. Chloride ions in the surrounding plasma shift into the RBC (chloride shift). This chemical process is reversed when the venous blood reaches the lungs so that CO_2 can diffuse across the alveolar membrane to be exhaled.

KEY POINTS

- Oxygen and CO_2 diffuse quickly across alveolar-capillary membranes. Complete equilibration of gases occurs in the first third of the capillary under normal conditions. Diffusion may be incomplete when the alveolar-capillary membrane is abnormally thickened or capillary blood flow is extremely rapid.
- Carbon dioxide is more soluble and diffuses more quickly than O_2. Disorders of diffusion often affect O_2 transfer earlier and more significantly than CO_2 transfer.
- Oxygen is carried in the blood in two forms: dissolved in solution and bound to hemoglobin. Significantly more O_2 is bound than dissolved. Low hemoglobin level and low hemoglobin saturation profoundly affect the O_2 content in the blood.
- The oxyhemoglobin saturation curve describes the relationship between Pao_2 and hemoglobin saturation. At a Pao_2 of 90 to 100 mm Hg, hemoglobin is fully saturated. An increase in Pao_2 above this level does not significantly improve O_2 content.
- The affinity of hemoglobin for O_2 is affected by temperature, acid–base status, 2,3-BPG levels, and CO_2 concentration. Affinity decreases at the tissue level because of increased concentrations of H^+ ions, 2,3-BPG, and CO_2. This "shift to the right" enhances the unloading of O_2 at the tissue. A "shift to the left" occurs at the lung, where the blood is more alkalotic and CO_2 levels are lower. Increased affinity of hemoglobin in the lung enhances oxygen binding.
- Carbon dioxide is transported in the blood in three major forms: dissolved in plasma, carbaminohemoglobin, and bicarbonate ion. The most important of these is bicarbonate ion, which is formed from the combination of CO_2 and H_2O, producing carbonic acid (H_2CO_3). Carbonic acid dissociates into HCO_3^- and H^+. At the lung, the reaction proceeds in the reverse direction to form CO_2, which diffuses into the alveoli.

ALTERATIONS IN PULMONARY FUNCTION

Partial pressures of arterial O_2 in the newborn (60 to 70 mm Hg) and elderly (70 to 80 mm Hg) are less than those in the adult. The lower O_2 pressure is well tolerated in the newborn because of the presence of fetal hemoglobin, which has decreased binding of 2,3-BPG, thus facilitating oxygen transfer by shifting the oxygen dissociation curve to the left. The newborn also has a higher hemoglobin concentration (20 to 21 g/dL) for the first few weeks after birth. Therefore oxygenation is not normally a problem. The lower Pao_2 of the newborn is also associated with an increased $Paco_2$. Other blood gas values (see Table 21.3) show little difference from those of adults unless an oxygenation problem is present, such as infant respiratory distress syndrome or congenital heart disease.

Hypoventilation and Hyperventilation

Hypoventilation occurs when delivery of air to the alveoli is insufficient to meet the need to provide oxygen and remove carbon dioxide. It is influenced by decreased rate and depth of respiration. Hypoventilation results in increased $Paco_2$ (>45 mm Hg) and resultant hypoxemia due to increased alveolar carbon dioxide, which displaces oxygen. Causes may be drugs, such as morphine or barbiturates (which depress the central respiratory drive), or disorders such as obesity (Pickwickian syndrome), myasthenia gravis, obstructive sleep apnea, chest wall damage, or paralysis of respiratory muscles (especially the diaphragm). Pain related to surgery of the thorax or abdomen often results in hypoventilation secondary to decreased inspiration.

Hyperventilation is an increase in the amount of air entering the alveoli, leading to hypocapnia ($Paco_2$ <35 mm Hg). A physiologic cause of hyperventilation is hypoxic stimulation of peripheral chemoreceptors. Pain, fever, and anxiety are common causes of hyperventilation. Less common causes include obstructive and restrictive lung diseases, sepsis, and brainstem injury (central neurogenic hyperventilation). Hyperventilation is a normal physiologic response to high altitude as a compensatory mechanism to decrease $Paco_2$. Low $Paco_2$ leads to a greater ability to bind oxygen to hemoglobin (shift to the left) despite low oxygen pressure in the inspired air at high altitude.

Ineffective gas exchange from ventilatory failure occurs when an adequate volume of gas is maldistributed, minute ventilation is decreased, and/or alveolar hypoventilation occurs. Maldistribution of gas occurs in patients with emphysema, in which gas exchange occurs only in some alveolar units. In the healthy lung, some maldistribution of gas occurs because of gravitational forces on the lung, as previously discussed. In addition to the gravitational forces, airway resistance affects distribution of gases. In obstructive pulmonary diseases, increased airway resistance develops in localized regions because of (1) obstruction of airways from increased sputum production, (2) mucosal hypertrophy and edema, (3) loss of structural integrity of the airway, and (4) narrowing of the airway from bronchial smooth muscle contraction, when there is hyperactivity of the airways. During expiration, air leaves the areas of least resistance first, thus creating areas of maldistribution of gas.

Hypoxemia and Hypoxia

Two terms frequently used in discussing decreased Pao_2 are *hypoxemia* and *hypoxia*. Hypoxemia refers to deficient levels of blood oxygen as measured by low arterial O_2 concentration and low hemoglobin saturation as measured by arterial blood gases or pulse oximetry (O_2 saturation). Hypoxia refers to a decrease in tissue oxygenation. Tissue hypoxia is difficult to measure, but may be assumed when either blood flow or Pao_2 is abnormally low. A decrease in blood flow leads to a decrease in oxygen delivery.

Resultant types of hypoxia can be classified into four categories: hypoxic hypoxia, anemic hypoxia, circulatory hypoxia, and histotoxic hypoxia. Hypoxic hypoxia occurs when the Pao_2 is decreased despite normal O_2-carrying capacity. Causes include high altitude, hypoventilation, and airway obstruction. Oxygen therapy usually provides adequate treatment.

Anemic hypoxia results from a decrease in O_2-carrying capacity. Any disorder resulting in low hemoglobin concentration can cause anemic hypoxia.

Circulatory hypoxia results from a low cardiac output state in which the O_2-carrying capacity is normal but blood flow is reduced. Examples of circulatory hypoxia include shock, cardiac arrest, severe blood loss, thyrotoxicosis, and congestive heart failure.

The final classification is histotoxic hypoxia, which occurs when interference of a toxic substance leads to the inability of tissues to utilize available oxygen. Cyanide poisoning is an example of histotoxic hypoxia.

Ineffective gas exchange occurs when ventilation and perfusion are mismatched, when diffusion abnormalities exist, and when a right-to-left shunt exists. During periods of normal perfusion not all capillaries are open; however, the capillary system has the ability to recruit (open up) more capillaries and to distend (expand) capillaries already in use (see

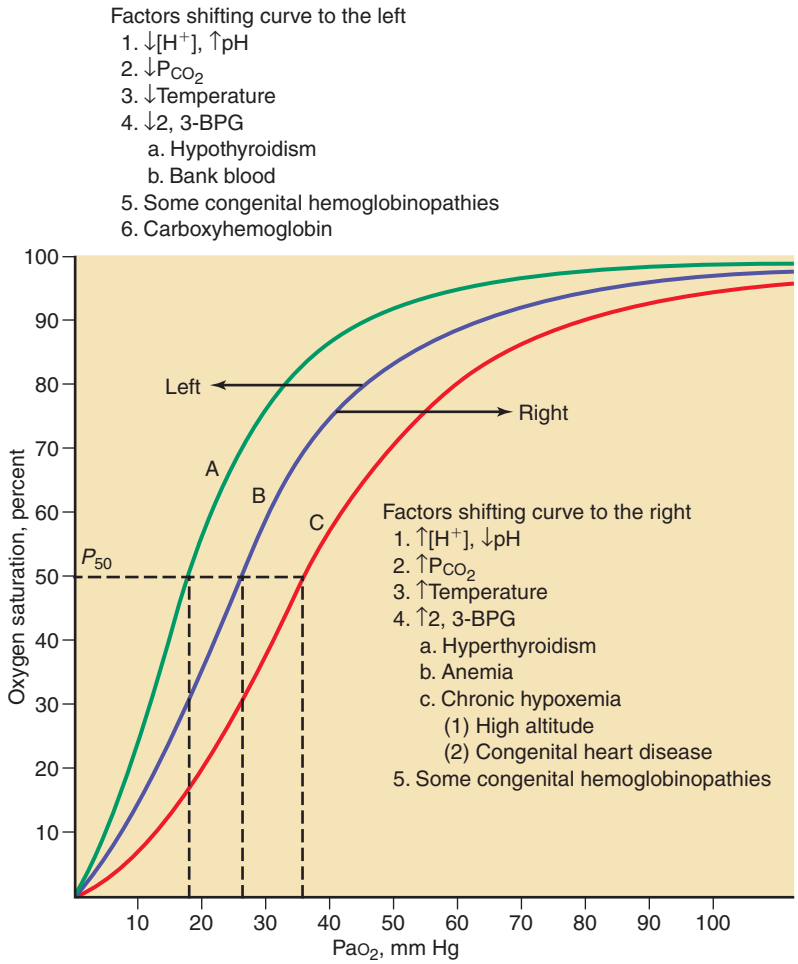


FIG 21.17 Oxyhemoglobin dissociation curve showing factors affecting hemoglobin's affinity for oxygen. Curve *B* is the standard curve under normal conditions. Curve *A* shows a shift to the left, which represents an increased affinity of hemoglobin for oxygen. Curve *C* demonstrates a shift to the right, which represents a decreased affinity. (From Gottlieb JE: Breathing and gas exchange. In Kinney MR, Packa DR, Dunbar SB, editors: *AACN's clinical reference for critical care nursing,* ed 4, New York, 1998, McGraw-Hill, p 672.)

Fig. 21.9) to increase alveolar blood flow when it is needed as a compensatory mechanism. In addition, 2% of the total blood flow in the lung is not oxygenated because the thebesian, pleural, and bronchial veins drain unoxygenated blood into the left side of the heart and into the pulmonary veins.

Areas of low ventilation-perfusion (see Box 21.1) may have normal perfusion but receive inadequate alveolar ventilation (Fig. 21.18, *A*). These areas are similar to shunting of unoxygenated pulmonary arterial blood through totally unventilated units, except that they are responsive to oxygen therapy. Areas of high ventilation-perfusion (see Fig. 21.18, *B*) may have adequate ventilation (high oxygen level in the alveoli) but have areas of decreased perfusion. This effect is similar to having increased dead space, clinically represented by areas of ventilation without blood flow. Although it is difficult clinically to differentiate diffusion defects from shunt effect, abnormalities occur in patients who have thickening of the alveolar-capillary membrane. Examples of diseases that cause thickening of the membranes include systemic lupus erythematosus, sarcoidosis, diffuse interstitial fibrosis, alveolar cell carcinoma, and Goodpasture syndrome, a rare autoimmune disease that affects the lungs and kidneys.

Ineffective gas exchange is also seen in patients with true pulmonary shunt (see Fig. 21.18, *C*). A shunt effect results from blood flowing from the right side to the left side of the heart without passing through ventilated areas of the lung. Anatomic shunts may occur in patients with ventricular septal defects, atrial septal defects, and patent ductus arteriosus. Localized pneumonia and ARDS result in intrapulmonary shunts because of $\dot{V}_A/\dot{Q}$ mismatch, in which alveoli are perfused but not ventilated.

Acute Respiratory Failure

ARF is defined as a state of disturbed gas exchange resulting in abnormal arterial blood gas values: a Pao_2 value less than 60 mm Hg (hypoxemia) and a $Paco_2$ value greater than 50 mm Hg (hypercapnia) with a pH less than 7.30 when the patient is breathing room air. Patients with respiratory failure can be divided into three categories: (1) those with failure of respiration or oxygenation leading to hypoxemia and normal or low carbon dioxide levels; (2) those with failure of ventilation leading to hypercapnia; and (3) those with a combination of respiratory and ventilatory failure.

Etiology

The precise pathophysiologic mechanism of ARF depends on the cause or causes of the disease process. A number of conditions may cause respiratory failure (Box 21.2), including disorders of the neuromuscular

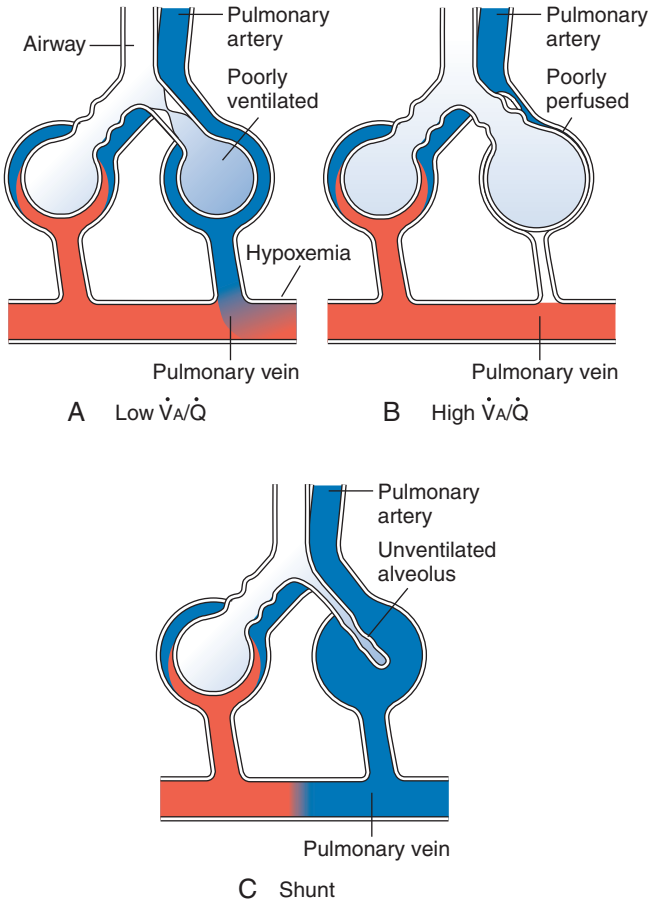


FIG 21.18 Ventilation-perfusion abnormalities. **A,** Low $\dot{V}_A/\dot{Q}$ areas that are well perfused but underventilated. **B,** High $\dot{V}_A/\dot{Q}$ areas that are well ventilated but underperfused. **C,** Shunt areas that have no ventilation but are perfused (blood flow passes unventilated alveoli).

chest apparatus (poliomyelitis, Guillain–Barré syndrome, quadriplegia, hemiplegia), disorders affecting the chest skeletal system (kyphoscoliosis), and chest trauma (rib and sternal fractures). Shock (e.g., septic, hypovolemic), pulmonary emboli (PE), and pulmonary edema may also lead to respiratory failure. Extreme obesity may lead to alveolar hypoventilation, resulting in respiratory failure. The most common lung diseases causing ARF are advanced emphysema, pneumonia, asthma, pulmonary edema, and ARDS.

In general, the development of hypoxemia is related to poorly matched ventilation and perfusion. The development of hypercapnia is related to inadequate alveolar ventilation in relation to production of carbon dioxide.

Pathogenesis. Respiratory failure can develop from any abnormality in the airways, alveoli, central nervous system, peripheral nervous system respiratory muscles, hypoperfusion, chest wall, diaphragm, and pleural space. Ventilatory demand exceeds ventilatory capacity due to one of the problems listed earlier.

Clinical Manifestations

Clinical features of ARF vary with the cause. General features of hypoxia and hypercapnia include headache, dyspnea, confusion, decreased level of consciousness, restlessness, agitation, dizziness, tremors, and initial hypertension, followed by hypotension and tachycardia. Early signs include rapid, shallow breathing with increased inspiratory muscle movement. Late findings include cyanosis, nasal flaring, and sternal and intercostal retractions. The increased work of breathing may lead to cool, clammy skin; dysrhythmias; and decreased capillary refill time.

Diagnosis

Diagnostic tests include measurement of arterial blood gases and chest radiography. A Pao_2 of less than 60 mm Hg and a $Paco_2$ of greater than 50 mm Hg on room air are common findings. Chest radiographic findings depend on the disease process. Other supporting tests include an electrolyte panel with evidence of electrolyte imbalance such as low potassium and low sodium concentrations, and a complete blood cell count with evidence of increased numbers of white blood cells associated with infection or decreased levels of red blood cells and hemoglobin attributable to anemia.

Treatment

Maintaining ventilatory support by maintaining airway patency and ensuring adequate alveolar ventilation is the primary goal of therapy. Mechanical ventilation may be the initial treatment, followed by management of the underlying cause. If a neuromuscular problem or skeletal weakness is present, assisted ventilation with a positive-pressure volume ventilator is indicated to maintain airway patency and ensure adequate alveolar ventilation.

The primary goal of therapy is to provide adequate oxygenation at the cellular level by maintaining a Pao_2 greater than 60 mm Hg (oxygen saturation, 90%). Specific interventions depend on the cause. If acute respiratory failure is caused by chronic obstructive pulmonary disease (COPD), then vigorous management of bronchospasm and possible infection is required using a combination of methylxanthines, β_2 agonists, corticosteroids (controversial), and antibiotics. Heart failure or hypotension may require drug therapy. Diuretics may be given for volume reduction depending on the fluid volume status of the patient. Hypotension should be managed promptly with volume replacement and/or vasopressors. The use of corticosteroids in high doses for the first 24 to 48 hours of the disease process is controversial because no conclusive evidence of efficacy is available.

General supportive care consists of providing adequate nutrition to maintain fluid and electrolyte balance, offering pain management and emotional support, and preventing complications of gastrointestinal stress and bed rest. Developing a method of communication with ventilated patients is also very important. High-calorie, high-protein, low-carbohydrate nutritional support is recommended. A diet high in carbohydrates should be avoided because of its tendency to increase carbon dioxide production. (See Chapter 23 for specific treatments for ARDS and infant respiratory distress syndrome.)

KEY POINTS

- Ventilatory failure occurs when alveolar ventilation is insufficient to accomplish adequate gas exchange. Ventilatory failure may result from decreased respiratory rate, decreased tidal volume, or increased dead space. Arterial blood gas analyses demonstrate hypercarbia and hypoxemia.
- A general deficiency of O_2 in the blood (hypoxemia) results from poor diffusion at the alveoli (hypoxic hypoxia) or anemia (anemic hypoxia). Tissue hypoxia may be due to general hypoxemia or poor perfusion (circulatory hypoxia) or result from poor uptake of O_2 by the tissue (histotoxic hypoxia).
- Oxygenation failure occurs when diffusion of gases across the alveolar-capillary interface is impaired. Oxygenation failure may be due to mismatching, right-to-left shunt, or excessive barriers to diffusion. Arterial blood gas values demonstrate hypoxemia but not necessarily hypercarbia.

BOX 21.2 Causes of Acute Respiratory Failure

Central Nervous System
Drug overdose (sedative, hypnotic, opioid, anesthetic)
Cerebral vascular accident (stroke)
Hypothyroidism
Central nervous system infections
Brain trauma
Brain tumor

Neuromuscular Diseases and Related Disorders
Guillain–Barré syndrome
Myasthenia gravis
Multiple sclerosis
Muscular dystrophy
Myxedema
Poliomyelitis
Polymyositis
Drug or toxin induced
- Botulism
- Aminoglycosides
- Organophosphates
- Neuromuscular blocking agents

Tetanus
Amyotrophic lateral sclerosis
Quadriplegia
Hemiplegia

Chest Wall and Diaphragm
Trauma (thoracic/abdominal)
Kyphoscoliosis
Upper abdominal or thoracic surgery
Pleural effusion
Hemothorax/pneumothorax/chylothorax
Massive ascites

Airways
Laryngospasm
Foreign body aspiration
Asthma
Acute exacerbation of chronic bronchitis or emphysema

Pulmonary Parenchymal Diseases
Lung contusion
Aspiration
Pneumonia
Interstitial lung diseases
Emphysema
Pulmonary fibrosis
Acute respiratory distress syndrome
Infant respiratory distress syndrome
Pulmonary emboli (blood, fat, air, amniotic fluid)
Cardiac and noncardiac pulmonary edema
Shock
Increased CO_2 production
- Fever
- Infection
- Hyperthyroidism
- Drugs

- Acute respiratory failure (ARF) is generally diagnosed from arterial blood gas disturbances. The usual defining values are a Pa_{O_2} less than 60 mm Hg and a Pa_{CO_2} greater than 50 mm Hg when the subject is breathing room air.
- Conditions that predispose an individual to hypoventilation, ventilation–perfusion mismatch, or right-to-left shunt may lead to respiratory failure (e.g., drugs, neuromuscular weakness, chest wall deformities or trauma, and parenchymal lung diseases).
- Manifestations of respiratory failure are due to tissue hypoxia and compensatory responses and include confusion, tremors, hypotension, depressed consciousness, tachypnea, and tachycardia.
- The goal of therapy is to reduce tissue hypoxia by maintaining Pa_{O_2} above 60 mm Hg. Depending on the underlying disease process, this may require mechanical ventilation, supplemental oxygen administration, nutritional supplementation, and utilization of bronchodilators and antibiotics.

DIAGNOSTIC TESTS

Pulmonary Function Testing

The primary criterion in diagnosing obstructive disease is the demonstration of obstruction to airflow in the lungs. Table 21.2 lists common ventilatory parameters referred to in spirometry.

Spirometry is performed by asking the patient to inhale deeply and then to exhale as quickly as possible until maximal air is exhaled. The total volume of air exhaled is known as the *forced vital capacity (FVC)*. To determine flow, the time required for exhaling the air is also measured. The volume exhaled in the first second is a reliable and reproducible index of obstructive airway disease. This value is the *forced expiratory volume in 1 second* (FEV_1). Fig. 21.19 presents spirogram examples of normal, restrictive, and obstructive graphs for FEV_1 and FVC. For all spirometric studies, normal values are based on large population studies of healthy volunteers and are adjusted for height, weight, age, and gender. Results are compared with predicted values and reported as percent of predicted.

A simple formula has been developed to define and quantify airflow obstruction. If the FEV_1/FVC ratio is 75% or greater, no significant obstruction of airflow is present. If the value obtained is between 60% and 70%, then mild obstruction of airflow is present. Moderate obstruction is defined as a value of 50% to 60%, and severe obstruction is present when the FEV_1/FVC ratio is less than 50%. Therefore using a spirometer and measuring both volume and time, the diagnosis of COPD can be made and the severity quantified.

From the spirometric ventilatory measurements (see Table 21.2), other determinations of airflow can be made from the middle to later parts of an FVC maneuver. These measures are helpful in determining the presence of small airway disease. Some investigators believe that small airway disease may be a precursor to the development of chronic bronchitis and emphysema.

Frequently, an inhaled bronchodilator, such as albuterol or metaproterenol, may be given, with testing repeated in 15 to 20 minutes. If the FEV_1 improves by 15% or more, the patient is considered to have a positive bronchodilator response, indicative of partially reversible bronchospasm of the smooth muscles of the airways. This is most often the case with asthma or asthmatic bronchitis.

A second pulmonary function test known as the *diffusion capacity* measures the ability of the alveolar gases to diffuse into the capillary

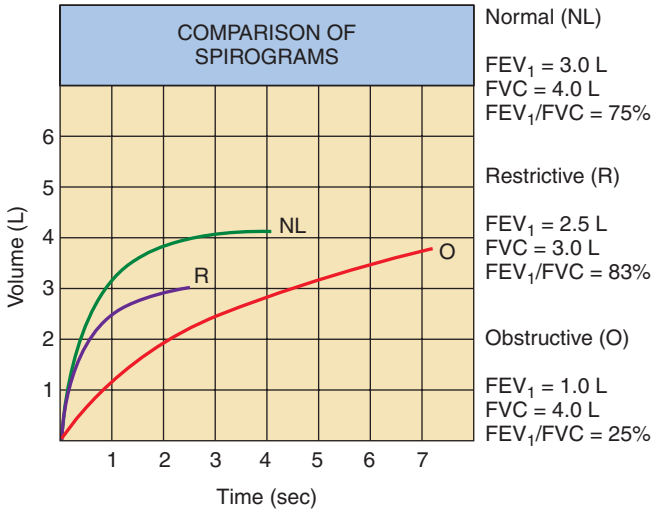


FIG 21.19 Spirometry examples for normal, obstructive, and restrictive patterns.

blood. The technical details of this test are beyond the scope of this book, but it is a valuable test for determining either thickening (fibrosis) of the alveolocapillary membrane or destruction (emphysema) of the membrane.

By breathing mixtures of an inert gas, such as helium, the total lung capacity (TLC) can be determined. This volume is composed of the FVC and the RV. The RV is the volume of air that remains in the lung after a person has forcefully exhaled all of the air from the lungs (see Fig. 21.10). RV/TLC is normally 30% to 35%. In some patients with airflow obstruction, air tends to get trapped in the lungs, thereby increasing the RV and resulting in overinflation of lung tissue.

Arterial blood gases are also useful as a pulmonary function measurement. Using these values, a careful assessment of both the oxygenation and the acid–base status can be determined. The normal pH is 7.40, the normal $Paco_2$ is 40 mm Hg, and the normal Pao_2 at sea level is 80 to 100 mm Hg. In COPD, especially in the severe stage, Pao_2 falls and $Paco_2$ rises. Table 21.3 lists normal arterial blood gas values for various groups. A thorough discussion of arterial blood gas analysis can be found in Chapter 25.

Bronchial Provocation Tests

The controlled induction of bronchospasm by inhalation of various agents is occasionally used to identify patients with hyperreactive airways and to prove whether certain inhaled substances can produce bronchospasm. Usually a series of inhalations is administered, followed by a series of ventilation measurements. Generally the test is stopped when the FEV_1 falls at least 20% more than the control measurement. This should only be done where emergency support services are available. Bronchoprovocation is contraindicated if the patient is already exhibiting symptoms or requires continual asthma medication. Allergens can be administered as solutions, dusts, or fumes. The amount administered should be no more than the patient would normally encounter in the environment. If symptoms occur, they can be readily reversed by two to four inhalations of albuterol or metaproterenol.

General hyperreactivity of the bronchi can be detected by having the patient inhale histamine phosphate solutions or methacholine (related to acetylcholine) or nebulized distilled water. A decline of more than 20% in the FEV_1 is indicative of hyperreactivity.

KEY POINTS

Obstructive disorders are associated with characteristic abnormalities on pulmonary function testing. These include the following:

- Decreased FEV_1
- Low FEV_1/FVC ratio (<70%)
- Improvement in FEV_1 after use of a bronchodilator (asthma)
- Increased residual volume
- Increased functional residual capacity

ALTERATIONS IN PULMONARY VASCULATURE

Pulmonary Hypertension

Etiology

Normally the pulmonary circulation is a high-flow, low-pressure system. Pulmonary hypertension is defined as a sustained increase in pulmonary artery pressure above 25 mm Hg systolic resting and above 30 mm Hg systolic with exercise. In some cases of pulmonary hypertension, systolic pressures may be as high as 60 to 110 mm Hg. Two broad types of pulmonary hypertension exist: primary (idiopathic) and secondary. Primary pulmonary hypertension (PPH) is relatively rare (1300 per 1 million people), rapidly progressive, and more common in women than in men (1.7:1 ratio). PPH usually presents in the third to fourth decade of life. The cause is unknown, but can be associated with portal hypertension of cirrhosis, use of appetite-suppressant drugs, and human immunodeficiency virus infection. Most cases are sporadic and may be familial (6% to 12% of cases); however, mutations in the genes that code for members of the tumor growth factor-β family of receptors on chromosome 2q33 have been found in some cases. The long-term prognosis is poor, and medical management is usually ineffective.

Secondary pulmonary hypertension results from a known disease process or pathophysiologic condition. Increased pulmonary blood flow, increased resistance to blood flow, and increased left atrial pressure are the three major mechanisms resulting in pulmonary hypertension. Of these, increased resistance to blood flow is the most common cause and is attributable to hypoxic vasoconstriction, as seen in chronic bronchitis and advanced emphysema. Box 21.3 lists the major causes of secondary pulmonary hypertension.

Pathogenesis

Chronic exposure to the mechanisms listed in Box 21.3 (except PPH) results in morphologic changes within the arterial lumen. Initially, the walls of the small pulmonary vessels thicken because of an increase in the muscle. This initial response is thought to occur as a result of local tissue hypoxia, acidosis, or both.

As the underlying pathologic process intensifies, the internal layer of the pulmonary artery wall becomes fibrotic, with further muscle thickening. In addition, muscle development occurs in vessels that are normally nonmuscular. Pulmonary atherosclerosis is present in major pulmonary vessels as well.

Sustained pulmonary hypertension (mean pulmonary arterial pressure of 27 to 60 mm Hg) results in the formation of plexiform (network of blood vessels) lesions. These nodular lesions are composed of irregular, interconnecting blood channels that further impede an already compromised pulmonary vasculature. Tissue necrosis and hemorrhage often result.

Clinical Manifestations

The clinical manifestations of pulmonary hypertension vary according to the severity and duration of the underlying pathologic process. Because of the normal distensibility of pulmonary capillaries and the ability of

BOX 21.3 Mechanisms of Secondary Pulmonary Hypertension

Increased Pulmonary Vascular Resistance

Vasoconstrictive
- Alveolar hypoxia attributable to bronchitis or emphysema
- Acidosis
- High altitude
- Thromboembolic causes from obstruction or release of histamine, serotonin, or catecholamines
- Hypoxia attributable to neuromuscular disease, obesity, obstructive sleep apnea, or kyphoscoliosis

Obstructive
- Embolism (blood clots, fat emboli, amniotic emboli, tumor cells, or foreign body)

Obliterative (loss of capillary bed)
- Emphysema
- Lung resection
- Pulmonary fibrosis
- Collagen vascular disease
- Vasculitis

Increased Left Atrial Pressure

Mitral stenosis, mitral regurgitation
Left ventricular failure
Constrictive pericarditis

Increased Pulmonary Blood Flow/Viscosity

Atrial septal defects
Ventricular septal defects
Polycythemia
Sickle cell disease
Patent ductus arteriosus
Congenital heart disease

Other

Portal hypertension/liver cirrhosis
Appetite-suppressant drugs
HIV
Schistosomiasis
Sarcoidosis

the lung to recruit additional reserve capillary beds with increased pressure or flow, the condition often remains asymptomatic until significant damage to pulmonary vasculature has occurred. Exercise intolerance (because of progressive loss of pulmonary capillary distention and recruitment capabilities) is often one of the earliest clinical symptoms. Patients may also experience syncope, increasing dyspnea, chest pain on exertion, fatigue, hemoptysis, and pulmonary edema. Eventually, cor pulmonale (right-sided heart enlargement secondary to primary lung disease) and right ventricular failure will develop if persistent, severe pulmonary hypertension continues, because of persistent backpressure to the right-sided heart chambers. (See the discussion of cor pulmonale and heart failure in Chapter 19.) Common signs and symptoms of pulmonary hypertension are dyspnea (60% of cases), syncope, chest pain, jugular venous distention, a systolic ejection click, narrowing or a splitting of S_2 and S_4, and accentuation of the pulmonary component of the second heart sound (P_2). In advanced cases, tricuspid and pulmonary valve insufficiency are present. Development of a hoarse voice attributable to compression of the recurrent laryngeal nerve by an engorged pulmonary artery (Ortner syndrome) may occur.

Diagnosis

Pulmonary artery catheters are used to obtain accurate pulmonary arterial pressure measurements in patients at rest and during exercise. Unfortunately, even if mild pulmonary hypertension is present, pulmonary arterial pressure values are usually normal at rest. Pulmonary arterial pressures measured in the exercising subject would be the optimal diagnostic tool. However, the feasibility of exercising a patient with invasive central line monitoring is problematic. The pulmonary artery catheter could become wedged and necrosis could occur, or the catheter could slip back into the ventricle and irritate the myocardium, causing ventricular dysrhythmias. Stress testing without exercise may be done in the cardiac catheterization laboratory by using medications to increase cardiac output.

A chest radiograph, although usually normal in cases of mild pulmonary hypertension, is one of the earliest diagnostic tools to suggest the presence of moderate-to-severe hypertension. Enlargement of the pulmonary arteries and right ventricle, as well as abnormal vessel contours, is indicative of hypertensive disease. The 12-lead electrocardiogram (ECG) shows evidence of right ventricular hypertrophy. The two-dimensional echocardiogram (a noninvasive technique) can also provide evidence of pulmonary hypertension. Echocardiography may reveal mitral stenosis, left atrial myxoma, and right heart enlargement. Further diagnostic tests should be done to exclude causes of secondary pulmonary hypertension.

Treatment

The major treatment for pulmonary hypertension is early identification and control of the underlying disease process. In the case of left-to-right shunts, surgical closure of an atrial septal defect or patent ductus arteriosus may be indicated. Because the most common cause of pulmonary hypertension is related to increased pulmonary vascular resistance, treatment is often directed at reversing vasoconstriction by administering supplemental oxygen and avoiding vigorous exercise and pregnancy. Depending on the stage of hypertension, vasodilators and diuretics are commonly used in an attempt to control the symptoms. These medical regimens have produced inconsistent results.

Advanced stages of PPH are irreversible. The only feasible intervention is lung or heart–lung transplantation. Continuous long-term infusion of prostacyclin (epoprostenol, treprostinil) is being used for its potent vasodilatory effect on the pulmonary vessels. It improves exercise capacity, quality of life, and long-term survival.

Pulmonary Venous Thromboembolism

Etiology

A PE is undissolved detached material that occludes blood vessels of the pulmonary vasculature. As a result, circulation distal to the obstructed area is impaired. Approximately 650,000 patients are affected annually, with an estimated mortality of 50,000 annually. Of those who experience fatal PE, 8% to 10% die within 1 hour of the onset of initial symptoms.

More than 90% of pulmonary thromboemboli originate in the deep veins of the lower extremities. Other sources of PE include fat, air, and amniotic fluid. The types of emboli and their causes are summarized in Table 21.4.

Virchow, a pathologist of the 1800s, discovered three physiologic factors that predispose patients to thrombus formation, increasing the risk of PE. The three factors, commonly referred to as *Virchow's triad,* are venous stasis (sluggish blood flow), hypercoagulability, and damage to the venous wall (intimal injury). Several predisposing factors enhance the probability of thrombus development and the subsequent risk for PE. Using Virchow's triad, predisposing factors have been categorized under each of the three components (Box 21.4). The most common

TABLE 21.4 Embolism Types and Causes

Embolism Type	Cause
Thrombotic	Blood clots develop in venous system, predominantly in thighs and legs
Fat	Globules of fat secondary to fractures of pelvis or long bones
Amniotic fluid	Collections of fluid, hair, or other debris related to complicated labor, especially in older, multiparous women
Air	Venous access through IV catheters
Tumor	Fragments from malignant tissue
Foreign material	Foreign bodies (bullets, sutures, catheter tips, orally prepared medications injected IV)
Septic	Infected tissue or related substances (fungal/bacterial)
Parasitic	Parasites present in lung vasculature

BOX 21.4 Factors Predisposing to Pulmonary Embolism of Virchow's Triad

Venous Stasis

Extended bed rest (delayed venous removal of activated clotting factors)
Postoperative state
Immobility (activated clotting factors)
Vascular disorders (thrombophlebitis of lower extremities and pelvic area)
Congestive heart failure (venous backflow/stasis)
Cardiac dysrhythmias (atrial fibrillation)
Dehydration
Prolonged air travel
Obesity

Hypercoagulability

Oral contraceptives (estrogen therapy), hormone replacement therapy
Pregnancy, early puerperium
Polycythemia (chronic high altitude; chronic pulmonary disease with decreased Pa_{O_2} and increased Pa_{CO_2})
Malignant pathologic processes, visceral cancer
Cigarette smoking
Inherited resistance to activated protein C
Deficiency of protein S
Deficiency of antithrombin III
Prothrombin gene mutation
Presence of antiphospholipid antibodies (lupus), anticoagulant and anticardiolipin antibodies

Damage to Vessel Wall (Intimal Injury)

Blunt trauma
Penetrating wounds
Bone fractures with soft tissue injury
Surgical procedures (hip, pelvic, abdominal, cardiovascular)
Obstetric manipulations during labor and delivery
Burns
Central venous catheter

risk factors for venous thromboembolus formation are immobility, trauma, pregnancy, cancer, heart failure, and estrogen use.

Pathogenesis

Thrombi are dislodged from their point of origin by multiple mechanisms, including direct trauma, exercise and muscle action, and changes in blood flow. Regardless of whether the emboli are blood clots or an alternative type of material (see Table 21.4), once they are released into the venous system, the undissolved material travels to the pulmonary vasculature. The lower lobes are frequently involved because of high blood flow. The impact of pulmonary emboli on the cardiopulmonary circulation depends on the size and cross-sectional area of circulatory impairment. If the embolus occludes less than 25% of the pulmonary vessels in a healthy individual, no physiologic changes may be seen. When the occlusive area approaches 25% to 30%, pulmonary arterial pressures may begin to rise, with potential right-sided heart failure. In the patient without any underlying pulmonary pathology, 50% of the cross-sectional pulmonary circulation must be impaired before dangerously high pulmonary arterial pressures are generated. Because of the large pulmonary capillary reserve, significant damage is necessary before pulmonary decompensation occurs.

Pulmonary arterial pressures increase because of vasoconstriction from actual mechanical obstruction of blood vessels and the release of serotonin and neural sympathetic stimulation in a combined neurohormonal response. Right-sided heart failure occurs because of the high resistance generated by the pulmonary vasculature. Eventually, hypotension occurs as a result of diminished cardiac output.

Actual pulmonary infarction (death of lung parenchyma) occurs only in about 10% to 15% of cases of PE. Pulmonary necrosis is rare because three sources are available for oxygen supply: the pulmonary arterial circulation, the bronchial arterial circulation, and the airways. Significant underlying pulmonary or cardiac impairment (COPD, mitral stenosis) increases the risk for occurrence of pulmonary infarctions.

Clinical Manifestations

Presenting symptoms depend on the size of the embolus, as well as on any underlying cardiopulmonary pathologic conditions. Initial symptoms may include restlessness, apprehension, and anxiety. The most common symptom is dyspnea (75% to 85% of patients). In addition, tachycardia (23%) and tachypnea (30%) are often present. Sudden dyspnea and severe chest pain are usually associated with medium-sized to massive pulmonary emboli. Chest pain may be nonpleuritic or pleuritic (infarction). Pain on inspiration is seen in 65% to 75% of patients. Hemoptysis may or may not occur. As the clinical picture worsens, patients experience heart failure, shock, and respiratory arrest.

Diagnosis

Because PE is often misdiagnosed, the clinician should have a high index of suspicion when clinical manifestations, combined with factors predisposing to PE (see Box 21.4), are noted. Although no simple noninvasive test has both high specificity and high sensitivity, the $\dot{V}_A/\dot{Q}$ lung scan is one of the tests performed to determine the presence of a PE. The scan determines whether a mismatch exists between ventilation and perfusion. Adequate ventilation with impaired perfusion (blood flow) to the pulmonary vasculature (mismatch) is indicative of PE if the scan is performed within 8 hours of symptom onset. Helical angiography is replacing lung scans as the initial diagnostic test for pulmonary thromboembolism. This noninvasive test requires administration of radiocontrast dye.

Other screening tools, such as arterial blood gas analyses, ECGs, chest radiographs, and cardiac enzyme determinations, are valuable for ruling out related pathologic processes. Arterial blood gases generally reveal decreased Pa_{O_2} and Pa_{CO_2} and increased pH. A chest radiograph may be normal or show an elevated diaphragm, pleural effusion, infiltrates, or atelectasis. An ECG is abnormal in 70% to 85% of cases with acute PE. Common electrocardiographic findings are sinus tachycardia, nonspecific T-wave and ST-segment changes, and T-wave

BOX 21.5 Wells Rules

1. Clinical signs/symptoms of deep vein thrombosis (DVT) (score = 3.0)
2. No alternative diagnoses likely or more likely than pulmonary embolism (PE) (score = 3.0)
3. Heart rate >100 beats/min (score = 1.5)
4. Immobilization or surgery in last 4 weeks (score = 1.5)
5. Previous history of DVT or PE (score = 1.5)
6. Hemoptysis (score = 1)
7. Cancer actively treated within last 6 months (score = 1.0)
8. The probability of a PE is high if total score is >6, moderate if 2 to 6, and lower if <2.

inversion. The probability of a PE can be estimated by the use of Wells rules (Box 21.5).

The conclusive diagnostic test for PE is pulmonary arteriography, but is rarely done. This invasive procedure involves the injection of radiopaque material into the pulmonary artery. If an intraluminal filling deficit can be identified, the test is considered diagnostic for PE. A duplex ultrasonography of the lower extremities to determine the site of deep vein thrombosis should also be performed.

Treatment

The primary intervention for PE is prevention. Patients who are at risk for developing one of the factors of Virchow's triad must be treated prophylactically. In the case of prolonged bed rest, active range-of-motion exercises as well as prophylactic low-dose subcutaneous sodium heparin or low-molecular-weight heparins may be used. Intraoperative and postoperative graded compression stockings and intermittent pneumatic leg compression are beneficial in reducing the risk of PE.

Patients with suspected or confirmed PE are given supplemental oxygen or ventilator support with immediate activity limitations to decrease oxygen demand. A continuous heparin IV drip is used as a mainstay of therapy. Although heparin does not dissolve the clot, formation of new clots is prevented. Heparin may also stimulate the intrinsic fibrinolytic system, enhancing the degradation of the PE. Thrombolytic therapy may be used to dissolve the emboli. However, thrombolytics create an increased risk of bleeding, and unless the patient has a massive embolism and is hemodynamically unstable, they are not used.

If patients are thought to be releasing multiple emboli despite adequate heparin therapy, an umbrella filter (Mobin-Uddin) or a "bird's nest" filter may be placed in the inferior vena cava to trap emboli as they migrate toward the pulmonary vasculature. Inferior vena cava filters reliably prevent recurrent PE with a recurrence rate of about 1% at 12 days. An embolectomy may be performed on an emergency basis if the hemodynamic consequences of the emboli are life threatening and the patient has refractory hypotension.

KEY POINTS

- Pulmonary hypertension usually results from conditions that increase the resistance of the pulmonary vasculature. Disorders that reduce the total cross-sectional area of the lung increase resistance and promote pulmonary hypertension. Destruction of capillaries (emphysema), blockage of vessels (emboli), and vasoconstriction (hypoxemia) are common examples.
- Pulmonary hypertension may occur when left atrial pressure is elevated. Pulmonary arterial pressure must increase to maintain the driving pressure necessary to propel blood through the pulmonary circulation. The excessive pulmonary blood flow that accompanies left-to-right shunting of blood through heart defects may also lead to pulmonary hypertension.
- Cor pulmonale (right ventricular hypertrophy) and right-sided heart failure may develop with sustained high pulmonary vascular resistance. Few symptoms of pulmonary hypertension are manifested until the right side of the heart is affected.
- Treatment centers on efforts to ameliorate the underlying cause if possible (e.g., closure of heart defects, administration of oxygen to reduce hypoxic vasoconstriction). Vasodilators and diuretics may be used to reduce pulmonary arterial pressure and decrease strain on the right side of the heart.
- Pulmonary emboli result in obstruction of blood flow through part of the pulmonary system. When emboli are large or multiple, a significant increase in pulmonary pressure may result, causing right ventricular failure.
- Emboli may be composed of fat, air, amniotic fluid, or thrombi (blood clots). Thrombi are the most common cause. Thrombi generally form in the leg under conditions of venous stasis, enhanced coagulation, or vascular trauma (Virchow's triad).
- Pulmonary embolism (PE) is suspected with sudden dyspnea and chest pain. Symptoms of right-sided heart failure may be present when emboli are large. A ventilation–perfusion scan or other imaging may be done to confirm the diagnosis.
- Prophylactic anticoagulation in persons at risk for thrombus formation is important to prevent PE. Postoperative pneumatic leg compression also reduces the risk of thrombus formation. Bed rest, oxygen administration, and thrombolytic and anticoagulation therapy are the mainstays of therapy for acute PE. Ventilator support and measures to improve the functioning of the right side of the heart may be necessary in severe cases.

PULMONARY MALIGNANCIES

Etiology

The incidence of lung cancer in the United States has been increasing in recent years, with more than 220,000 new cases per year. The four major types of lung cancer are large cell carcinoma, small cell carcinoma, squamous cell carcinoma, and adenocarcinoma. Another type of lung cancer is bronchoalveolar, which comprises 5% of lung cancers. Lung cancer is responsible for more than 32% of cancer deaths in men and more than 25% of cancer deaths in women. Among women there has been a 600% increase in the incidence of lung cancer over the last 80 years. Tobacco smoking is the major cause (85%) of lung cancer, with approximately 155,000 deaths reported per year. About 3800 deaths per year are reported in nonsmokers who receive secondary smoke from the environment. The remaining percentage of lung cancer is not attributable to smoking. Individuals at highest risk for developing lung cancer are those who started smoking before the age of 25 years or longer, have smoked one or more packs of cigarettes a day for 20 years, work under conditions of asbestos exposure, and are older than 50 years. (See Chapter 7 for further discussion of cancer biology.)

Pathogenesis

Squamous cell epidermoid carcinoma (20% to 30% of cases) usually originates (in two thirds of cases) in the central bronchi near the hilus as an intraluminal growth. Cytologic examination of sputum reveals the squamous cell carcinoma, leading to earlier detection of this cancer than with other lung neoplasms. The tumor normally doubles its volume in 100 days and, as it advances, metastasizes to regional lymph nodes in the area.

Adenocarcinomas (35% to 40% of cases) usually appear in the periphery of the lung and are not as amenable to early detection as squamous cell carcinoma. Adenocarcinomas are characterized by acinar bronchoalveolar and papillary tumors. Doubling time is about 180 days, with metastasis occurring to distant organs, which may be due to aerosol transmission in the case of bronchoalveolar (2% of cases) carcinoma.

Large cell carcinomas (10% to 15% of cases) develop in the lung periphery and are similar to adenocarcinoma. The tumor cells are large and are arranged in nests or clusters. The tumor doubles in size about every 100 days and metastasizes to distant organs.

Small cell (oat cell) carcinoma (15% to 20% of cases) tends to originate extrinsically in the central bronchus region, thus compressing and narrowing the bronchi. The narrowing may lead to signs and symptoms of obstruction of a central airway, leading to wheezing. This type of tumor is associated with a lesion on chromosome 3 and grows rapidly, doubling in about 33 days. Widespread metastasis is common with small cell carcinomas and they are the most resistant to therapy.

Bronchoalveolar (5%) carcinoma tends to originate in the periphery and metastasize through the lymphatics. There is no correlation with cigarette smoking.

Clinical Manifestations

Clinical features vary according to the type and the location of the tumor and whether it has metastasized. Approximately 10% to 25% of cases are asymptomatic. Signs and symptoms can be classified as intrathoracic or extrathoracic.

Extrathoracic manifestations are weight loss, fatigue, anorexia, anemia, and clubbing. Facial and upper extremity edema is noted in cases of tumor compression of the superior vena cava. Superior vena cava syndrome is most commonly caused by bronchogenic carcinoma.

Intrathoracic manifestations include dyspnea, cough, chest pain, hemoptysis, and increased sputum production (with bronchoalveolar carcinoma). Hoarseness may be evident and is caused by pressure of the tumor on the recurrent laryngeal nerve. Phrenic nerve involvement (1% of lung cancer patients) leads to paralysis of the hemidiaphragm on the affected side and the potential for development of atelectasis and pneumonia. Clinical findings of endobronchial obstruction include atelectasis, postobstructive pneumonia, pleural effusion (12% to 33%), and Horner syndrome (miosis, ipsilateral ptosis, and dyshidrosis). Abdominal breathing measures are taught to the patient who does not recover diaphragmatic function from damage to the phrenic nerve. Extension of the cancer cells to the pleural cavity may cause pleural effusion.

Diagnosis

Pulmonary function tests may show increased volumes in moderately advanced cases of bronchial carcinoma. Because the tumor blocks the airway, an obstructive pattern of pulmonary disease may lead to increased or decreased functional residual capacity attributable to the effect of the mass lesion.

The definitive diagnosis of cancer requires positive cytologic or histologic findings. Bronchoscopy washings are a common method of diagnosing lung cancer in patients whose lesions are centrally located. Pleural fluid samples show positive findings in 50% to 65% of patients with malignant pleural effusion. Histologic examination of tissue after biopsy of the pleura, lung tissue, or mediastinal lymph nodes may also be helpful in diagnosing lung cancer.

Chest radiographs show abnormal findings in nearly all patients with lung cancer. Common findings are hilar (squamous cell) and/or peripheral (adenocarcinoma) masses, atelectasis, mediastinal widening, infiltrates, pleural effusions, and cavitation (squamous cell epidermoid carcinoma). Chest studies are helpful in evaluating tumor size and nodal involvement. Computed tomography (CT) scans of the chest are used for staging and for follow-up study after treatment. Positron emission tomography scanning with F-labeled fluorodeoxyglucose, a metabolic marker of malignant tissue, is superior to a CT scan in detecting mediastinal and distant metastases in non–small cell lung cancer.

Treatment

Primary treatment options for pulmonary neoplasms are surgery, chemotherapy, radiation therapy, laser therapy, and tissue-specific therapies for airway lesions. Patients are also strongly encouraged to stop smoking.

The treatment of choice for non–small cell carcinoma is surgery. For nonoperable neoplasms, radiation therapy is the secondary choice. Radiotherapy improves survival in patients with nonresectable non–small cell carcinoma. Combination chemotherapy is the therapy of choice in patients with extensive small cell carcinoma. Combined chemotherapy and chest radiation therapy has proved effective as a cure for patients in whom disease was detected early. Radiation therapy is also used for palliation of symptoms as with other cancers, immunotherapies may be effective adjuncts. (See Chapter 7.)

KEY POINTS

- Cigarette smoking is the major cause of lung cancer. Lung cancer is usually disseminated at the time of diagnosis and is associated with a high mortality.
- Lung cancers can develop in the bronchial tree (small cell, squamous cell) or in the parenchyma (large cell, adenocarcinoma).
- Lung cancer may be advanced before symptoms become troublesome. Manifestations include cough, hemoptysis, hoarseness, chest pain, and pleural effusion. The diagnosis is based on examination of cells from bronchial secretions or tissue biopsy. Pulmonary masses may be detected by plain radiography or computed tomography (CT) of the chest.
- As with other cancers, treatment may include surgical removal of resectable tumors followed by radiation therapy, chemotherapy, and immunotherapy.

SUMMARY

The primary function of the respiratory system is oxygenation of the tissues. This function is accomplished by the movement of O_2 from the atmosphere through the airways to alveolar sacs. The inhaled air is warmed, humidified, and filtered in the upper airway on its way to the alveoli. Once in the alveoli, diffusion of O_2 and CO_2, a gaseous waste product, occurs and O_2 is transported by means of hemoglobin molecules to the tissues. The respiratory system has numerous control mechanisms that influence its function. For example, the respiratory control centers may be inhibited when a person has a brain injury, thus producing inadequate respiration. Failure of oxygenation is seen in acute respiratory failure. Acute respiratory failure, defined as a Pao_2 less than 60 mm Hg and a $Paco_2$ greater than 50 mm Hg at room air, occurs in individuals who have developed ventilation–perfusion mismatching, right-to-left shunt, or hypoventilation. Diseases affecting the pulmonary vasculature include pulmonary hypertension and PE. Pulmonary hypertension is associated with disease processes (e.g., emphysema, PE, hypoxemia) that increase pulmonary vascular resistance. PEs result in obstruction to blood flow in the pulmonary vasculature. The health or disease of a patient's cardiovascular, renal, and hematologic systems also affects the functioning of the respiratory system. Health care professionals have a key role in the prevention and management of respiratory disease and in patient and family education.

RESOURCES

Pulmonary System Development and Physiology

Allen SC: The respiratory system. In Fillit HM, Rockwood K, Woodhouse K, editors: *Brocklehurst's textbook of geriatric medicine and gerontology*, ed 7, New York, 2010, Churchill Livingstone.

Barrett KE, Barman SM, Boltan SS, et al: *Ganong's review of medical physiology*, ed 25, New York, 2015, McGraw Hill.

Clouter MM, Throll RS: The respiratory system. In Koeppen BM, Stanton BA, editors: *Berne & Levy physiology*, ed 6, updated edition. Philadelphia, 2010, Mosby-Elsevier.

Gadara H, Hirbe A, Nassif M, et al: *The Washington manual of medical therapeutics*, ed 34, Philadelphia, 2013, Lippincott Williams & Wilkins.

Hall JE: *Guyton and Hall textbook of medical physiology*, ed 13, Philadelphia, 2015, Elsevier.

Kliegman RM, Stanton BF, St Geme JW: *Nelson textbook of pediatrics*, ed 20, Philadelphia, 2015, Elsevier.

Kumar V, Abbas AK, Aster J: *Robbins & Cotran pathologic basis of disease*, ed 9, Philadelphia, 2015, Elsevier.

Mackenzie CF: Compromises in the choice of orotracheal or nasotracheal intubation and tracheostomy. *Heart Lung* 12:485–492, 1983.

Moore KL, Persaud TVN, Torchia M: *The developing human: clinically oriented embryology*, ed 10, Philadelphia, 2015, Elsevier.

Rossi A, et al: Aging and the respiratory system. *Aging* 8(3):143–161, 1996.

West JB: *Respiratory physiology: the essentials*, ed 9, Philadelphia, 2012, Lippincott Williams & Wilkins.

Pulmonary Disorders

Arshad R: Pulmonary arterial hypertension. In Ferri FF, editor: *Ferri's clinical advisor:5 books in 1*, Philadelphia, 2016, Elsevier, pp 1044–1046.

Chestnutt MS, Prendergast TJ: Pulmonary disorders. In McPhee SJ, Papadakis MA, Rabow MW, editors: *Current medical diagnosis and treatment*, ed 55, New York, 2016, Lange/McGraw-Hill, pp 242–321.

Fine B: Pulmonary embolism. In Ferri FF, editor: *Ferri's clinical advisor: 5 books in 1*, Philadelphia, 2016, Elsevier, pp 1049–1052.

Goroll AJ, Mulley AG: Management of chronic obstructive pulmonary disease. In Goroll AH, Mulley AG, editors: *Primary care medicine: office evaluation and management of the adult patient*, ed 7, Philadelphia, 2014, Lippincott Williams & Wilkins, pp 382–398.

Lapner ST, Kearon C: Venous thromboembolism. In Bope ET, Kellerman RD, editors: *Conn's current therapy 2016*, Philadelphia, 2016, Elsevier, pp 423–431.

Murphy DP, Boller WA: Chest pain (non-cardiac). In Buttaro TM, Trybulski J, Polgar-Bailey P, et al, editors: *Primary care: a collaborative practice*, ed 4, Philadelphia, 2012, Elsevier, pp 433–439.

West JB: *Pulmonary physiology and pathophysiology: an integrated case-based approach*, ed 2, Philadelphia, 2007, Lippincott Williams & Wilkins.

West JB: *Pulmonary pathophysiology: the essentials*, ed 8, Philadelphia, 2013, Lippincott Williams & Wilkins.

Whelan CA: Plogar-Bailey P: Pulmonary hypertension. In Buttaro TM, Trybulski J, Polgar-Bailey P, et al, editors: *Primary care: a collaborative practice*, ed 5, Philadelphia, 2016, Elsevier, pp 114–115.

Pulmonary Malignancies

Diatrick KB, Long J, Chang AC: Thoracic wall, pleura, mediastinum, and diaphragm. In Doherty GM, editor: *Current surgical diagnosis and treatment*, ed 14, New York, 2015, Lange/McGraw-Hill, pp 331–389.

Ferri FF: Lung neoplasm, primary. In Ferri FF, editor: *Ferri's clinical advisor:5 books in 1*, Philadelphia, 2016, Elsevier, pp 751–754.

Kratzke RA, Patel MR: Primary lung cancer. In Bope ET, Kellerman RD, editors: *Conn's current therapy 2016*, Philadelphia, 2016, Saunders, pp 409–414.

Winland-Brown JE, Porter BO, Thomas DJ: Respiratory problems. In Dunphy LM, Winland-Brown JE, Porter BO, et al, editors: *Primary care: the art and science of advanced practice nursing*, ed 4, Philadelphia, 2015, FA Davis.

22

Obstructive Pulmonary Disorders

Benjamin J. Miller and Lorna L. Schumann

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- What are the clinical manifestations and common causes of acute airway obstruction?
- What are the pathogenesis and clinical manifestations for conditions that cause obstruction of the airway lumen?
- What is the role of inflammation in the development of asthma?
- How does the underlying genetic defect in cystic fibrosis lead to pulmonary and exocrine gland dysfunction?
- How does smoking cause both the alveolar destruction of emphysema and the bronchial damage of chronic bronchitis?
- What is the rationale for using drugs such as β_2 agonists, acetylcholine antagonists, leukotriene inhibitors, corticosteroids, and mast cell stabilizers to manage obstructive pulmonary disorders?

CHAPTER OUTLINE

Obstruction From Conditions in the Wall of the Lumen, 479
Asthma, 479
Etiology, 479
Pathogenesis, 480
Clinical Manifestations, 481
Diagnosis, 481
Treatment, 481
Acute Bronchitis, 483
Etiology, 483
Pathogenesis, 485
Clinical Manifestations, 485
Diagnosis, 485
Treatment, 485
Chronic Bronchitis, 485
Etiology, 485
Pathogenesis, 486
Clinical Manifestations, 486
Diagnosis, 487
Treatment, 487
Obstruction Related to Loss of Lung Parenchyma, 488
Emphysema, 488
Etiology, 488
Pathogenesis, 489
Clinical Manifestations, 490
Diagnosis, 490
Treatment, 491
Obstruction of the Airway Lumen, 491
Bronchiectasis, 491
Etiology, 491
Pathogenesis, 491
Clinical Manifestations, 492
Diagnosis, 492
Treatment, 492
Bronchiolitis, 493
Etiology, 493
Pathogenesis, 493
Clinical Manifestations, 493
Diagnosis, 493
Treatment, 493
Cystic Fibrosis, 494
Etiology, 494
Pathogenesis, 494
Clinical Manifestations, 495
Diagnosis, 495
Treatment, 495
Acute Tracheobronchial Obstruction, 495
Etiology, 495
Pathogenesis, 495
Clinical Manifestations, 495
Diagnosis, 495
Treatment, 495
Epiglottitis, 496
Etiology, 496
Pathogenesis, 496
Clinical Manifestations, 496
Diagnosis, 496
Treatment, 496
Croup Syndrome, 496
Etiology, 496
Pathogenesis, 496
Clinical Manifestations, 496
Diagnosis, 496
Treatment, 496

Obstructive lung diseases are manifested by increased resistance to airflow. Obstructive diseases of the lung can be classified into those involving (1) obstruction from conditions in the wall of the lumen (e.g., asthma, bronchitis), (2) obstruction resulting from increasing pressure around the outside of the airway lumen (e.g., emphysema secondary to loss of lung tissue and elasticity, enlarged lymph node, or tumor), and (3) obstruction of the airway lumen (e.g., presence of a foreign body, excessive secretions, aspiration of fluids).

These classifications are mainly terms of convenience because many respiratory disease processes involve several areas of the pulmonary system. Involvement of the airways produces narrowing of the passages so that airflow obstruction occurs. The major obstructive airway diseases are asthma, bronchitis, and emphysema.

OBSTRUCTION FROM CONDITIONS IN THE WALL OF THE LUMEN

Asthma

Etiology

Asthma is a lung disease characterized by (a) airway obstruction that is reversible (but not completely in some patients), (b) airway inflammation, and (c) increased airway reactivity to a variety of stimuli. In terms of symptoms, asthma is defined by paroxysms of diffuse wheezing, dyspnea, and cough resulting from spasmodic contractions of the bronchi. Airway inflammation leads to epithelial denudation, collagen deposition beneath the basement membrane, mast cell activation, mucosal edema, increased viscid secretions, and smooth muscle contraction. With proper treatment, most patients with asthma can control the disease and prevent development of emphysema or bronchitis. Asthma occurs in about 7% to 14% of the U.S. population and is common among children and adults, with an annual mortality rate of 10.7 per 1 million people, or approximately 3400 people. The annual direct and indirect cost of asthma is estimated at more than $56 billion. Asthma is the most common chronic disease of childhood with a disproportionally high prevalence rate in African Americans, inner-city dwellers, and premature or low-birth-weight children. The pathophysiology of both intrinsic (nonallergic, sometimes referred to as *adult onset*) and extrinsic (allergic, sometimes referred to as *pediatric onset*) asthma is thought to involve inflammation of the airways. Most cases of asthma can be triggered both by allergens and by stimuli, such as exercise and exposure to cold air. The terms *intrinsic* and *extrinsic* are still used, but many prefer the terms *nonallergic* and *allergic.* The clinical features of all forms are similar.

Asthma is associated with the release of inflammatory chemicals from mast cells in the airways. The mechanisms stimulating mast cell release are immunoglobulin E (IgE)–mediated triggers for extrinsic/allergic asthma (Fig. 22.1). Intrinsic/nonallergic asthma occurs in patients who have no history of allergy. Allergic asthma (extrinsic) comprises approximately one third to one half of all cases. Asthma is often associated with a history of hay fever or eczema (atopy), a positive family history of the disease, and positive skin test reactions to allergens (dust mites, cat/dog dander, industrial chemicals). Pharmacologic therapy, allergen-specific immunotherapy, and environmental control are usually beneficial. Refer to Chapter 10 for details about IgE-mediated mechanisms and hyposensitization methods.

Intrinsic/nonallergic asthma frequently develops in middle age and has a less favorable prognosis. Respiratory tract infections or psychological factors appear to be contributory, whereas antigen–antibody reactions appear to have less of a role in the disease process, although IgE levels may be elevated. Attacks are often severe, and patients have a variable response to medical therapy. Allergen-specific immunotherapy and environmental control measures are *not* usually helpful. Airways are hyperreactive, and patients may present with extreme dyspnea, orthopnea, and agitation.

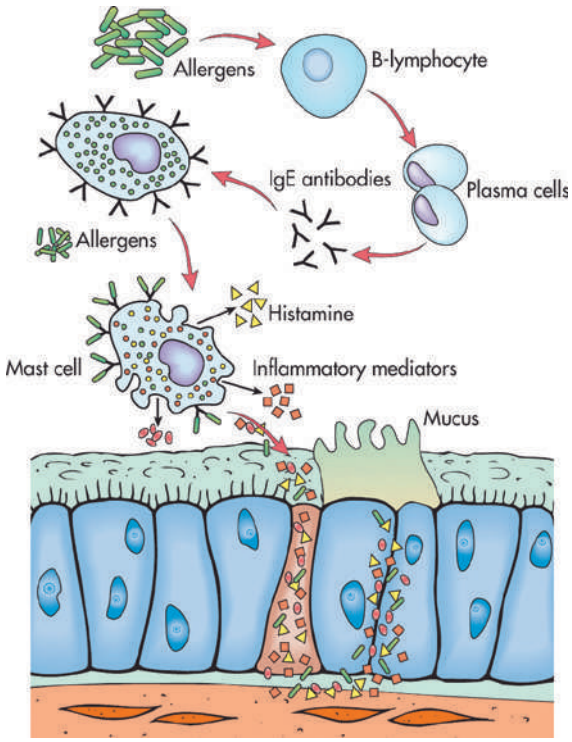


FIG 22.1 Allergic asthma is triggered when an allergen cross-links IgE receptors on mast cells, which are then activated to release histamine and other inflammatory mediators (early-phase response). A late-phase response may occur due to further inflammation. (From Lewis SL et al: *Medical-surgical nursing: assessment and management of clinical problems,* ed 9, St Louis, 2014, Elsevier.)

Exercise-induced asthma is common, especially in children and adolescents. Bronchospasm often occurs within 3 minutes after the end of exercise and usually resolves in 60 minutes. Heat loss, water loss, and increased osmolarity of the lower respiratory mucosa are believed to stimulate mediator release from basophils and tissue mast cells. This mediator release produces airway smooth muscle contraction. Running, jogging, and playing tennis are the most common instigators of exercise-induced asthma. Bicycling and swimming are much less likely to induce symptoms.

Occupational asthma may be accompanied by positive skin test reactions to protein allergens in the work environment. Occupational exposures to allergens, such as fumes from plastic, formaldehyde, isocyanates, some metals, textiles, engine exhaust, sulfur dioxide, fluoride, and western red cedar dust, *do not* provoke skin reactions. To prove hypersensitivity, it may be necessary to conduct challenge tests in the patient by inhalation of the suspected dust or fumes in a controlled environment. The individual affected by occupational asthma tends to have progressively more severe attacks with subsequent exposures. Symptoms may clear over a weekend or vacation and recur when the individual returns to the work environment. This repeated history often is sufficient to establish the diagnosis. Hyposensitization in most cases of occupational asthma is ineffective because of lack of an IgE antibody reaction and because the chemicals that cause symptoms usually are toxic when injected.

Drug-induced asthma can produce symptoms ranging from mild rhinorrhea to respiratory arrest requiring mechanical ventilation. In

patients with nasal polyps, sinusitis, and asthma, ingestion of aspirin may induce severe or occasionally fatal asthmatic attacks. Sometimes anaphylactoid reactions cause a decrease in blood pressure, itching (pruritus), rhinorrhea, or a rash after aspirin ingestion. Aspirin intolerance with asthma usually occurs in adults. Attacks may occur within minutes of ingestion or may be delayed up to 12 hours. Nonsteroidal antiinflammatory drugs such as indomethacin (Indocin), ibuprofen (Motrin, Advil), and related drugs may also induce asthma in the aspirin-intolerant patient. Aspirin reactions are not immunologically mediated. Therefore skin testing is not useful for diagnosing aspirin intolerance. Because aspirin and nonsteroidal antiinflammatory drugs inhibit the conversion of arachidonic acid to prostaglandins, it is possible that aspirin shunts arachidonic acid breakdown products to the leukotriene system. Leukotrienes, released from mast cells, are slow-reacting substances of anaphylaxis with powerful bronchoconstriction activity (see Fig. 22.1). Avoidance is the most practical approach to this problem because testing can be dangerous.

Asthma can occur from ingestion of food additives. Tartrazine (yellow dye no. 5), which is used to color pharmaceuticals, hair products, and food products, may also produce severe asthma in susceptible persons. A complete list of drugs containing tartrazine can be obtained from the Food and Drug Administration.

Monosodium glutamate, used as a flavor enhancer in foods, can produce faintness, nausea, sweating, a fall in blood pressure, and, occasionally, asthma. Sodium or potassium metabisulfite, used to preserve fruits, vegetables, and meats, can cause anaphylactoid reactions. A challenge with the chemical may be necessary to establish a diagnosis, as metabisulfites are widespread in our society.

Hops in beer have also been implicated in causing severe bronchospasm. Skin reactivity does not occur, and the mechanism of the problem is not IgE mediated. The diagnosis involves a history of exposure followed by symptoms.

Gastroesophageal reflux disease can trigger an asthma attack. It is assumed that the gastric acid reflux in the esophagus is aspirated into the lungs, resulting in bronchoconstriction.

Pathogenesis

The immunohistopathologic features of asthma include denudation of airway epithelium, collagen deposition beneath the basement membrane, edema, mast cell activation, and inflammatory cell infiltration by neutrophils, eosinophils, and lymphocytes. Inflammation of the airway contributes to acute bronchospasm (bronchoconstriction), mucosal edema, mucous plug formation, and airway wall remodeling (Fig. 22.2). Genetic predisposition (chromosomes 5, 11, 14) for atopy and structural predisposition (smaller airways) are the strongest predisposing factors for developing asthma. There is a strong association of the *ADAM33* gene with asthmatic bronchial hyperresponsiveness.

With allergic asthma, an IgE-mediated response is common and is manifested by elevated IgE levels, allergic rhinitis; eczema; a positive family history of allergy; and attacks associated with seasonal, environmental, or occupational exposure. The mechanism of action is initiated by exposure to a specific antigen that has previously sensitized mast cells in airway mucosa. When the antigen reacts with the antibody on the surface of the mast cell, packets of chemical mediator substances stored in the cell are released. The chemical mediators that are released include histamine, slow-reacting substances of anaphylaxis (leukotrienes), prostaglandins, bradykinins, eosinophilic chemotactic factor, serotonin, and others. Fig. 22.1 depicts common chemicals that are released by the mast cell and the physiologic effect of these chemicals. Cytokines are probably the most important inflammatory mediators, particularly those associated with T_H2 helper T-cell activation (granulocyte-macrophage colony–stimulating factor and interleukins 3, 4, 5, and 13). These cytokines may be responsible for modulating inflammatory and immune cell function. Other inflammatory mediators are arachidonic acid metabolites such as leukotrienes and prostaglandins, platelet-activating factor, neuropeptides, reactive oxygen species, histamine, and adenosine. With the release of chemical mediators, the normal respiratory epithelium is denuded and replaced by goblet cells, resulting in mucosal edema, production of inflammatory exudates, and hyperresponsiveness of the airway (bronchoconstriction and leakage). Alterations in epithelial integrity lead to increased microvascular permeability. A secondary mediator response occurs 6 to 12 hours after the primary asthma attack and is more refractory to treatment. Neutrophil chemotactic factor may be the cause of this secondary response.

Histologic changes in the epithelial basement membrane occur over time. The basement membrane is a complex structure that separates endothelial cells from underlying stroma. The membrane provides tensile strength and physical support to surrounding structures. It also functions as a filter and as a site for cell attachment. In a classic study by Hogg in 1982, the width of the basement membrane was shown to thicken in

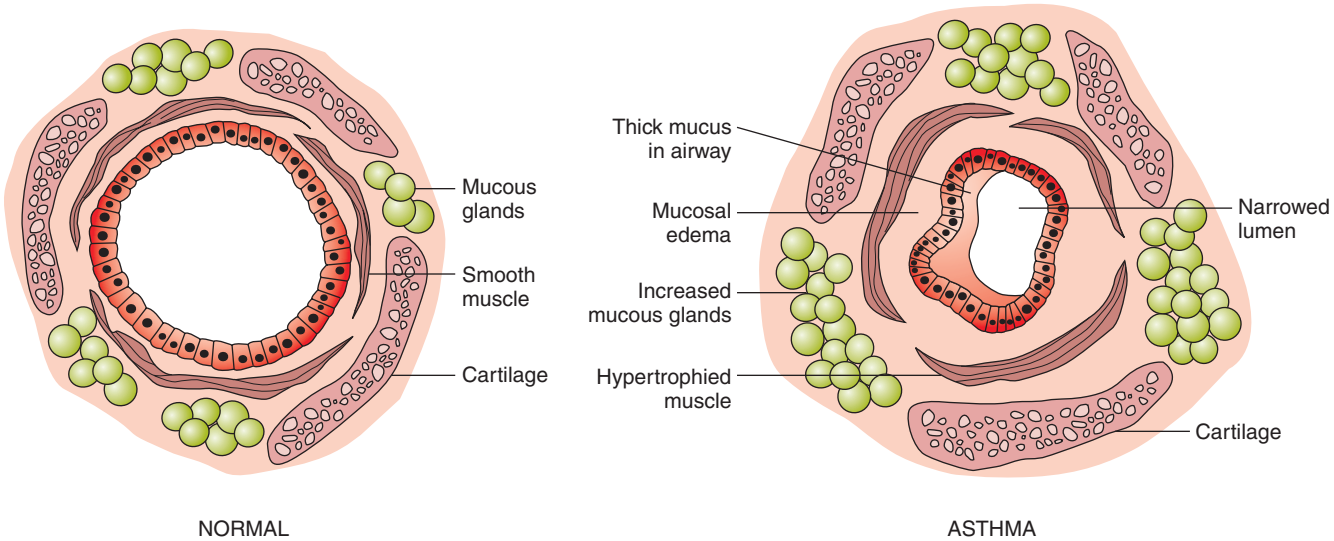


FIG 22.2 Common bronchial wall remodeling changes in asthma are hypertrophied smooth muscle, edema, mucous gland hyperplasia, and mucus in the lumen.

asthmatic patients over time. The width seen in asthmatic patients is 17.5 µm, whereas that seen in healthy subjects is 7 µm. Airway remodeling has been detected pathologically. Declines in pulmonary function over time can progress to chronic obstructive pulmonary disease (COPD). Fig. 22.1 depicts the pathogenesis of asthma in relation to mast cell release and parasympathetic stimulation by way of the vagus nerve. Vagal stimulation leads to edema, mucus hypersecretion, and bronchoconstriction. The nerve endings of asthmatic patients have been found to be devoid of the bronchodilator neuropeptide vasoactive intestinal peptide.

Clinical Manifestations

Common symptoms are wheezing, feelings of tightness of the chest, dyspnea, cough, and increased sputum production. Some patients have only a chronic dry cough, and others have a productive cough. Especially in children, cough is often the earliest sign of exacerbation of asthma. Wheezing is caused by vibration in narrowed airways, which act like the vibrating reed of a wind instrument, yielding a musical sound. Because airways naturally widen with inspiration, inspiratory wheezes reflect increased constriction. Sputum is often thick, tenacious, scant, and viscid (sticky). Physical findings vary with the severity of the attack. A mild attack may be associated with a random monophonic expiratory wheezing associated with airway narrowing, tachycardia, and tachypnea. Random monophonic wheezes are located throughout the chest and are intermittent on examination. The area in which the wheezes are heard best is indicative of the area of obstruction (e.g., if they are heard best at the mouth, this is indicative of large airway obstruction). Tachycardia is an early sign of hypoxemia. A more severe attack requiring medical assistance may be accompanied by the use of accessory muscles of respiration, intercostal retractions, distant breath sounds with inspiratory wheezing, orthopnea, agitation, tachypnea, and tachycardia. In the severe state, the patient may appear cyanotic, agitated, restless, and confused. The intensity of wheezing is *not* a reliable indicator of blockage of airflow. The measurement of peak expiratory flow rate (PEFR) is the best indicator of reduction in airflow (see discussion under "Diagnosis"). PEFRs are affected by weight, height, age, gender, ethnicity, posture, effort, smoking, and circadian rhythm. A PEFR of less than 80 L/min indicates severe obstruction. When obstruction is the tightest, the patient cannot move enough air with enough velocity to make wheezing sounds. Isolated inspiratory wheezing may be an indicator of large airway obstruction caused by mucus or laryngeal obstruction. A patient with severe respiratory distress, prolonged expiration (indicating that the person is having difficulty moving air out of the lungs), neck and intercostal retractions, and minimal air sounds is critically ill and requires emergency intervention.

Status asthmaticus is a severe form of asthma that fails to respond to the typical use of inhaled bronchodilators. This is often life threatening and requires immediate medical attention.

Diagnosis

The diagnosis of asthma is based on history, physical findings, sputum examination, pulmonary function tests, blood gas analysis, and chest radiography. Radiographic findings may be normal or may show evidence of hyperinflation with flattening of the diaphragm in progressive disease. Abnormal physical findings include cough, wheezing, a hyperinflated chest, and decreased breath sounds. Asthmatic sputum samples may reveal Charcot–Leyden crystals (formed from crystallized enzymes from eosinophilic membranes), eosinophils, and Curschmann spirals (mucous casts of bronchioles).

Forced expiratory volumes decrease during asthma attacks. PEFR is measured to determine the index of airway function. The PEFR is the maximal flow of expired air attained during a forced vital capacity (FVC) procedure. The evaluation of asthma should include the measurement of forced expiratory volume over 1 second (FEV_1), FVC, and the FEV_1/FVC ratio before and after administration of a short-acting bronchodilator. Airflow obstruction is indicated by an FEV_1/FVC ratio of less than 75%. Classification of asthma severity and control (Figs. 22.3 and 22.4) is based on presenting symptoms, frequency of nighttime symptoms, and lung function. Fig. 22.5 shows stepped therapy for asthma based on classification.

Arterial blood gas values may be normal during a mild attack, but as the bronchospasm increases in intensity, respiratory alkalosis and hypoxemia become prominent findings. Elevation of arterial partial pressure of carbon dioxide ($Paco_2$) is a poor prognostic sign, indicating that the patient's ability to continue breathing at a rapid rate has diminished and that exhaustion is imminent.

Respiratory failure may be manifested by severe respiratory distress in a patient who shows no radiographic evidence of pneumothorax. As the patient improves, the wheezing becomes louder. When wheezing is no longer heard after an asthma attack, pulmonary function tests may continue to show obstructive changes for several weeks. Some patients have a slight monophonic wheeze continuously between asthma bouts and still are comfortable and functional.

Determination of allergens is done by skin testing or inhalation of suspected allergens. Skin testing is usually more helpful in young patients who have extrinsic asthma. Bronchial provocation testing with histamine or methacholine may be useful in confirming the diagnosis of asthma in certain cases (see the Diagnostic Tests section in Chapter 21).

A complete blood cell count can show an elevated number of white blood cells (WBCs) with an increased number of eosinophils. Eosinophils are prominent in the cellular infiltrate of the bronchioles, the sputum, and the peripheral blood. A decline in the total eosinophil count is a valuable measure of effectiveness of corticosteroid treatment. With effective treatment, the total eosinophil count is depressed below 10/µL.

Treatment

Patients should be advised to avoid the objects in the environment that trigger asthma attacks. Environmental control includes control of dust; removal of allergens such as feathers, molds, and animal dander; and, in some cases, removal of rugs and carpets. Other environmental control factors that help some patients include the use of air purifiers and air conditioners. The patient should also be taught preventive therapy in regard to smoking cessation and avoidance of passive smoke, aerosols, and odors. Patients should seek early treatment for respiratory tract infections.

Pharmacologic therapy for all three major obstructive disorders is similar and focuses on decreasing inflammation and bronchoconstriction, including β_2 agonists, corticosteroids, leukotriene modifiers, and mast cell inhibitors (Fig. 22.6). Other therapies used in patients with more severe asthma include home oxygen therapy and home administration of small-volume nebulizer treatments via intermittent positive-pressure ventilation. At home, peak flow monitoring is helpful to parents or patients in determining a treatment plan and when to seek medical assistance. Peak flow meters are also helpful in monitoring progress of the patient with around-the-clock therapy.

Allergen-specific immunotherapy (hyposensitization) may be used as an adjunct to other therapies. The allergen is first identified by testing with purified allergens using the scratch, prick, or intradermal method. Desensitization therapy has been shown in controlled studies to reduce the frequency and severity of asthmatic episodes when a single offending allergen can be identified.

Status asthmaticus (severe attack unresponsive to routine therapy) requires more rapid and intense therapy, which may include epinephrine, subcutaneous terbutaline, and/or aminophylline. Once airflow has improved, aerosol bronchodilating inhalers may be used. Intravenous

Components of Severity		Classification of Asthma Severity ≥12 years of age			
			Persistent		
		Intermittent	Mild	Moderate	Severe
Impairment **Normal FEV_1/FVC:** **8–19 yr 85%** **20–39 yr 80%** **40–59 yr 75%** **60–80 yr 70%**	Symptoms	≤2 days/week	>2 days/week but not daily	Daily	Throughout the day
	Nighttime awakenings	≤2x/month	3–4x/month	>1x/week but not nightly	Often 7x/week
	Short-acting $beta_2$-agonist use for symptoms control (not prevention of EIB)	≤2 days/week	>2 days/week but not daily, and not more than 1x on any day	Daily	Several times per day
	Interference with normal activity	None	Minor limitation	Some limitation	Extremely limited
	Lung function	• Normal FEV_1 between exacerbations • FEV_1 >80% predicted • FEV_1/FVC normal	• FEV_1 >80% predicted • FEV_1/FVC normal	• FEV_1 >60% but <80% predicted • FEV_1/FVC reduced 5%	• FEV_1 <60% predicted • FEV_1/FVC reduced >5%
Risk	Exacerbations requiring oral systemic corticosteroids	0–1 year (see note)	≥2/year (see note) →		
		← Consider severity and interval since last exacerbation. → Frequency and severity may fluctuate over time for patients in any severity category. Relative annual risk of exacerbations may be related to FEV_1.			
Recommended Step for Initiating Treatment **(See Figure 22-5 for treatment steps.)**		**Step 1**	**Step 2**	**Step 3** and consider short course of oral systemic corticosteroids	**Step 4 or 5** and consider short course of oral systemic corticosteroids
		In 2–6 weeks, evaluate level of asthma control that is achieved and adjust therapy accordingly.			

Key: *FEV_1*, forced expiratory volume in 1 second; *FVC*, forced vital capacity; *ICU*, intensive care unit

Notes:

- The stepwise approach is meant to assist, not replace, the clinical decision making required to meet individual patient needs.
- Level of severity is determined by assessment of both impairment and risk. Assess impairment domain by patient's/caregiver's recall of previous 2-4 weeks and spirometry. Assign severity to the most severe category in which any feature occurs.
- At present, there are inadequate data to correspond frequencies of exacerbations with different levels of asthma severity. In general, more frequent and intense exacerbations (e.g., requiring urgent, unscheduled care, hospitalization, or ICU admission) indicate greater underlying disease severity. For treatment purposes, patients who had ≥2 exacerbations requiring oral systematic corticosteroids in the past year may be considered the same as patients who have persistent asthma.

FIG 22.3 Classifying asthma severity and initiating treatment in youths ≥12 years of age and adults who are not currently taking long-term control medicines. (From National Institute of Health's National Heart, Lung, and Blood Institute: 2007 National Asthma Education and Prevention Program: *Expert Panel Report 3: guidelines for the diagnosis and management of asthma,* p 344, Fig. 4.6.)

corticosteroids are the mainstay of abortive therapy. Oxygen therapy, with or without mechanical ventilation, may be necessary in severe cases.

The more patients understand about their asthma, the better they are at self-managing their symptoms. Educational materials are available from the American Lung Association, the Asthma and Allergy Foundation of America, and the National Institute of Allergy and Infectious Diseases.

KEY POINTS

- An asthma episode may range in severity from mild to life threatening, depending on the degree of airway obstruction. With intense narrowing of the bronchi, severe hypoxemia may result.
- Several types of asthma have been identified. Nonallergic (intrinsic) asthma is precipitated by exercise, stress, and exposure to pulmonary irritants, but no specific allergen can be identified. Drugs such as aspirin and exposure to occupational allergens have also been identified as etiologic agents.
- Allergic (extrinsic) asthma is mediated by IgE, which is produced in response to specific antigens. The IgE binds to mast cells and causes them to release inflammatory chemicals in response to antigen. Skin testing may be helpful in identifying suspected allergens.
- Prevention of asthma attacks is an important part of therapy. Avoidance of precipitating factors and use of prophylactic drug therapy are recommended. Bronchodilators, corticosteroids, and oxygen therapy are mainstays of treatment for an acute attack.

Components of Control		Classification of Asthma Control (≥12 years of age)		
		Well Controlled	Not Well Controlled	Very Poorly Controlled
Impairment	Symptoms	≤2 days/week	>2 days/week	Throughout the day
	Nighttime awakenings	≤2x/month	1–3x/week	≥4x/week
	Interference with normal activity	None	Some limitation	Extremely limited
	Short-acting beta$_2$-agonist use for symptom control (not prevention of EIB)	≤2 days/week	>2 days/week	Several times per day
	FEV_1 or peak flow	>80% predicted/ personal best	60–80% predicted/ personal best	<60% predicted/ personal best
	Validated questionnaires ATAQ ACQ ACT	 0 ≤0.75* ≥20	 1–2 ≥1.5 16–19	 3–4 N/A ≤15
Risk	Exacerbations requiring oral systemic corticosteroids	0–1 year	≥2/year (see note)	
		Consider severity and interval since last exacerbation		
	Progressive loss of lung function	Evaluation requires long-term follow-up care		
	Treatment-related adverse effects	Medication side effects can vary in intensity from none to very troublesome and worrisome. The level of intensity does not correlate to specific levels of control but should be considered in the overall assessment of risk.		
Recommended Action for Treatment (See Figure 22-5 for treatment steps.)		• Maintain current step. • Regular follow-ups every 1–6 months to maintain control. • Consider step down if well controlled for at least 3 months	• Step up 1 step and reevaluate in 2–6 weeks. • For side effects, consider alternative treatment options.	• Consider short course of oral systemic corticosteroids, • Step up 1–2 steps, and reevaluate in 2 weeks. • For side effects, consider alternative treatment options.

*ACQ values of 0.76–1.4 are indeterminate regarding well-controlled asthma.
Key: *EIB*, exercise-induced bronchospasm; *ICU*, intensive care unit

Notes:

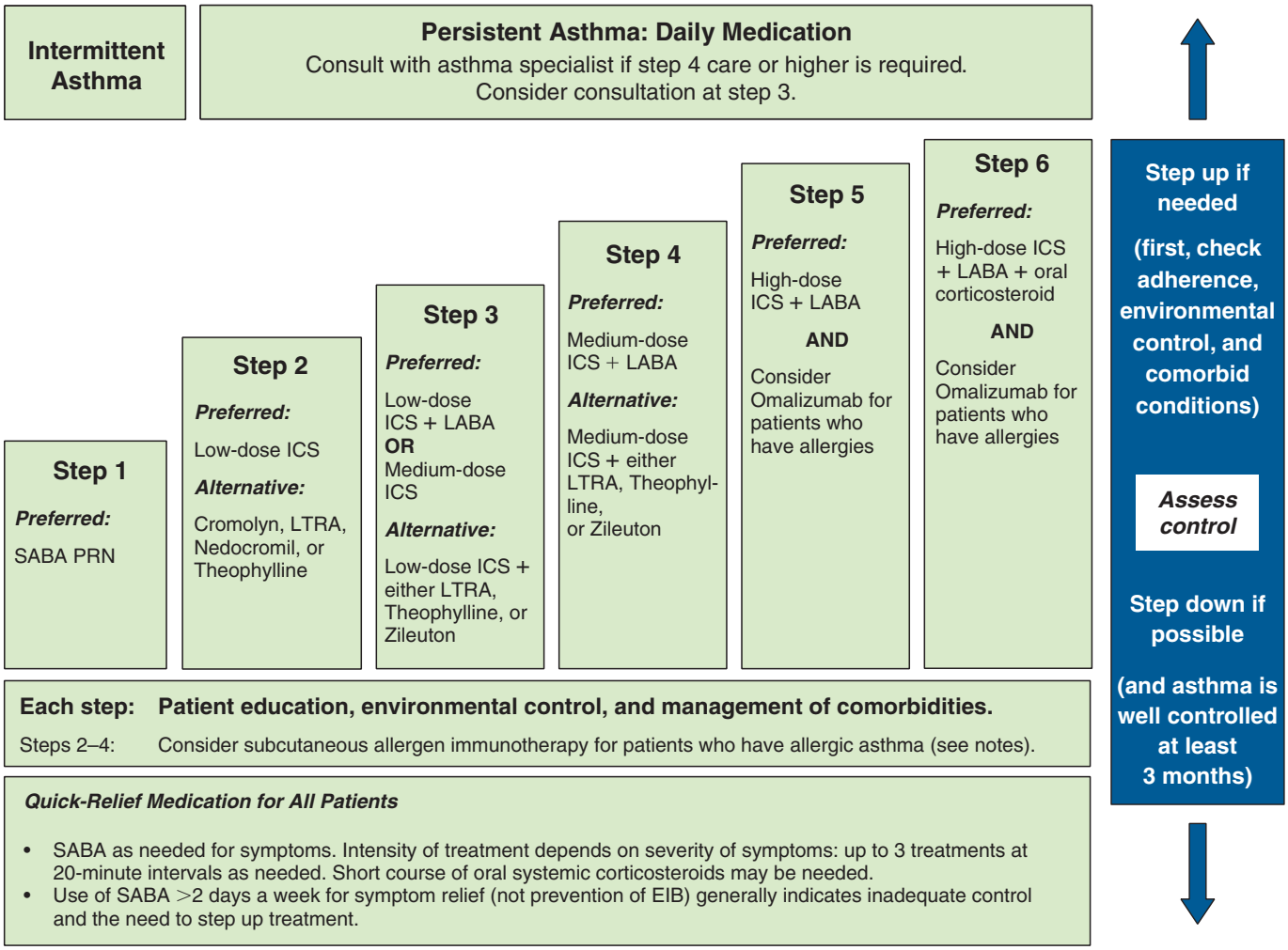
- The stepwise approach is meant to assist, not replace, the clinical decision making required to meet individual patient needs.
- The level of control is based on the most severe impairment or risk category. Assess impairment domain by patient's recall of previous 2–4 weeks and by spirometry/peak flow measures. Symptom assessment for longer periods should reflect a global assessment, such as inquiring whether the patient's asthma is better or worse since the last visit.
- At present, there are inadequate data to correspond frequencies of exacerbations with different levels of asthma control. In general, more frequent and intense exacerbations (e.g., requiring urgent, unscheduled care, hospitalization, or ICU admission) indicate poorer disease control. For treatment purposes, patients who had ≥2 exacerbations requiring oral systemic corticosteroids in the past year may be considered the same as patients who have not-well-controlled asthma, even in the absence of impairment levels consistent with not-well-controlled asthma.
- Check NIH website for information on ATAQ, ACQ, and ACT questionnaires.
- Before step up in therapy:
  - — Review adherence to medication, inhaler technique, environmental control, and comorbid conditions.
  - — If an alternative treatment option was used in a step, discontinue and use the preferred treatment for that step.

FIG 22.4 Assessing asthma control and adjusting therapy in youths ≥12 years of age and adults. (From National Institute of Health's National Heart, Lung, and Blood Institute: 2007 National Asthma Education and Prevention Program: *Expert Panel Report 3: guidelines for the diagnosis and management of asthma,* p 345, Fig. 4.7.)

Acute Bronchitis

Etiology

Acute inflammation of the trachea and bronchi is produced most commonly (80% of the 12 million cases per year in the United States) by a variety of viruses such as influenza virus A or B, parainfluenza virus, respiratory syncytial virus, coronavirus, rhinovirus, Coxsackie virus, and adenovirus. Nonviral causes include *Streptococcus pneumoniae, Haemophilus influenzae,* mycoplasma, moraxella, and *Chlamydia pneumoniae.* Numerous other pathogens as well as heat, smoke inhalation, inhalation of irritant chemicals (e.g., sulfur dioxide or chlorine, bromine, or fluorine gases), and allergic reactions have also been identified. Highest incidences are noted in smokers, young children, and the elderly, with a prevalence in the winter months. The swelling of bronchial mucosa in children associated with obstruction, respiratory distress, and wheezing



Key: **Alphabetical order is used when more than one treatment option is listed within either preferred or alternative therapy.** *EIB*, exercise-induced bronchospasm; *ICS*, inhaled corticosteroid; *LABA*, long-acting inhaled beta$_2$-agonist, *LTRA*, leukotriene receptor antagonist; *PNR,* as needed; *SABA*, inhaled short-acting beta$_2$-agonist.

Notes:

- The stepwise approach is meant to assist, not replace, the clinical decision making required to meet individual patient needs.
- If alternative treatment is used and response is inadequate, discontinue it and use the preferred treatment before stepping up.
- Zileuton is a less desirable alternative due to limited studies as adjunctive therapy and the need to monitor liver function. Theophylline requires monitoring of serum concentration levels.
- In step 6, before oral systemic corticosteroids are introduced, a trial of high-dose ICS + LABA + either LTRA, theophylline, or zileuton may be considered, although this approach has not been studied in clinical trials.
- Steps 1, 2, and 3 preferred therapies are based on Evidence A; step 3 alternative therapy is based on Evidence A for LTRA, Evidence B for theophylline, and Evidence D for zileuton. Step 4 preferred therapy is based on Evidence B, and alternative therapy is based on Evidence B for LTRA and theophylline and Evidence D for zileuton. Step 5 preferred therapy is based on Evidence B. Step 6 preferred therapy is based on (EPR—2 1997) and Evidence B for omalizumab.
- Immunotherapy for steps 2–4 is based on Evidence B for house-dust mites, animal danders, and pollens; evidence is weak or lacking for molds and cockroaches. Evidence is strongest for immunotherapy with single allergens. The role of allergy in asthma is greater in children than in adults.
- Clinicians who administer immunotherapy or omalizumab should be prepared and equipped to identify and treat anaphylaxis that may occur.

FIG 22.5 Stepwise approach for managing asthma in youths ≥12 years of age and adults. (From National Institute of Health's National Heart, Lung, and Blood Institute: 2007 National Asthma Education and Prevention Program: *Expert Panel Report 3: guidelines for the diagnosis and management of asthma,* p 343, Fig. 4.5.)

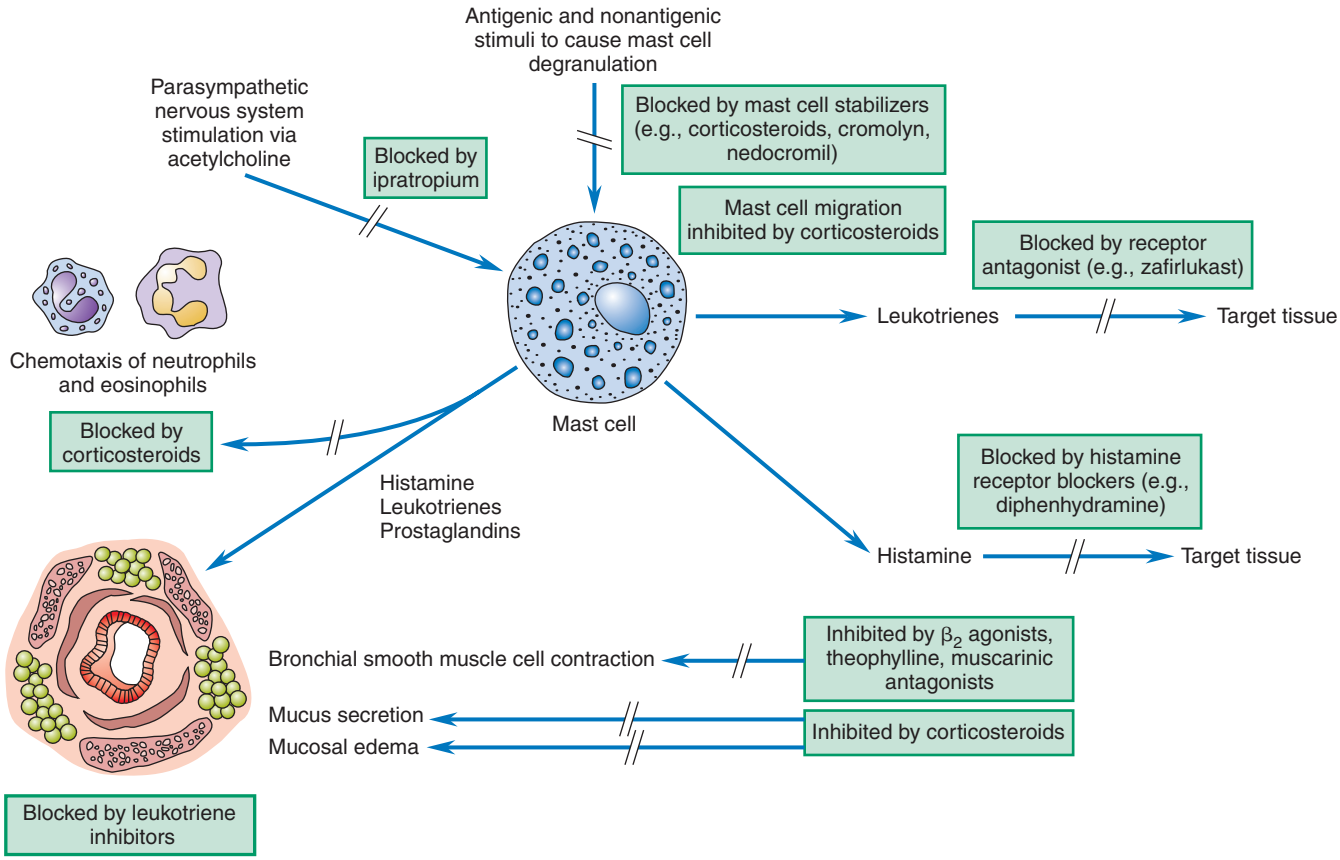


FIG 22.6 Pathophysiologic basis of asthma and site of action of drugs used in management.

is known as *asthmatic bronchitis.* Acute bronchitis in children seldom occurs as a primary bacterial infection, but is usually secondary to a cold or another communicable disease. Acute bronchitis differs from bronchiolitis in the size of the airways affected (i.e., trachea and bronchi as opposed to the small bronchioli).

Pathogenesis

The airways become inflamed and narrowed from capillary dilation, swelling from exudation of fluid, infiltration with inflammatory cells, increased mucus production, loss of ciliary function, and loss of portions of the ciliated epithelium. Many viruses and mycoplasmal bacteria inhibit macrophages and lymphocytes, temporarily promoting secondary bacterial invasion. Microorganisms may also induce long-lasting hyperirritability of the respiratory tract with associated episodes of bronchospasm.

Clinical Manifestations

The presentation of acute bronchitis is usually mild and self-limited, requiring only supportive treatment. Cough may be productive or nonproductive. Associated symptoms include low-grade fever, substernal chest discomfort, sore throat, postnasal drip, and fatigue. In children, the smaller airways are easily obstructed by inflammation so that severe obstruction may occur. The smallness of airways in proportion to body size is due to a smaller lumen in relation to the vessel wall. Associated inflammation of the larynx and trachea produces croup (see the "Croup Syndrome" section in this chapter for further details).

Diagnosis

Diagnosis of acute bronchitis is usually based on the clinical presentation, with recent onset of cough being the distinctive hallmark. Neither the appearance of purulent sputum nor the determination of an increased WBC count is a reliable diagnostic indicator. A chest radiograph may be helpful to distinguish acute bronchitis (normal radiograph) from pneumonia (pulmonary infiltrates on radiograph).

Treatment

Acute bronchitis is predominantly caused by viruses (rhinovirus, coronavirus, adenovirus, influenza virus). Viral infections do not respond to antimicrobial therapy, and symptoms resolve spontaneously in most normal, otherwise healthy individuals. Acute bronchitis caused by bacterial organisms responds well to antibiotic therapy. The primary goal in management is symptom management with antitussive agents, beta-2 agonists, and other classes of bronchodilators in selective patients. Codeine-containing medications are helpful in relieving the cough associated with bronchitis that interferes with sleep. Nonpharmacologic recommendations are to increase fluid intake, avoid smoke, and use a vaporizer in the bedroom.

The dangers of acute bronchitis include the potential for bacterial invasion, which can worsen symptoms in patients with COPD and precipitate serious infections in elderly patients or those with debilitating disease.

Chronic Bronchitis

Etiology

The next two sections of this chapter present chronic bronchitis and emphysema. Characteristic pathologic and clinical findings are described for each of these classifications. Clinically, pure forms of emphysema and chronic bronchitis are rare, and most patients present with a combination of both of these obstructive processes. Patients with emphysema and chronic bronchitis constitute most cases of COPD.

The major causes of chronic bronchitis are cigarette smoking (90% of cases), repeated airway infections, genetic predisposition, and inhalation of physical or chemical irritants.

Chronic bronchitis (also referred to as *type B COPD*) is diagnosed *symptomatically* by hypersecretion of bronchial mucus and a chronic or recurrent productive cough of more than 3 months' duration and occurring each year for 2 or more successive years in patients in whom other causes have been excluded. For patients with chronic bronchitis and emphysema, airway obstruction is persistent and irreversible. The National Center for Health Statistics reports a 3:1 ratio of annual cases of chronic bronchitis to emphysema.

Pathogenesis

Pathologic changes in the airway include chronic inflammation and swelling of the bronchial mucosa resulting in scarring, increased fibrosis of the mucous membrane, hyperplasia of bronchial mucous glands and goblet cells, hypertrophy of bronchial glands and goblet cells, and increased bronchial wall thickness, which potentiates obstruction to airflow. Inflammation appears to predominantly be the result of neutrophil activity. Interleukin-8 levels are elevated, indicating sustained attraction of neutrophils to the site of inflammation. CD8 T-lymphocyte levels are also elevated. During acute exacerbations, bronchial biopsy specimens have a 30-fold increase in the number of eosinophils. Figs. 22.7 and 22.8 show the histologic changes seen in chronic bronchitis.

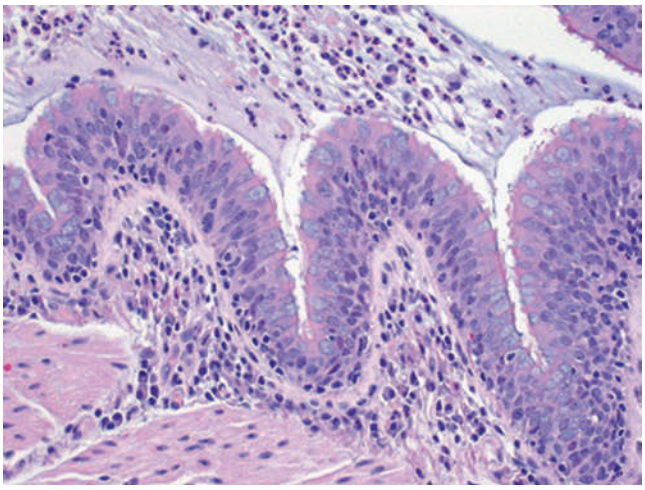


FIG 22.7 Chronic bronchitis. Low-power view shows bronchus with goblet cell and mucous gland hyperplasia. Note that some of the mucous glands are small due to adjacent fibrosis. (From Husain A: *Thoracic pathology,* St Louis, 2012, Saunders.)

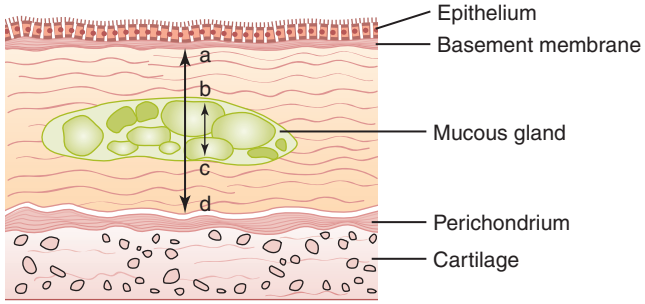


FIG 22.8 Structure of a normal bronchial wall. In chronic bronchitis, the thickness of the mucous glands increases and can be expressed as the Reid index, given by the following formula: (b – c)/(a – d). The ratio is normally less than 0.4. A ratio of 0.7 indicates severe bronchitis.

Hypertrophy of mucosal glands and goblet cells leads to increased mucus production; the mucus then combines with purulent exudate to form bronchial plugs. Chronic bronchitis patients often display bacterial colonization with *H. influenzae* and *S. pneumoniae.* The mucociliary clearance action is impaired or lost, and some areas of ciliated columnar epithelium are replaced by squamous cells. Ciliary dysfunction occurs because of a decreased number of cilia and decreased action of available cilia.

Often the inflammatory and fibrotic changes extend into the surrounding alveoli. The narrowed airways and the mucous plugs prevent proper oxygenation and potentiate airway obstruction. High airflow resistance increases the work of breathing, leading to increased oxygen demands. In areas of greater obstruction to airflow, alveoli empty and fill more slowly, leading to ventilation–perfusion ($\dot{V}_A/\dot{Q}$) mismatch, thus lowering arterial oxygenation. The chronic bronchitis patient may appear as the "blue bloater" (Fig. 22.9), characterizing the pathophysiologic process of oxygen desaturation (cyanosis) and edema associated with right-sided heart failure in advanced disease or exacerbations.

The involvement of small pulmonary arteries related to inflammation in the bronchial walls and the compensatory vasoconstriction of pulmonary blood vessels from hypoxia produce pulmonary hypertension. In addition, widespread bronchial narrowing and mucous plugging produce ventilation–perfusion mismatch with hypoxemia and hypercarbia from impeded ventilation. The combination of hypoxia and hypercarbia increases pulmonary artery resistance and pulmonary hypertension. While the process of pulmonary hypertension continues, right ventricular end-diastolic pressures increase, leading to right ventricular dilation (cor pulmonale) and right-sided heart failure. An enlarged right heart results in increased venous pressure, liver engorgement, and dependent edema. Manifestations of heart failure may occur during exacerbations of bronchitis and subside with appropriate treatment.

Destruction of bronchial walls results in dilation of airway sacs. This is termed *bronchiectasis.* Causes of bronchial wall destruction include infection from severe streptococcal or staphylococcal pneumonia, repeated bouts of acute bronchitis, infection with the mold *Aspergillus fumigatus,* presence of mucous plugs or foreign bodies, or deficiencies in immunologic response. (Refer to the "Bronchiectasis" section later in this chapter for a more detailed description of this disease process.) The dilated sacs contain pools of infected secretion that do not clear themselves and serve as sources of further infection that can spread to adjacent lung fields by the lymphatics or venous drainage to other areas of the body, commonly the brain. If bronchiectatic lesions are localized, surgical resection of the affected portions of lung may be helpful.

Clinical Manifestations

The typical patient is an overweight man or woman (1:2 male-to-female ratio) in his or her 30s or 40s (or older) who presents with shortness of breath on exertion, excessive amounts of sputum, chronic cough, evidence of excess body fluids (edema, hypervolemia), and a history of smoking. In addition, the patient often complains of chills, malaise, muscle aches, fatigue, loss of libido, and insomnia.

Sputum production may be variable and worsens with respiratory tract infection. Cough and sputum production are most severe in the mornings. Gradually, patients develop progressive shortness of breath on exertion. Most patients do not seek help until dyspnea becomes troublesome. By the time dyspnea on exertion is present, the disease is well advanced.

In the end-stage disease process, the patient presents with signs of right-sided heart failure (distended neck veins, right ventricular heave, right ventricular gallop, and peripheral edema). Hypoxia leads to pulmonary hypertension. Cyanosis is a late sign.

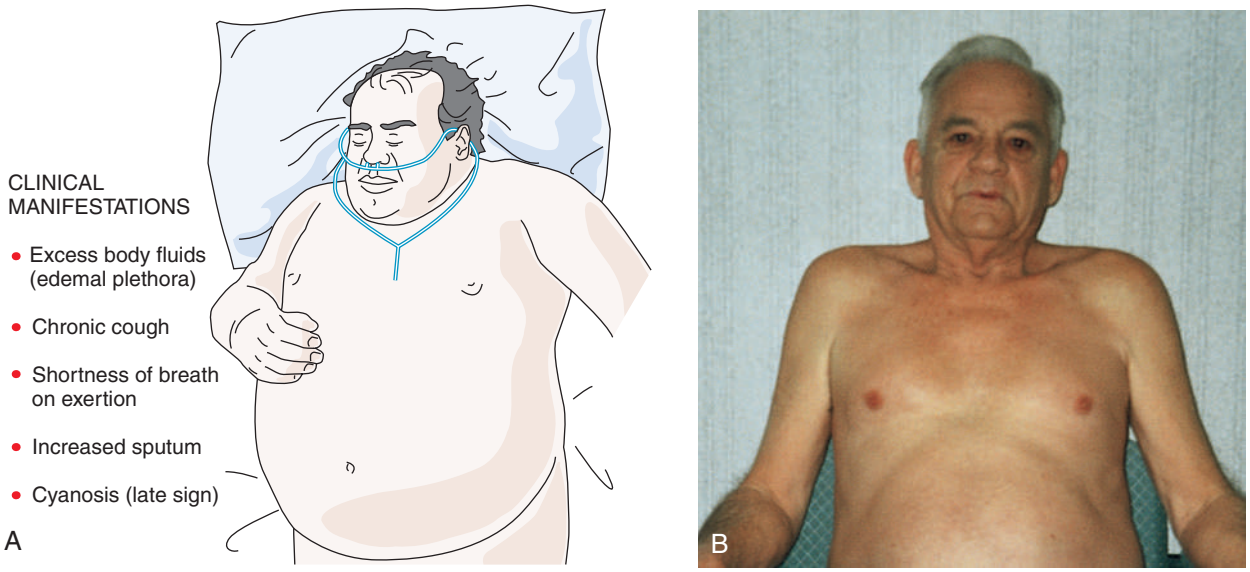


FIG 22.9 A, A hypoxemic patient with edema from right-sided heart failure. **B,** A patient with chronic obstructive bronchitis. Note the stocky build and the presence of pursed-lip breathing. The slight gynecomastia is a side effect of corticosteroid therapy. The patient's shoulders are raised because of shortness of breath and increased work of breathing. (B, From Black JM, Hawks SJ: *Medical-surgical nursing: clinical management for positive outcomes,* ed 8, Philadelphia, 2008, Saunders, p 1581.)

Diagnosis

Measures used to confirm the diagnosis include chest radiography, which may show increased bronchial vascular markings, congested lung fields, an enlarged horizontal cardiac silhouette, and evidence of previous pulmonary infection. Pulmonary function tests show normal total lung capacity (TLC), increased residual volume (RV), and decreased FEV_1. Early pulmonary function testing before the onset of symptoms shows increased closing volume and a decrease in the maximal midexpiratory flow rate. Arterial blood gas (ABG) evaluation may show elevated $Paco_2$ and decreased Pao_2 (often below 65 mm Hg); abnormal ABGs develop early in the disease process. The electrocardiogram may reveal atrial dysrhythmias and evidence of right ventricular hypertrophy. Secondary polycythemia (increased numbers of red blood cells) related to continuous or nocturnal hypoxemia is common. Hypoxemia leads to a compensatory production of red blood cells in an attempt to carry more oxygen to the body tissues.

Depending on the severity of the disease, the physical examination may reveal scattered crackles, rhonchi, and wheezes; use of accessory muscles to breathe; jugular vein distention; clubbing; and pedal and ankle edema. Table 22.1 lists the distinguishing features of both emphysema and chronic bronchitis.

Treatment

Because bronchitis and emphysema are most frequently seen in combination, the therapies are similar. The overall goals are to (1) block the progression of the disease, (2) return the patient to optimal respiratory function, and (3) return the patient to usual activities of daily living.

Pharmacologic treatment involves the use of inhaled short-acting β_2 agonists and inhaled anticholinergic bronchodilators, cough suppressants, and antimicrobial agents for infections. Inhaled or oral corticosteroids may also be used in the treatment of some patients for acute exacerbations. Theophylline products are used less frequently because of their narrow therapeutic range and toxicity. However, many patients derive significant benefits from theophylline.

Low-dose oxygen therapy is recommended for patients with Pao_2 levels less than 55 mm Hg. Mechanical ventilation may become necessary to get the patient over a crisis period of acute exacerbation. Although traditionally the mechanism of carbon dioxide retention with oxygen therapy was thought to be related to a diminished ventilatory drive, current research suggests that oxygen therapy may instead cause increased $\dot{V}_A/\dot{Q}$ imbalance, precipitating a rise in carbon dioxide concentration. It is important to remember that not all patients with a history of COPD are carbon dioxide retainers and most can use oxygen safely. Home oxygen therapy has been demonstrated to retard the development of pulmonary hypertension and cor pulmonale in chronic bronchitis. Portable oxygen saturation monitors for evaluating the effectiveness of oxygen administration at home may also be used.

Smoking cessation is essential to decreasing the progression of the disease. A reduction in exposure to inhaled pulmonary irritants is also advised. Supportive therapies include adequate rest, proper hydration (8 to 12 glasses of water per day unless the patient has congestive heart failure), and physical reconditioning programs using a treadmill or stationary bicycle. Alternating rest and exercise improves results on pulmonary function tests. Walking has proved to be the best form of exercise for increasing duration and intensity of activity. All COPD patients also benefit from yearly influenza vaccines and pneumococcal vaccine.

KEY POINTS

- Acute bronchitis results from temporary inflammation of the tracheobronchial tree. Inflammation may be due to viral, bacterial, fungal, or chemical causes. Symptoms are caused by narrowing of inflamed airways and increased mucus production. Dyspnea on exertion and cough are common.
- Chronic bronchitis is an inflammatory disorder of the airways that most commonly results from long-term cigarette smoking. It is defined as a productive cough lasting more than 3 months per year for 2 or more consecutive years. Resultant airway damage is not reversible.

- Chronic bronchitis is associated with persistent narrowing of the airways attributable to chronic inflammation, scarring, and excessive mucus production. Airway obstruction leads to poor ventilation of alveoli and impaired exchange of oxygen and carbon dioxide. Blood gases are characterized by low Pao_2 and high $Paco_2$ values. Persistent hypoxemia causes a compensatory increase in red blood cell production (polycythemia). Cyanosis may be evident.
- Alveolar hypoxia leads to generalized pulmonary vasoconstriction, pulmonary hypertension, and right ventricular hypertrophy (cor pulmonale) in the person with chronic bronchitis. Right-sided heart failure may occur because of the high pulmonary resistance.
- The management of chronic bronchitis centers on removing the etiologic factors (e.g., cigarette smoke), providing bronchodilator therapy, removing secretions, preventing respiratory muscle fatigue, and providing low-dose supplemental oxygen. High-dose oxygen must be used cautiously because it may increase $\dot{V}_A/\dot{Q}$ imbalance and $Paco_2$ levels in some patients.

OBSTRUCTION RELATED TO LOSS OF LUNG PARENCHYMA

Emphysema

Etiology

Emphysema (also referred to as *type A COPD*) is defined *pathologically* by destructive changes of the alveolar walls and abnormal enlargement of the distal air sacs. Emphysema is frequently associated with chronic bronchitis. According to the National Center for Health Statistics, 15 million Americans have COPD, with an associated mortality of more than 120,000 deaths each year. Within the two distinct subgroups of COPD, the majority (approximately 10 million) have chronic bronchitis and the remainder have emphysema. The etiologies of emphysema include smoking, air pollution, certain occupations (e.g., welding, mining, and working with or near asbestos), and α_1-antitrypsin deficiency (1%). Emphysema tends to develop over a long period and thus is seen more

TABLE 22.1 Common Distinguishing Features of Emphysema and Chronic Bronchitis*

Patient Data	Emphysema (COPD Type A)	Bronchitis (COPD Type B)
History		
Lifestyle	Smoker	Smoker
Weight	Weight loss	Overweight
Onset of symptoms	Usually after age 50 years	Usually after age 40 years
Sputum	Mild, mucoid	Excessive, purulent
Cough	Minimal or absent	Chronic; more severe in mornings
Dyspnea	Progressive exertional dyspnea	Mild to moderate, but may gradually progress to severe exertional dyspnea
Patient complaints	Dyspnea on exertion, fatigue, insomnia	Chronic cough with mucopurulent sputum, chills, malaise, muscle aches, fatigue, insomnia, loss of libido
Physical Signs		
Edema	Absent	Present
Central cyanosis	Absent	Present in advanced disease
Use of accessory muscles to breathe	Present	Absent until end stage
Body build	Thin, wasted	Stocky, overweight
Anteroposterior chest diameter	"Barrel chest," 1:1 ratio anteroposterior chest diameter	Normal
Auscultation of chest	Decreased breath sounds, decreased heart sounds, prolonged expiration	Wheezes, crackles, rhonchi, depending on severity of disease
Percussion	Hyperresonance	Normal
Jugular vein distention	Absent	Present
Other	Pursed-lip breathing	Evidence of right-sided heart failure (cor pulmonale)
General Diagnostic Tests		
Chest radiography	Narrowed mediastinum; normal or small vertical heart; hyperinflation; low, flat diaphragm; presence of blebs or bullae	Congested lung fields, increased bronchial vascular markings, enlarged horizontal heart
Arterial blood gas analysis	Decreased Pao_2 (60–80 mm Hg); increased $Paco_2$ with advancing disease	Decreased Pao_2 (<65 mm Hg); increased $Paco_2$
Electrocardiography	Normal or tall symmetric P waves; tachycardia, if hypoxic	Right axis deviation, right ventricular hypertrophy, atrial dysrhythmias
Hematocrit	Normal	Polycythemia
Pulmonary Function Tests		
Functional residual capacity	Increased	Normal or slight increase
Residual volume	Increased	Increased
Total lung capacity	Increased	Normal
Forced expiratory volume	Decreased	Decreased
Vital capacity	Decreased	Normal or slight decrease
Static lung compliance	Increased	Normal

COPD, Chronic obstructive pulmonary disease.
*Clinically features of bronchitis and emphysema are not clear-cut because most patients with COPD have a combined disease process.

frequently in persons older than 50. Cigarette smoking in excess of 70 pack-years is highly predictive of COPD. The normal aging process, starting at about age 30, reflects changes similar to those seen in emphysema, including a loss of alveoli, an increase in the size of alveolar ducts, a loss of gas-exchanging surface area (4% per decade), and a decrease in bronchiolar musculature.

When emphysema occurs in young to middle-aged adults or before the age of 50 in a smoker, it may be associated with a deficiency of α_1-antitrypsin activity in the lung. α_1-Antitrypsin deficiency is a hereditary disorder characterized by low serum levels (25 to 50 mg/dL) of α_1-antitrypsin. α_1-Antitrypsin is a protective enzyme that inhibits proteolytic breakdown of alveolar tissue. The protease enzymes (neutrophil-derived elastase) that break down lung protein are released from neutrophils that migrate to the lung during inflammation, causing alveolar wall destruction.

Emphysema may follow bacterial lung infections that involve secretion of proteases that destroy the elastin proteins responsible for the normal elasticity of the lung tissue. Bacterial infections block mechanisms that normally inhibit the release of proteolytic enzymes from degenerating neutrophilic granulocytes.

Pathogenesis

The pathologic changes leading to alveolar destruction are associated with the release of proteolytic enzymes from inflammatory cells such as neutrophils and macrophages. Smoking is commonly associated with emphysema. Smoking causes alveolar damage in two ways: (1) it leads to inflammation in the lung tissue (parenchyma), thus initiating a chain of events leading to the release of proteolytic enzymes that directly damage alveolar tissue; and (2) it inactivates α_1-antitrypsin, which normally acts to protect the lung parenchyma. Fig. 22.10 illustrates the pathogenesis of emphysema.

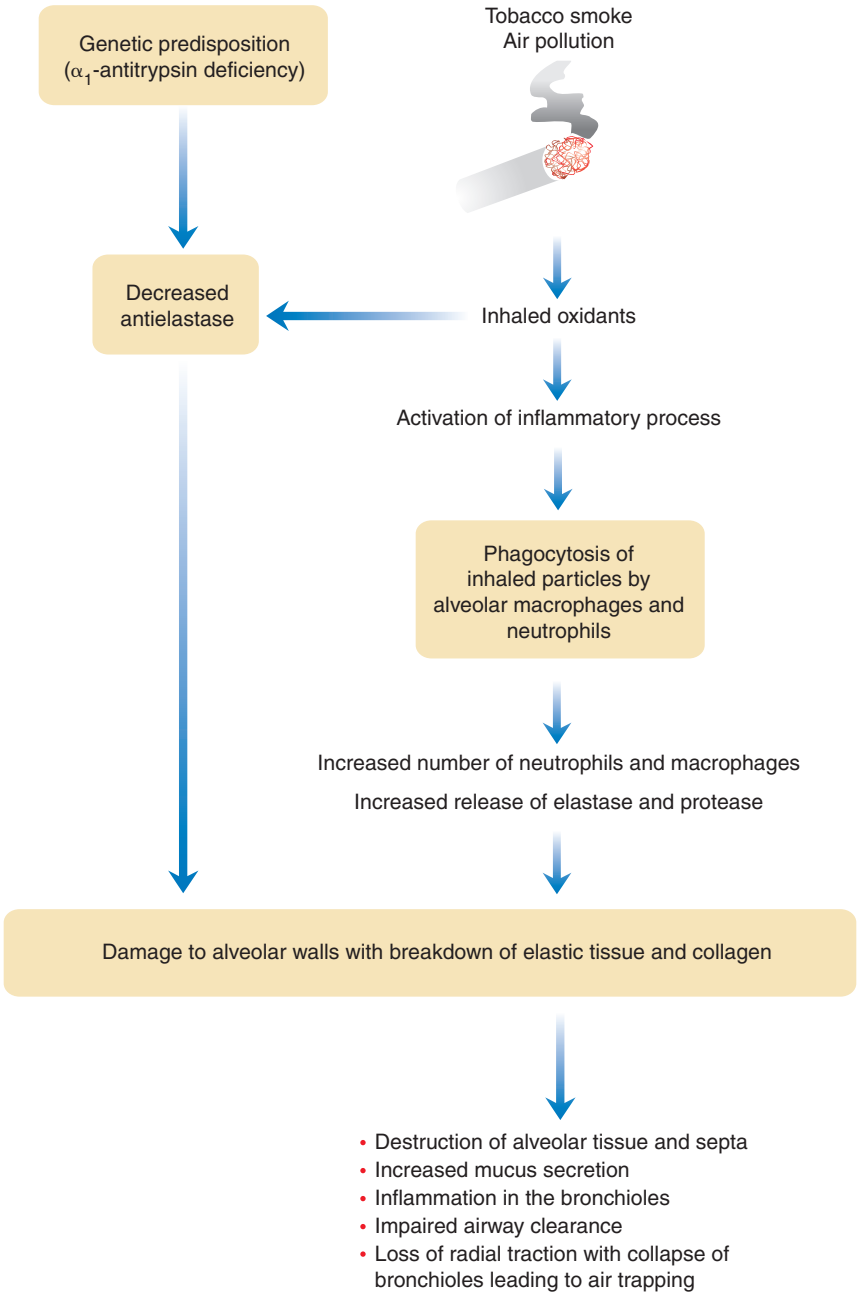


FIG 22.10 Pathogenesis of smoke-induced emphysema.

With the loss of alveolar walls, there is also a marked reduction in the pulmonary capillary bed, which is essential for exchange of oxygen and carbon dioxide between the alveolar air and capillary blood. There is also a loss of elastic tissue in the lung, which leads to a decrease in the size of the smaller bronchioles. The loss of lung tissue leads to a loss of *radial traction,* which normally holds the airway open, and to increasing pressure around the outside of the airway lumen, which in turn increases airway resistance and decreases airflow. Fig. 22.11 shows the effect of decreased radial traction on the size of small bronchioles. Air then becomes trapped in distal alveoli, leading to distended air sacs, which adds to the collapsing pressure on more proximal bronchi and increases airway obstruction. Loss of alveolar walls and air trapping leads to the formation of bullae (large, thin-walled cysts in the lung) that further rob the lung of its gas transport function. The histologic appearance of the lung and lung tissue from typical emphysematous patients is shown in Figs. 22.12 and 22.13.

Three major classifications of emphysema exist: (1) centriacinar (also called *centrilobular*), which is associated with both smoking and chronic bronchitis and destroys the respiratory bronchioles; (2) panacinar (also called *panlobular*), which destroys the alveoli; and (3) paraseptal, which affects the peripheral lobules. Some of the classifications of emphysema and the topographic distribution of emphysema in lung tissue are shown in Fig. 22.14.

Clinical Manifestations

Patients with emphysema commonly seek help because of progressive exertional dyspnea. The typical patient with advanced disease is a thin man or woman around 55 years of age who has complained of increasing shortness of breath for the past 3 to 4 years. As with chronic bronchitis, the incidence of emphysema is increasing in women who smoke. Patients become thin due to increased respiratory effort and resulting caloric expenditure with decreased ability to consume adequate calories. The difficulty in breathing is evidenced by the use of accessory muscles to breathe, progressive dyspnea, and the use of pursed-lip breathing in an effort to exhale more air over a longer period before the small airways collapse. Cough may be minimal or absent. Digital clubbing is common. The appearance of overinflation (barrel chest) is from an increase in lung volume, which can be confirmed by pulmonary function testing. With the loss of alveolar walls and the formation of bullae, the patient is at risk for developing a pneumothorax and may present with chest pain on the affected side and dyspnea. Decreased arterial oxygen saturation remains minor until late in the course of the disease. Late in the disease process, the major symptom is dyspnea on exertion. These patients may be referred to as "*pink puffers*" (Fig. 22.15), a term related to the physiologic matching of ventilation and perfusion that allows near-normal gas exchange. Ventilation–perfusion matching and a sustained high respiratory effort produce a relatively normal arterial oxygen level until late stages of the disease.

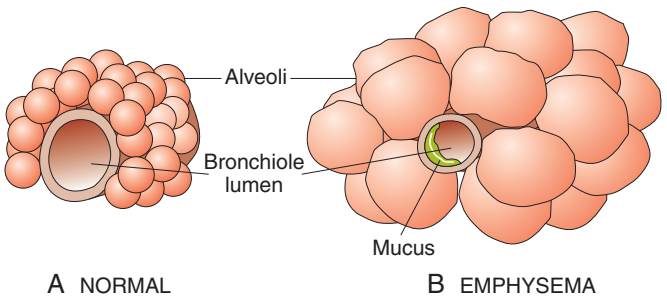


FIG 22.11 Loss of radial traction in emphysema leads to airway collapse. **A,** Terminal bronchiole in cross-section. **B,** Terminal bronchiole with narrowed lumen resulting from loss of surrounding alveoli, leading to decreased radial traction and airway collapse.

Diagnosis

The diagnosis of emphysema is based on the patient's history and physical findings, pulmonary function tests, chest radiographs, ABGs, and electrocardiogram. Changes seen on pulmonary function tests include an increased functional residual capacity, increased RV, increased TLC, decreased FEV_1, and decreased FVC. Chest radiographs show hyperinflation; a low, flat diaphragm; the presence of blebs or bullae; a narrow mediastinum; and a normal or small "vertical" heart (see Table 22.1). Electrocardiographic findings may be normal or show tall P waves. Sinus tachycardia may be the first sign of decreased oxygenation. Supraventricular dysrhythmias (atrial tachycardia, atrial flutter, and atrial fibrillation) and ventricular irregularities may also occur. ABG values typically reveal a mild decrease in Pao_2 (60 to 80 mm Hg) and a low or normal $Paco_2$ until late stages.

Physical examination shows a thin, wasted individual who is using accessory muscles to breathe and sits slightly hunched forward in an effort to breathe better. Auscultation and percussion of the lung fields reveal decreased breath sounds and lack of crackles and rhonchi,

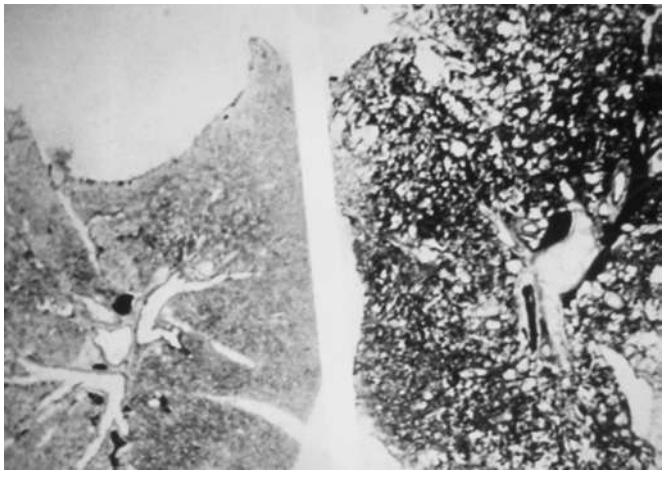


FIG 22.12 Gross appearance of emphysematous lung. *Left,* Normal lung tissue from a nonsmoker. *Right,* Lung tissue from a smoker who has developed emphysema.

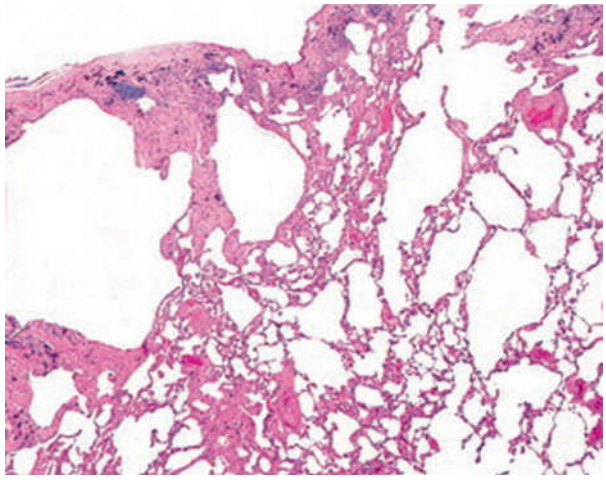


FIG 22.13 Pulmonary emphysema. There is a marked enlargement of air spaces with thinning and destruction of alveolar septa. (From Young B, Stewart W, O'Dowd G: *Wheater's basic pathology: a text, atlas, and review of histopathology,* ed 5, 2011, Churchill Livingstone.)

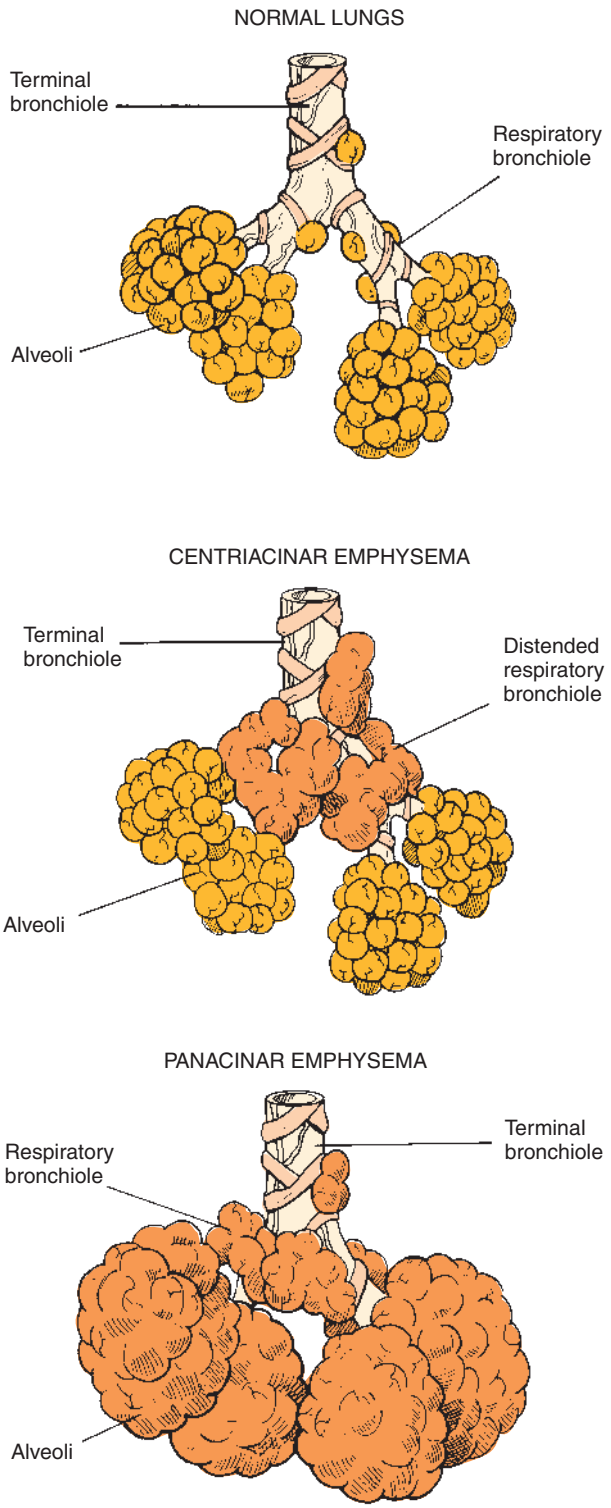


FIG 22.14 Types of emphysema. (From Black JM, Hawks SJ: *Medical-surgical nursing: clinical management for positive outcomes,* ed 8, Philadelphia, 2008, Saunders, p 1580.)

decreased heart sounds, prolonged expiration, decreased diaphragmatic excursion, and hyperresonance of the chest. Pursed-lip breathing, chronic morning cough because of mucus buildup at night, and an increased anteroposterior chest diameter (barrel chest) are also common findings. Weight loss occurs because of anorexia and lack of energy to eat. Bronchoconstriction leads to wheezing.

Treatment

Refer to the "Treatment" section under "Chronic Bronchitis" earlier in this chapter for detailed treatment modalities common to both chronic obstructive lung diseases. Poor prognosis is associated with weight loss, so treatment is focused on maintaining proper nutrition.

KEY POINTS

- Emphysema is a form of COPD that results from destruction of alveoli and small airways. Emphysema occurs primarily in cigarette smokers and is often seen in association with chronic bronchitis.
- Alveolar destruction is due to release of inflammatory proteolytic enzymes that degrade lung proteins. Smoking also inhibits a protective enzyme, α_1-antitrypsin, that normally inhibits the proteolytic enzymes. Genetic deficiency of α_1-antitrypsin is an uncommon possible cause of emphysema.
- Emphysema causes two major problems with respiration: (1) a decrease in surface area for gas exchange and (2) airway collapse attributable to loss of radial traction. Airway collapse is greater on expiration, resulting in air trapping and hyperinflation.
- Emphysema is characterized by dyspnea; weight loss; use of accessory muscles to breathe; a low, flat diaphragm; and a barrel chest. Cyanosis is not present until late stages of the disease. By sustaining high ventilatory effort, a patient can have blood oxygen levels that are generally maintained at near normal. Carbon dioxide levels may be normal or low as a result of hyperventilation until late in the disease.
- Therapy for emphysema is similar to that for chronic bronchitis. Cessation of smoking is necessary to prevent progression of the disease. Present damage is irreversible. Oxygen therapy improves activity tolerance and quality of life.

OBSTRUCTION OF THE AIRWAY LUMEN

Bronchiectasis

Etiology

Bronchiectasis means dilation of bronchi. It is either acquired or congenital and is classified as both an obstructive and a suppurative (pus-forming) disorder. Acquired bronchiectasis is now rare in the United States because of rapid diagnosis and management of bronchopulmonary infections. Fifty percent of the cases of bronchiectasis are associated with cystic fibrosis. Children are at higher risk for development of bronchiectasis because of anatomic factors such as small, soft, elastic bronchi. Bronchi in children are easily damaged by overinflation and distention from inflammation and infection.

Bronchiectasis can be classified according to bronchial shape: saccular (with cavity-like dilatations) or cylindrical and with widening of the bronchial walls. A fusiform shape is a combination of saccular and cylindrical changes. These anatomic changes are shown in Fig. 22.16. Little clinical or pathophysiologic difference in the three types has been demonstrated.

Pathogenesis

Bronchiectasis is characterized by recurrent infection and inflammation of bronchial walls, which leads to persistent dilatation of the medium-sized bronchi and bronchioles. Inflammation results in destruction of the walls of central bronchi and obliteration of peripheral bronchi and bronchioles. *H. influenzae* is the most common cause of bacterial infections. The destructive process leads to loss of ciliated columnar epithelium, with transformation to a squamous cell and pus formation, which in turn leads to obstruction of airflow. Lung tissue of a patient

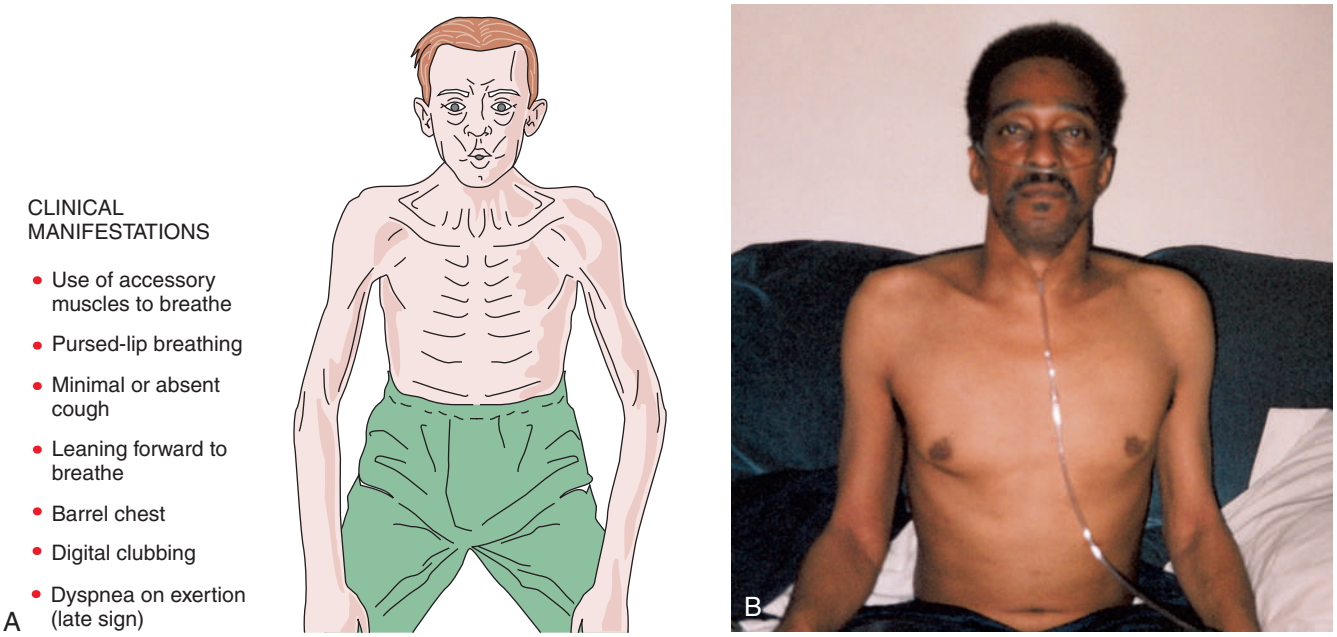


FIG 22.15 **A,** A patient with near-normal PaO_2. Note the use of accessory muscles and pursed-lip breathing in an effort to get more air out of the lungs. **B,** A patient with emphysema. Note the thin appearance and the presence of continuous oxygen therapy. The use of accessory muscles of respiration (neck and shoulder muscles) reflects the patient's shortness of breath and increased work of breathing necessary to increase minute ventilation and to maintain adequate arterial blood gas values. (B, From Black JM, Hawks SJ: *Medical-surgical nursing: clinical management for positive outcomes,* ed 8, Philadelphia, 2008, Saunders, p 1581.)

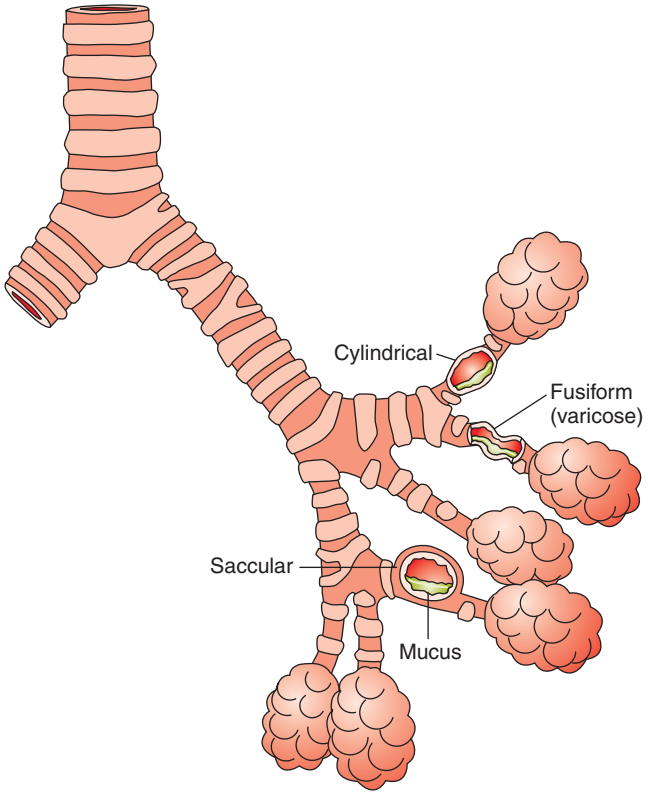


FIG 22.16 Bronchial dilatations attributable to bronchiectasis. The saccular form occurs in the segmental bronchi, which are severely dilated and end blindly. The varicose form resembles varicose veins with irregular dilatations and constriction. The cylindrical form shows uniform slight dilatation.

with cystic fibrosis complicated by varicose bronchiectasis is shown in Fig. 22.17.

Clinical Manifestations

The child usually presents with a chronic productive cough with copious amounts of purulent, foul-smelling, green or yellow sputum. The sputum has the characteristic of separating into three distinct layers in a sputum cup. Other clinical features are hemoptysis; fever; night sweats; moist crackles, rhonchi; halitosis (bad breath); skin pallor; and, infrequently, digital clubbing. Clubbing is caused by prolonged decreased oxygenation, which leads to fibrous tissue hyperplasia in the area between the nail and distal portion of each digit. Clubbing is associated with lymphocytic extravasation, increased vascularity, and edema. The severity of clubbing parallels the severity of pulmonary disease. Digital clubbing can be identified by two methods, as seen in Fig. 22.18. Hypoxemia is seen in severe cases. Complications of bronchiectasis are malnutrition, recurrent pneumonia, right ventricular failure, and secondary visceral abscesses.

Diagnosis

Generally, the diagnosis of bronchiectasis is based on a history of chronic productive cough. The patient complains about producing copious amounts of foul-smelling, purulent sputum. Radiographic abnormalities may reveal small cysts, thickening of bronchial walls, and increased bronchial markings (areas of intensity showing bronchi, which are usually not distinct). Pulmonary function tests show decreased airflow and vital capacity in advanced cases. ABG analyses reveal decreased Pao_2 and increased $Paco_2$ values from obstruction to airflow. High-resolution computed tomography is the test of choice for diagnosing bronchiectasis.

Treatment

Antibiotic therapy accompanied by inhalation of bronchodilators followed by vigorous chest percussion and postural drainage is the mainstay of treatment. Proper hydration and nutrition are important in promoting

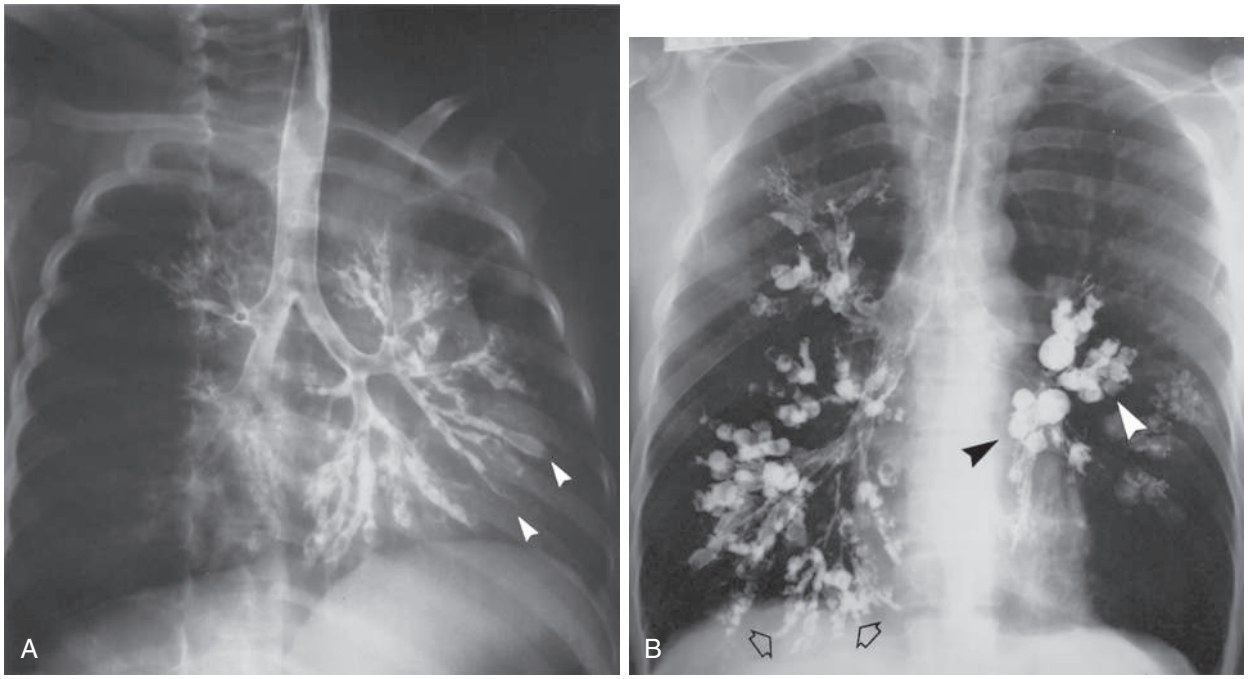


FIG 22.17 Bronchographic features of varicose and cystic bronchiectasis. **A,** A left tracheobronchogram in a shallow posterior oblique projection reveals mild dilatation and slightly irregular bronchi that terminate after four to six generations of branchings from the trachea in a squared or bulbous appearance *(arrowheads).* The findings are those of varicose bronchiectasis. **B,** A bilateral tracheobronchogram in the anteroposterior projection demonstrates a multitude of contrast material–filled cystic spaces resembling a cluster of grapes *(arrowheads),* a characteristic feature of cystic bronchiectasis. Note that the cystic spaces appear after only two to three bronchial generations. Less severe bronchiectasis of the varicose type is present in the right lower lobe *(open arrows).* (From Fraser RG et al: *Diagnosis of diseases of the chest,* ed 3, vol 3, Philadelphia, 1990, Saunders, p 2199.)

liquefaction of secretions and preventing increased susceptibility to infection resulting from malnutrition. Maintaining adequate nutrition is problematic because of fatigue and the energy required for eating. (Refer to the "Cystic Fibrosis" section later in the chapter for further discussion on treatment.) In severe cases, when other measures fail, bronchoscopy with bronchial lavage may be necessary to remove thick, purulent secretions. In the child with severe saccular bronchiectasis, removal of the affected area of the lung may be necessary. Patient education materials can be obtained from the Cystic Fibrosis Foundation. Childhood immunizations have led to a decreased incidence of bronchiectasis attributable to pertussis.

Bronchiolitis

Etiology

Bronchiolitis is characterized by widespread inflammation of bronchioles attributable to infectious agents such as respiratory syncytial virus (RSV) (50% of cases), influenza virus (type A, B, or C), or bacteria (*H. influenzae,* pneumococci, or hemolytic streptococci) and occasionally is produced by allergic reactions. RSV infection is a common cause of hospitalization in infants. Other organisms that may cause bronchiolitis include mycoplasma, chlamydia, ureaplasma, and *Pneumocystis (carinii) jiroveci.* RSV occurs in yearly epidemics in winter to spring, usually in children younger than 2 years. The average incubation period is 5 days, with inoculation occurring through the nose and eyes. In adults, bronchiolitis is commonly associated with smoking, toxic fumes, and immunosuppression.

Pathogenesis

Once initiated by the causal agent, proliferation and necrosis of bronchiolar epithelium occur, producing obstruction and increased mucus production. Production of thick, tenacious mucus leads to airway obstruction, atelectasis, and hyperinflation. Three possible mechanisms of airway obstruction may follow the inflammatory process. They include (1) development of inflammatory exudate, which may displace surfactant, leading to airway obstruction; (2) release of chemical mediators, which may produce bronchiolar constriction; and (3) development of inflammation, which may induce fibrosis and narrowing of the airway. Goblet cell metaplasia and increased bronchial muscle mass may also occur, resulting in further airway narrowing.

Clinical Manifestations

The severity and course of the disease are variable, ranging from mild to fatal. Common clinical features include wheezing attributable to bronchospasm, crackles, decreased breath sounds, retractions, increased sputum, dyspnea, tachypnea (rapid, shallow respirations), and low-grade fever. Otitis media is a common complication often associated with *S. pneumoniae.*

Diagnosis

Patients commonly have an elevated WBC count. The chest radiograph may show enlarged air sacs, interstitial infiltrates, atelectasis, or severe hyperinflation. Pulmonary function tests reveal severe obstruction to airflow. Rapid diagnosis of RSV may be made by identifying the viral antigen from nasal washings or nasal swab culture of secretions, using an enzyme-linked immunosorbent assay or immunofluorescent assay.

Treatment

Adequate oxygenation is maintained by providing humidified oxygen; monitoring blood gases or oxygen saturation; and administering oral,

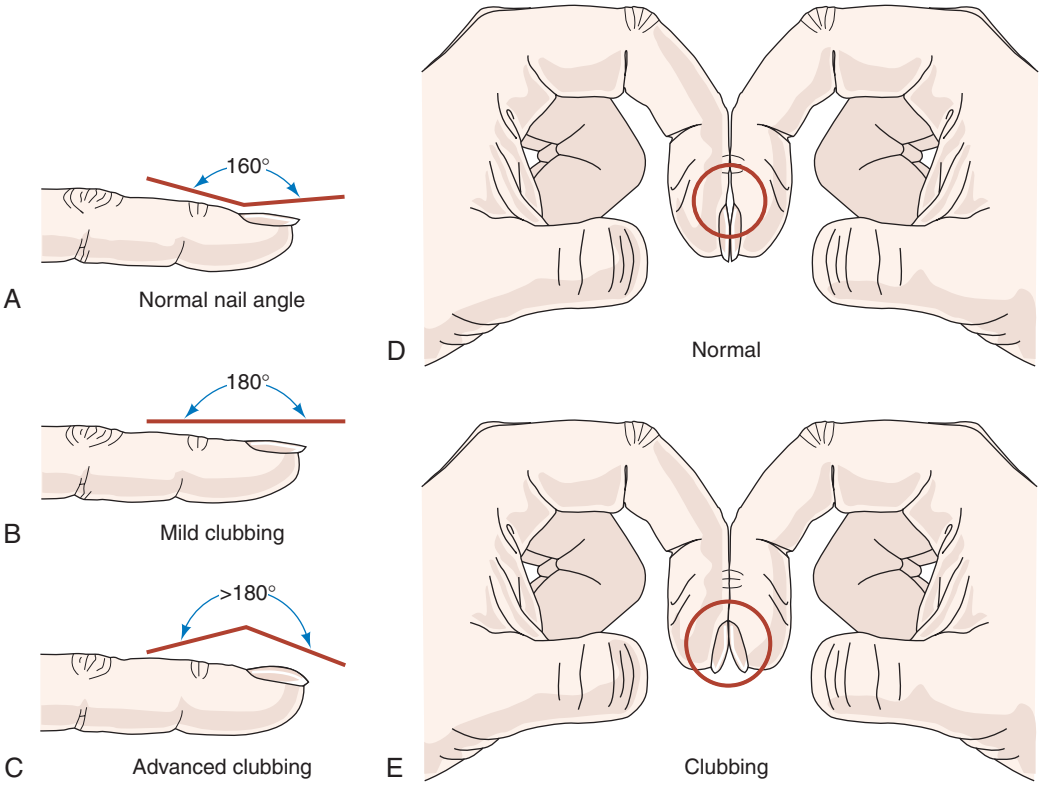


FIG 22.18 Clubbing. **A,** Normal fingernail angle is 160 degrees. **B,** Early mild clubbing appears as a flattened angle between nail and skin (180 degrees). **C,** Advanced clubbing shows a rounded (clubbed) fingertip and nail. To assess clubbing by Schamroth's diagnostic method (**D** and **E**), place the nails of the second digits together. Obliteration of the normal diamond-shaped space between the nails is an abnormal finding, signifying clubbing.

inhaled, or intravenous bronchodilator agents and, in selected cases, corticosteroids. General information about pharmacologic agents commonly used in the management of various respiratory disorders is listed in other texts. Use of these agents depends on the severity of the diagnosis and prescriber preference.

Other therapies include sedation for anxiety, hydration, and the administration of appropriate antivirals and antibiotics. Patients are encouraged to stop smoking and to avoid passive smoke exposure. The use of eye–nose goggles by health care workers is recommended to control the spread of RSV. The virus is spread through the air or by contact with secretions from the eye, nose, or mouth, and transmission may not be prevented by the use of masks and gowns.

Cystic Fibrosis

Etiology

Cystic fibrosis (mucoviscidosis) is an autosomal-recessive disorder of the exocrine glands. It is the most common genetic lung disease in the United States, with an incidence of 1 in 2000 to 3000 Caucasian births. One in 26 Caucasians is a heterozygous carrier of the cystic fibrosis gene. The incidence in African Americans is rare (1 in 10,000 live births), and in Native Americans the incidence is 1 in 80,000 births. It is almost never seen in the Asian population. About 35% of the 30,000 cases of cystic fibrosis in the United States involve individuals older than 18 years. Cystic fibrosis can be classified either as an airflow obstructive disorder or as a suppurative (pus-forming) disorder. Hypersecretion of abnormal, thick mucus that obstructs exocrine glands and ducts is a characteristic finding in the disease.

With advances in antibiotic therapy and early recognition and management of complications, patients with cystic fibrosis are living longer into adulthood. The median survival age is now 31 years. Some patients are now having families.

Pathogenesis

Cystic fibrosis is classified as an autosomal-recessive disorder. More than 800 mutations in the gene that encodes for the cystic fibrosis transmembrane conductance regulator (CFTR) have been described. One genetic defect associated with cystic fibrosis involves deletion of three base pairs in codon 508 (AF508) that code for phenylalanine on chromosome 7 (band q31). With the loss of these three base pairs, the *CFTR* gene is dysfunctional. This is the most common genetic mutation causing cystic fibrosis and occurs in 60% to 75% of cystic fibrosis patients tested. *CFTR* encodes a membrane chloride channel and is expressed in the sweat glands, the lungs, and the pancreas. Mutations in the *CFTR* gene result in alteration in chloride and water transport across the apical surface of epithelial cells. Cystic fibrosis primarily affects the pancreas, intestinal tract, sweat glands, and lungs, and in males causes infertility. The mucus-producing glands in the gastrointestinal tract enlarge, generating excessive secretions. The thick eosinophilic mucous secretions plug the glands and ducts of the pancreatic acini, intestinal glands, intrahepatic bile ducts, and the gallbladder, causing dilation and fibrosis. These changes result in decreased production of pancreatic enzymes necessary for digestion of fats, carbohydrates, and proteins, thus leading to increased fat and protein in the stool.

The bronchopulmonary system is also affected by the thick, tenacious mucus that results from failure of chloride channels to function in the apical membranes of mucosal cells. Decreased flow of ions and water results in viscid mucus. High concentrations of DNA in airway secretions (attributable to inflammation and lysis of neutrophils)

increase sputum viscosity. The thick mucus causes airway obstruction, atelectasis, and hyperinflation and also decreases ciliary action, thus contributing to mucus stasis, which provides a medium for pulmonary infection. Sweat glands, salivary glands, and lacrimal glands are also affected, leading to high concentrations of sodium and chloride in these secretions.

Clinical Manifestations

Typical findings include a history of cough in a young adult or child; thick, tenacious sputum; recurrent pulmonary infections (commonly *Pseudomonas aeruginosa*); and recurrent episodes of bronchitis. These processes ultimately progress to pneumonia and bronchiectasis, right-sided heart failure (cor pulmonale), and exercise intolerance.

Physical examination may reveal digital clubbing (late sign), dyspnea, tachypnea, sternal retractions, unequal breath sounds, moist basilar crackles and rhonchi, and a barrel chest that is hyperresonant to percussion. Other findings that may be present are pancreatic insufficiency (85% to 90%), cirrhosis of the liver (15% to 20%), diabetes mellitus (8% to 15%), gallstones (30% to 35%), nasal polyps (15%), and failure of development of the vas deferens in males. Infants frequently present with a history of multiple respiratory tract infections, meconium ileus (stool retained in intestine), failure to thrive, jaundice, salt depletion, and edema.

Nutritional assessment reveals depleted fat stores, steatorrhea (fatty stools), anorexia, decreased growth rate in children (weight, height, head circumference), and decreased midarm indices.

Diagnosis

The diagnosis of cystic fibrosis is based on clinical and laboratory findings. Diagnostic studies that are routinely performed include ABG measurements, pulmonary function tests, sputum culture and sensitivity with Gram stain, and chest radiography. Specific diagnostic tests for cystic fibrosis include stool examination for fat, pilocarpine iontophoresis (sweat test), and genetic testing. A 72-hour stool collection combined with the dietary history during that time is used to determine fat absorption and fecal fat excretion. A coefficient of fat absorption of less than 95% (85% in infants) can be used to define steatorrhea (fatty stools). ABG analyses commonly show hypoxemia and hypercapnia because of airway obstruction. Pulmonary function tests reveal decreased vital capacity, decreased airflow rates, increased airway resistance, increased functional residual capacity, and decreased tidal volume. Chest radiographs show evidence of patchy atelectasis, bronchiectasis, obstructive emphysema, cystic lung fields, and peribronchial thickening.

The quantitative pilocarpine iontophoresis sweat test reveals elevated sodium and chloride levels, with more than 98% of patients having levels greater than 60 mEq/L in children and greater than 80 mEq/L in adults. A diagnostic blood test for the genetic marker AF508 may be useful for confirming the diagnosis and providing genetic information to the family.

Treatment

Management of cystic fibrosis involves an interdisciplinary approach. A comprehensive program that focuses on multiorgan derangements is recommended. Because pulmonary disease accounts for the majority of morbidity and mortality associated with cystic fibrosis, treatment is aimed at aggressive pharmacologic management of pulmonary infection. Treatment includes the use of bronchodilators, and mobilization of the thick mucus by postural drainage and chest physiotherapy (percussion and vibration) is a priority. Alternative methods for mucus removal include the forced expiratory technique, which involves coughing (huffing) with an open glottis. Recombinant human deoxyribonuclease I (dornase alfa) acts by digesting extracellular DNA (released from lysed neutrophils) present in the viscid sputum of cystic fibrosis patients to decrease the viscoelasticity of sputum, thus improving pulmonary function and decreasing the risk of infection.

High-dose antibiotic therapy is used for acute exacerbations of respiratory tract infections to decrease bacterial growth in the lungs. An annual influenza vaccine is recommended because of the increased risk of complications associated with infection.

Nutritional therapy includes unrestricted fat consumption (approximately 30% of caloric intake), ingestion of a high-protein diet, and use of vitamin supplements (especially the fat-soluble vitamins A, D, E, and K). Other pharmacologic therapy related to nutrition is aimed at replacement of pancreatic enzymes (pancreatin or pancrelipase). Maintenance of weight in children with cystic fibrosis often requires an intake of 150% of the normal calories recommended for healthy children. In some cases, enteral feedings or intravenous nutrition may be necessary on a short-term basis. Salt supplementation may be necessary in hot weather.

Heart–lung or lung transplantation is currently the only definitive treatment. More than 200 cystic fibrosis patients worldwide have undergone transplantation, with a 3-year survival of 55%. Patients receiving transplants showed marked improvement in mobility, energy, and quality of life.

Identification of the disease-related gene, *CFTR,* has advanced prospects for corrective gene therapy. One limiting factor is that the therapeutic gene has a short-term expression.

Acute Tracheobronchial Obstruction

Etiology

Acute tracheobronchial obstruction requires immediate treatment. Causes frequently include aspiration of a foreign body (e.g., a piece of meat, peanut, coin), malpositioned endotracheal tube, laryngospasm, epiglottitis, trauma, swelling from smoke inhalation, postsurgical blood clot, and compression of the bronchus or trachea by tumors or enlarged lymph nodes. With inhaled foreign bodies, the right side of the lung is affected more often than the left because of the angle of the anatomic extension of the right main bronchus from the trachea.

Pathogenesis

Obstruction by one of the etiologic agents listed earlier can be partial or complete. The health care worker must be prepared to assess the situation rapidly and act immediately to clear the airway.

Clinical Manifestations

With complete obstruction, no air movement will be heard on auscultation, but the patient may still be making inspiratory chest movements. Other clinical features of complete obstruction include inability to talk, tachycardia, cyanosis, and rapid progression to unconsciousness unless the problem is quickly reversed.

With partial obstruction of the airway, the patient usually presents with stridor, sternal and intercostal retractions, wheezing, nasal flaring, tachypnea, dyspnea, tachycardia, and use of accessory muscles to breathe. Cyanosis is a late sign that usually indicates exhaustion or complete obstruction.

Diagnosis

The diagnosis of airway obstruction is based on clinical features and ABG analyses. ABG values frequently show hypoxemia and hypercarbia. Chest radiographs may reveal the location of the obstruction.

Treatment

Treatment involves opening the obstructed airway as quickly as possible. Blows to the patient's back or use of abdominal thrusts (previously called the *Heimlich maneuver*) may be necessary for the foreign body

to be expelled. Aspirated contents occluding the airway are suctioned to relieve obstruction. If these methods are unsuccessful, an emergency tracheostomy should be performed in the case of a suspected upper airway obstruction in the subglottic region or above.

Epiglottitis

Etiology

Epiglottitis is a rapidly progressive cellulitis of the epiglottis and adjacent soft tissues. Acute epiglottitis is suspected when odynophagia (pain with swallowing) seems out of proportion to pharyngeal findings. Inability to swallow saliva with evidence of drooling is common. Epiglottitis is classified as a subtype of croup. The causative organism is primarily *H. influenzae* type B (Hib). It is most often seen in children 2 to 4 years old. Pneumococci, streptococci, and staphylococci are also causal agents. The role of viruses in epiglottitis is unclear.

Pathogenesis

The infecting agent localizes in the supraglottic area in the epiglottis and pharyngeal structures, causing rapid and potentially fatal inflammation with swelling and airway obstruction.

Clinical Manifestations

The patient frequently presents with acute respiratory difficulty that has progressed rapidly over several hours. Common signs and symptoms include drooling, dysphagia, rapid onset of fever, dysphonia, inspiratory stridor, and inspiratory retractions. The child often sits in a "sniffing dog" position, which provides the best airway patency. The oropharynx is edematous and cherry red.

Diagnosis

Definitive diagnosis is obtained by direct or fiberoptic visualization of the epiglottis. Lateral neck radiographs assist in making a definitive diagnosis and reveal a classic "thumbprint sign" (swollen epiglottis that looks like a thumbprint). A complete blood count may reveal leukocytosis with a shift to the left.

Treatment

This condition is a true medical emergency and may necessitate intubation. Antibiotic therapy should be started immediately. Preventive treatment with the Hib vaccine has been the key to decreasing the incidence of this disease.

Croup Syndrome

Etiology

Croup syndrome describes a number of acute viral and inflammatory diseases of the larynx. Croup diseases include laryngotracheobronchitis (viral croup) and bacterial tracheitis. Viral croup affects the larynx, trachea, and bronchi. It is often caused by parainfluenza virus type 1. Other potential infecting organisms include parainfluenza types 2 and 3, RSV, influenza virus, adenovirus, and *Mycoplasma pneumoniae.* Croup usually occurs in the fall and early winter, affecting children ages 6 months to 3 years.

Pathogenesis

The infectious agent causes inflammation along the entire airway, leading to edema formation in the subglottic area.

Clinical Manifestations

The child presents with a history of upper respiratory tract infection or cold that has developed into a barking cough with stridor. Fever is low grade or absent. In severe cases the child may present with stridor at rest, retractions, and cyanosis.

Diagnosis

Diagnosis is based on clinical manifestations and lateral neck films to rule out epiglottitis. Direct laryngoscopy is also used to confirm the presence of epiglottitis because the clinical presentation is similar to that of croup. Lateral neck radiographs show subglottic narrowing and a normal epiglottis. The classic steeple sign associated with viral croup shows narrowing below the vocal cords.

Treatment

Supportive treatment is used for viral croup. Mist therapy, oral hydration, and avoidance of stimulation are used in outpatient therapy. Hospitalized children are managed with oxygen therapy and pulse oximetry. Nebulized epinephrine is effective in relieving airway obstruction. Endotracheal intubation may be required for children with respiratory failure.

KEY POINTS

- Obstructive disorders are associated with increased resistance to airflow, particularly during exhalation.
- Bronchiectasis is associated with recurrent inflammation of the bronchial walls, chronic cough, and aneurysm-like dilatations of the bronchioles. These bronchiolar dilatations serve as pockets of infection, producing purulent, foul-smelling sputum. Treatment centers on use of antibiotic therapy and removal of secretions.
- Bronchiolitis refers to widespread bronchiolar inflammation, often associated with smoking and a number of infectious agents. Inflammation results in mucosal swelling, excessive mucus production, and bronchial muscle constriction—all of which narrow the airway lumen and may lead to wheezing and dyspnea. Treatment centers on administration of bronchodilating agents and management of the underlying cause.
- Cystic fibrosis is an autosomal-recessive disorder of exocrine glands and mucus cells. Secretions are excessively thick because of insufficient chloride and water transport. Thick secretions cause airway obstruction, atelectasis, and air trapping. Associated symptoms resulting from dysfunction of the exocrine pancreas are apparent. Treatment centers on removal of secretions and provision of antibiotic therapy for complicating respiratory tract infections.
- Obstruction of the trachea or large bronchi may occur acutely, requiring immediate treatment. Usual causes include foreign body aspiration, trauma, and inflammation. With complete obstruction, no movement of air occurs, even though inspiratory efforts may be observed. Partial airway obstruction is associated with wheezing, retractions, and stridor. Treatment centers on removing the obstruction, if possible, or creating a patent airway by a tracheostomy.
- Epiglottitis is a medical emergency. *H. influenzae* type B, the primary organism associated with epiglottitis, invades the supraglottic structures (epiglottis and arytenoids), causing inflammation and edema, leading to obstruction. Key points in the clinical diagnosis are rapid onset of fever, pain and difficulty swallowing, and drooling. Lateral neck x-ray films reveal a classic thumbprint sign, which is indicative of epiglottal swelling. Airway maintenance via endotracheal intubation or tracheostomy and antibiotic therapy are the primary treatments. The Hib vaccine has greatly decreased the number of cases seen in the pediatric population.
- Croup is usually from a viral infection of the subglottic area. Children ages 6 months to 3 years present with cough and stridor after an upper respiratory tract infection. Humidification, oxygenation, and inhaled epinephrine are the primary treatment modalities.

SUMMARY

Health care professionals have a key role in the management of respiratory disorders in the hospital and in the community. Obstructive pulmonary diseases presented in this chapter include airway obstruction, obstruction from conditions affecting the tracheobronchial walls, and loss of lung parenchyma (emphysema). Obstructive pulmonary disorders are characterized by increased resistance to airflow. With bronchiectasis, obstruction is due to inflammation, infection, and dilatation of the bronchioles. Similarly, bronchiolitis is associated with inflammation; however, in this situation, inflammation leads to mucosal edema and excessive mucus production. Airway obstruction from cystic fibrosis is related to production of excessive, thick secretions. Obstruction of the airway in croup is the result of edema and increased secretions caused by viral infection. Similarly, epiglottitis is an infectious process requiring emergency treatment. The primary organism causing epiglottitis is *H. influenzae.* The incidence of epiglottitis has been greatly reduced with the advent of the Hib vaccine.

An inflammatory process is also seen in asthma and bronchitis. The inflammation is associated with increased mucus production and edema of the tracheal bronchial mucosa in asthma and bronchitis. Bronchospasm of the tracheobronchial tree attributable to exposure to allergens, pulmonary irritants, stress, and exercise may result in hypoxemia. Obstruction to airflow in emphysema is due to loss of alveoli and small airways. The most common cause is cigarette smoking.

RESOURCES

Asthma and Acute Obstruction

Akinbami LJ, Moorman JE, Bailey C, et al: Trends in asthma prevalence, healthcare use, and mortality in the United States 2001–2010. *National Center for Health Statistics; Centers for Disease Control. (94)* 2012.

Barrett KE, Baman SM, Boitano S, Brooks HL: *Ganong's review of medical physiology*, ed 25, New York, 2015, McGraw-Hill.

Binstadt BA, Schneider L: Allergic disorders and immunodeficiency. In Graef JW, editor: *Manual of pediatric therapeutics*, ed 7, Philadelphia, 2007, Lippincott-Raven, pp 500–516.

Bonini M, Usmani OS: The role of the small airways on the pathophysiology of asthma and chronic obstructive pulmonary disease. *Ther Adv Respir Dis* 9(6):281–293, 2015. doi:10.1177/1753465815588064.

Brooks AM: Asthma. In Garfunkel LC, Kaczorowski J, Christy C, editors: *Mosby's pediatric clinical advisor: instant diagnosis and treatment*, St Louis, 2002, Mosby, pp 171–173.

CDC; Summary Health Statistics for US Adults: National Health Interview Survey, 2012. Series 10 (260) 2014.

Chesnutt MS, Prendergast TJ: Pulmonary Disorders. In Papadakis MA, McPhee SJ, Rabow MW, editors: *Current medical diagnosis & treatment 2016*, New York, NY, 2016, McGraw-Hill.

Covar RA, Fleisher DM, Cho C, Boguniewicz M: Allergic disorders. In Hay WW, et al, editors: *Current diagnosis and treatment: pediatrics*, ed 23, New York, 2016, Lange/McGraw-Hill, pp 1114–1150.

Ferri FF: Asthma. In Ferri FF, editor: *Ferri's clinical advisor: instant diagnosis and treatment*, St Louis, 2016, Mosby, pp 152–161.

Gelb AF, Nadel JA: Understanding the pathophysiology of the asthma-chronic obstructive pulmonary disease overlap syndrome. *J Allergy Clin Immunol* 136(3):553–555, 2015.

Global Initiative for Asthma: Global Strategy for Asthma Management and Prevention, 2017. Available from www.ginasthma.org.

Goroll AH, Mulley AG: Management of asthma. In Goroll AH, Mulley AG, editors: *Primary care medicine: office evaluation and management of the adult patient*, ed 7, Philadelphia, 2014, Lippincott Williams & Wilkins, pp 399–415.

Hogg JC: The pathophysiology of asthma. *Chest* 82:s8–s11, 1982.

Kormis WA: Approach to the patient with acute bronchitis or pneumonia in the ambulatory setting. In Goroll AH, Mulley AG, editors: *Primary care medicine: office evaluation and management of the adult patient*, ed 7, Philadelphia, 2014, Lippincott Williams & Wilkins.

Mainous AG, Hueston WJ: Acute respiratory infections. In Sloane PD, et al, editors: *Essentials of family medicine*, ed 5, Philadelphia, 2011, Lippincott Williams & Wilkins, pp 769–785.

Pollat SM, Kolb A: Asthma in children. In Bope ET, Kellerman RD, editors: *Conn's current therapy, 2016*, ed 1e, Philadelphia, 2016, Saunders, pp 1096–1103.

Schatz M: Asthma in Adolesents and Adults. In Bope ET, Kellerman RD, editors: *Conn's current therapy, 2016*, ed 1e, Philadelphia, 2016, Saunders, pp 369–377.

Vura-Weis DE: Allergies and asthma. In Sloane PD, Slatt LM, et al, editors: *Essentials of family medicine*, ed 5, Philadelphia, 2011, Lippincott Williams & Wilkins, pp 745–768.

West JB: *Pulmonary physiology and pathophysiology: an integrated case-based approach*, ed 2, Philadelphia, 2007, Lippincott Williams & Wilkins.

West JB: *Pulmonary pathophysiology: the essentials*, ed 8, Philadelphia, 2013, Lippincott Williams & Wilkins.

Chronic Obstructive Pulmonary Disease

Federico MJ, Baker CD, Deboer EM, et al: Respiratory tract and mediastinum. In Hay WW, et al, editors: *Current diagnosis and treatment: pediatrics,*, ed 23, New York, 2016, Lange/McGraw-Hill, pp 500–550.

Goroll AH, Mulley AG: Management of chronic obstructive pulmonary disease. In Goroll AH, Mulley AG, editors: *Primary care medicine: office evaluation and management of the adult patient*, ed 7, Philadelphia, 2014, Lippincott Williams & Wilkins, pp 383–398.

Hanania N, Sharafkhaneh A: Chronic obstructive pulmonary disease. In Bope ET, Kellerman RD, editors: *Conn's current therapy*, Philadelphia, 2016, Saunders, pp 385–390.

Harper GM, Johnston CB, Landefeld CS: Geriatric medicine. In Papadakis MA, McPhee SJ, Rabow MW, editors: *Current medical diagnosis & treatment 2016*, New York, NY, 2016, McGraw-Hill.

National Center for Health Statistics: Chronic obstructive pulmonary disease (COPD). Available at: www.cdc.gov/nchs/fastats/copd.htm. (Accessed October 2015).

Weiss EF: Clubbing. In Greene HL, et al, editors: *Clinical medicine*, ed 2, St Louis, 1996, Mosby, pp 563–566.

Cystic Fibrosis and Pulmonary Infections

Behrman RE, Kliegman RM, Jenson HB: *Pocket companion to accompany Nelson textbook of pediatrics*, ed 16, Philadelphia, 2001, Saunders, pp 505–506.

Chen S: Respiratory syncytial virus/bronchiolitis. In Garfunkel LC, Kaczorowski J, Christy C, editors: *Mosby's pediatric clinical advisor: instant diagnosis and treatment*, St Louis, 2007, Mosby.

Cystic Fibrosis Foundation: 6931 Arlington Rd, Suite 2000, Bethesda, MD 20814; 800-344-4823. Available at: www.cff.org.

Ferri FF: Bronchiectasis. In Ferri FF, editor: *Ferri's clinical advisor: instant diagnosis and treatment*, St Louis, 2016, Mosby, pp 253–254.

Fishman DS, Bousuares A: Management of nutritional gastrointestinal and hepatic disorders. In Graef JW, editor: *Manual of pediatric therapeutics*, ed 7, Philadelphia, 2007, Lippincott-Raven, p 359.

Fort GG: Epiglottitis. In Ferri FF, editor: *Ferri's clinical advisor: instant diagnosis and treatment*, St Louis, 2016, Mosby, p 477.

Giosti R: Cystic fibrosis. In Bope ET, Kellerman RD, editors: *Conn's current therapy*, Philadelphia, 2016, pp 392–395.

Jackson MA, Vahle H: Croup. In Garfunkel LC, Kaczorowski J, Christy C, editors: *Mosby's pediatric clinical advisor: instant diagnosis and treatment*, St Louis, 2007, Mosby.

Jaskiewicz J: Cystic fibrosis. In Garfunkel LC, Kaczorowski J, Christy C, editors: *Mosby's pediatric clinical advisor: instant diagnosis and treatment*, St Louis, 2007, Mosby.

Roig IL, Shandera WX: Infectious diseases: viral and rickettsial. In Papadakis MA, McPhee SJ, Rabow MW, editors: *Current medical diagnosis & treatment 2016*, New York, NY, 2016, McGraw-Hill.

Salwen MJ, Siddiqi HA, Gress FG, Bowne WB: Laboratory diagnosis of gastrointestinal and pancreatic disorders. In McPherson RA, editor: *Henry's clinical diagnosis and management by laboratory methods*, ed 22, Philadelphia, 2011, Saunders, pp 312–328.

Voynow JA, Mascarenhas M, Kelly A, Scanlin TF: Cystic fibrosis. In Grippi MA, Elias JA, Fishman JA, et al, editors: *Fishman's pulmonary diseases and disorders*, ed 5, New York, NY, 2015, McGraw-Hill.

Yusen RD, et al: Pulmonary diseases. In Godara H, et al, editors: *The Washington manual of medical therapeutics*, ed 34, Philadelphia, 2013, Lippincott Williams & Wilkins.

Yusen RD, Lefrak SS: Pulmonary II: diseases. In Godara H, et al, editors: *The Washington manual of medical therapeutics*, ed 34, Philadelphia, 2013, Lippincott Williams & Wilkins.

23

Restrictive Pulmonary Disorders

Lorna L. Schumann and Benjamin J. Miller

http://evolve.elsevier.com/Banasik/pathophysiology//

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- How do fibrotic lung disorders develop?
- How is the pathogenesis of acute (adult) respiratory distress syndrome similar to that of infant respiratory distress syndrome?
- How do abnormal accumulations in the pleural space affect lung function?
- What neuromuscular disorders are associated with reduced lung compliance?
- What chest wall deformities are associated with reduced lung compliance?
- What is the pathogenesis of pneumonia?
- What is the pathogenesis of severe acute respiratory syndrome (SARS)?
- What is the pathogenesis of Middle East respiratory syndrome (MERS)?
- What is the pathogenesis of tuberculosis?

CHAPTER OUTLINE

Lung Parenchyma Disorders, 500
Fibrotic Interstitial Lung Diseases, 500
Diffuse Interstitial Lung Disease, 500
Sarcoidosis, 501
Hypersensitivity Pneumonitis, 502
Occupational Lung Diseases, 503
Atelectatic Disorders, 504
Acute (Adult) Respiratory Distress Syndrome, 504
Infant Respiratory Distress Syndrome, 507
Pleural Space Disorders, 508
Pneumothorax, 508
Pleural Effusion, 510
Neuromuscular, Chest Wall, and Obesity Disorders, 511
Neuromuscular Disorders, 511
Poliomyelitis, 511
Amyotrophic Lateral Sclerosis, 511
Muscular Dystrophies, 511
Guillain–Barré Syndrome, 511
Myasthenia Gravis, 511
Chest Wall Deformities, 511
Kyphoscoliosis, 511
Ankylosing Spondylitis, 512
Flail Chest, 513
Disorders of Obesity, 513
Infection or Inflammation of the Lung, 514
Pneumonia, 514
Severe Acute Respiratory Syndrome 516
Middle East Respiratory Syndrome 516
Pulmonary Tuberculosis, 516

Restrictive pulmonary diseases result from decreased expansion of the lungs attributable to alterations in the lung parenchyma, pleura, chest wall, or neuromuscular function. These disorders may be classified as pulmonary or extrapulmonary and represent acute or chronic patterns of lung dysfunction, rather than a single clinical disease. Table 23.1 lists the various disease processes that can be classified as restrictive. These diseases are characterized by a decrease in total lung capacity (TLC), vital capacity (VC), functional residual capacity (FRC), and residual volume (RV). The greater the decrease in lung volume, the greater the severity of the disease. Blood gas analysis often shows decreased arterial partial pressure of oxygen (PaO_2) and normal or decreased arterial partial pressure of carbon dioxide ($PaCO_2$), resulting in increased pH (respiratory alkalosis).

This chapter presents information related to restrictive pulmonary diseases, including lung parenchyma disorders, pleural space disorders, neuromuscular and chest wall disorders, pneumonia, and tuberculosis (TB). Specific lung parenchyma disorders, including interstitial fibrosis, sarcoidosis, hypersensitivity pneumonitis, and pneumoconiosis, as well as atelectatic disorders, including acute (adult) respiratory distress syndrome (ARDS) and infant respiratory distress syndrome (IRDS), are presented. Pleural space disorders, divided into pneumothorax and pleural effusions, are discussed. The section on neuromuscular and

chest wall disorders is divided into neuromuscular weakness, chest wall deformities, and obesity. The final section presents etiologic factors, pathogenesis, clinical manifestations, diagnosis, and management of pneumonia, TB, and severe acute respiratory syndrome (SARS). Table 23.2 describes variations in respiratory parameters that affect restrictive lung disease in infant and elderly populations.

TABLE 23.1 Restrictive Pulmonary Disorders

Disorder Type	Representative Examples
Diseases of the Lung Parenchyma	
Neoplastic disease	—
Pneumonia	Pneumonia (viral, bacterial, fungal), hypersensitivity pneumonitis
Granulomatous disease	Sarcoidosis, tuberculosis, coccidioidomycosis, blastomycosis
Pneumoconioses	Occupational lung disease
Acute interstitial pneumonitis	—
Collagen disease	Rheumatoid arthritis, scleroderma, systemic lupus erythematosus
Atelectasis	—
Pulmonary resection	—
Vascular Diseases	Pulmonary edema, pulmonary embolism
Acute respiratory distress syndrome	—
Diseases of Extrapulmonary Restriction	
Chest wall disease	Kyphoscoliosis, ankylosing spondylitis, obesity
Neuromuscular disease	Quadriplegia, hemiplegia, Guillain–Barré syndrome, myasthenia gravis, amyotrophic lateral sclerosis, muscular dystrophy
Pleural diseases	Pleural effusion, hemothorax, pneumothorax, chylothorax
Other	Abdominal distention, surgery, pregnancy

LUNG PARENCHYMA DISORDERS

Fibrotic Interstitial Lung Diseases

The term *interstitial lung disease* describes a group of more than 180 disorders characterized by acute, subacute, or chronic infiltration of alveolar walls by cells, fluid, and connective tissue. If left untreated, the inflammatory process may progress to irreversible fibrosis. The incidence of interstitial lung disease is 20 cases per 100,000 persons in the general population and 175 per 100,000 in people more than 75 years of age.

Diffuse Interstitial Lung Disease

Etiology. Diffuse interstitial lung disease *(diffuse interstitial pulmonary fibrosis)* is the name typically used for restrictive diseases characterized by thickening of the alveolar interstitium. Synonyms frequently presented in the literature include *interstitial pneumonia, diffuse parenchymal lung disease, Hamman–Rich syndrome, intrinsic fibrosing alveolitis, cryptogenic fibrosing alveolitis,* and *idiopathic pulmonary fibrosis.*

Pathogenesis. Pathogenesis of the disease is not well understood, but is possibly related to an immune reaction that usually begins with injury to the alveolar epithelial or capillary endothelial cells. Pathophysiologic changes may include interstitial and alveolar wall thickening and increased collagen bundles in the interstitium (Fig. 23.1). Lung tissue becomes infiltrated by lymphocytes, macrophages, and plasma cells. Persistent alveolitis may lead to obliteration of alveolar capillaries, reorganization of the lung parenchyma, and irreversible fibrosis. These changes in turn lead to the formation of large air-filled sacs (cysts) accompanied by dilated terminal and respiratory bronchioles. The immune response noted in interstitial lung disease is characterized by three pathologic patterns in the alveoli: inflammation, fibrosis, and destruction.

The inflammatory pattern occurs early and is potentially reversible. The triggering event (occupational exposure, tobacco abuse, drug ingestion, connective tissue disease) causes an inflammatory response leading to increased numbers of inflammatory cells (neutrophils, lymphocytes, macrophages). An associated injury to the alveolar capillary basement membrane from the triggering event leads to increased membrane permeability and movement of fluid and debris into the alveoli. The initial injury, in association with the inflammatory pattern, leads to fibroblastic proliferation and deposition of large amounts of

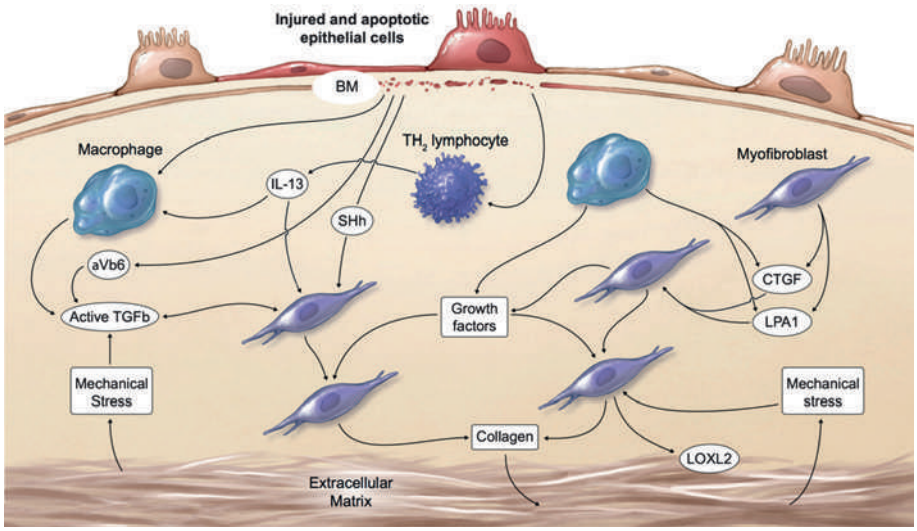


FIG 23.1 Schematic diagram of concepts in the pathogenesis of idiopathic pulmonary fibrosis (IPF) and potential targets to modulate fibroblast proliferation and extracellular matrix deposition in the alveolar wall. *BM,* Basement membrane. (From O'Riordan TG et al: Development of novel agents for idiopathic pulmonary fibrosis. *Chest* 2015;148(4):1083-1092, The American College of Chest Physicians.)

TABLE 23.2 Age-Related Features Contributing to Restrictive Lung Disease

Anatomic Site	Impact on Restrictive Disease
Infant	
Sternum and ribs are cartilaginous with soft chest wall	Diminishes effect of restrictive disease in infants
Ribs are horizontally oriented so that ribs move in and out easily	Diminishes effect of restrictive disease in infants
Accessory muscles of respiration are poorly developed	Majority of respiratory movement relies on diaphragm; restrictive diseases that compromise diaphragmatic excursion affect respiratory status; e.g., thoracic or abdominal surgery, paralysis, and abdominal masses affect diaphragmatic excursion
Diaphragm rests horizontally and draws lower ribs inward in supine position so that diaphragmatic excursion is decreased	Leads to compromised effort of breathing
Cartilage of infant larynx is soft, so airway is compressed when neck is flexed or hyperextended	Increases airway resistance
During first month of life, neonate is obligate nose breather	Nasal obstruction may lead to respiratory distress from decreased airflow
Small diameter of airway leads to increased resistance to airflow	Mucus or edema in the airway may lead to significant increase in resistance and decrease in airway diameter
Fewer alveoli than in adults, leading to decreased radial traction applied to airways	Increased tendency of airways to collapse
Pores of Kohn and channels of Lambert are underdeveloped, leading to fewer collateral ventilation pathways	May lead to respiratory compromise, reducing ventilatory support with restrictive diseases
Elderly	
Decreased ciliary activity	Increased incidence of infection; decreased mucus clearance in all types of respiratory disorders
Decreased chest wall compliance and decreased lung elasticity in some areas of lung	Leads to a reduction in lung volume; leads to decreased expansion of lungs and to decreased matching of ventilation and perfusion
Decreased stress tolerance	Increased incidence of disease and trauma with age
Decreased muscle tone	Decreased physical conditioning
Impaired immunity as evidenced by decreased T-cell function; increased autoantibodies	Decreased resistance to infection
Decreased oxygen uptake	Decreased oxygen level in blood
Decreased vital capacity	Decreased alveolar expansion
Decreased cough reflex	Impaired ability to clear secretions and inhaled particulate matter

collagen. The fibrotic pattern is manifested by increases in the number of mesenchymal cells and fibroblasts in the interstitium, and alveolar walls become thickened with increased amounts of fibrous tissue. Physiologic restriction leads to reduced compliance and increased elastic recoil. The lung destruction pattern is manifested by loss of alveolar walls. Radiographically, this appears as a "honeycomb lung" and indicates end-stage disease. Ground-glass appearance on chest radiograph is often an early finding. The fibrotic and honeycomb patterns respond poorly to treatment.

Clinical manifestations. The most common patient complaint is progressive dyspnea with nonproductive cough. Clinical features also include rapid, shallow breathing; dyspnea; clubbing of the nail beds (40% to 80% cases); bibasilar end-expiratory crackles (Velcro rales); and marked dyspnea with exercise. Cyanosis is a late finding. Anorexia and weight loss are noted on physical examination. While the disease progresses, patients exhibit an inability to increase cardiac output with exercise, as evidenced by low maximal heart rate and high peripheral vascular resistance. Arterial oxygen desaturation occurs with exercise.

Diagnosis. Chest radiographs show a honeycomb appearance and a coarse reticular pattern indicating late stage of disease. Ground-glass haziness is indicative of the presence of infiltrates. High-resolution computed tomography and bronchoalveolar lavage are the primary diagnostic tests used to evaluate interstitial lung disease. Open lung biopsy or transbronchial biopsy and gallium-67 scanning may be used for diagnostic evaluation. Results of pulmonary function tests are usually consistent with restrictive lung disease (decreased VC, reduced TLC, and decreased diffusing capacity).

Treatment. The patient should be encouraged to avoid tobacco use and environmental exposure to offending agents. Primary therapy consists of administration of antiinflammatory and immunosuppressive agents. Immunosuppressive agents have been useful in reducing the dosage of corticosteroids required. Oxygen therapy is needed in patients with hypoxemia. Lung transplantation has been successful in selected patients.

Sarcoidosis

Etiology. Sarcoidosis is categorized as an acute or chronic systemic disease of unknown cause, although an immunologic basis appears likely. A common feature of sarcoidosis is the presence of $CD4^+$ T cells. Activation of the alveolar macrophage from an unknown antigen trigger is a possible cause. The acute process occurs more commonly in women in the second or third decades of life. The chronic form is seen more commonly in the second to fourth decades of life, with the highest incidence seen in North American blacks (35.5/100,000) and Northern European whites (11.9/100,000). Having a first-degree relative with sarcoidosis increases the risk for disease fivefold.

Pathogenesis. The disease is characterized by the development of multiple uniform, noncaseating epithelioid granulomas that affect multiple organ systems, most commonly lymph nodes and lung tissue (90% of cases). Noncaseating granulomas are fibrotic and surrounded by large histiocytes. Sarcoid granulomas may also develop in the bronchial

airways. Abnormal T-cell function is noted with this disease. Other systems/organs frequently involved are the skin, eyes, spleen, liver, kidney, and bone marrow.

Clinical manifestations. Sarcoidosis is characterized by malaise, fatigue, weight loss, fever, chest discomfort, dyspnea of insidious onset, and a dry, nonproductive cough. Other features include erythema nodosum (lesions marked by the formation of painful nodes on the lower extremities); macules, papules, hyperpigmentation, and subcutaneous nodules; hepatosplenomegaly; and lymphadenopathy. Patients with acute disease usually present with enlarged lymph nodes and arthritis, although some patients experience no symptoms. Skin lesions and lacrimal and parotid gland involvement are also noted in the acute process. Iritis, uveitis (65% of patients), blurred vision, conjunctivitis, and ocular discomfort may develop.

Diagnosis. Common laboratory findings in patients with sarcoidosis include leukopenia, anemia, increased eosinophil count, elevated sedimentation rate, and increased calcium levels (seen in 11% of patients). Serum levels of liver enzymes may also be elevated. Approximately 70% of patients exhibit anergy (decreased sensitivity to specific antigens such as *Trichophyton, Candida,* mumps virus, and tuberculin). Patients with active disease also demonstrate elevated levels of angiotensin-converting enzyme (40% to 80% of cases). Chest radiographs can be used to differentiate stages of the disease process: stage 0, normal; stage I, hilar adenopathy alone; stage II, hilar adenopathy and bilateral pulmonary infiltrates; and stage III, pulmonary infiltrates without adenopathy. Stage IV is characterized by advanced fibrosis with evidence of honeycombing, hilar retraction, bullae, cysts, and emphysema. Gallium-67 scans will localize areas of granulomatous infiltrates. Pleural effusion is noted in 10% of cases of sarcoidosis. Pulmonary function test results may be normal or show evidence of restrictive disease and/or obstructive disease. Transbronchial lung biopsy demonstrates noncaseating granulomas, thus providing a definitive diagnosis (75% to 90% cases). Bronchoalveolar lavage may be used to monitor cell content in patients with sarcoidosis. The lavage fluid is characterized by increased lymphocytes and a high CD4/CD8 cell ratio.

Treatment. Administration of corticosteroids and management of symptoms are the mainstays of treatment for patients whose disease process does not resolve spontaneously and in whom progressive lung disease or evidence of extrapulmonary sarcoidosis develops. For patients with progressive disease that does not respond to corticosteroids, immunosuppressive agents may be used. Hydroxychloroquine is effective for treatment of disfiguring skin lesions, hypercalcemia, and neurologic involvement. The prognosis is best for stage I disease. Death attributable to pulmonary insufficiency occurs in about 5% to 7% of patients.

Hypersensitivity Pneumonitis

Etiology. Hypersensitivity pneumonitis, also called *extrinsic allergic alveolitis,* is classified as a restrictive and occupational disease. Numerous (more than 300) inhaled organic agents are responsible for the inflammatory process. Table 23.3 lists various allergens related to the disease. Unlike other pulmonary diseases, hypersensitivity pneumonitis has a predominance in nonsmokers (80% to 95% of cases).

Pathogenesis. The causative agent is suggested by the patient's history and confirmed by demonstration of precipitating antibodies in the serum directed to the causative antigen. The causative antigen combines with the serum antibody in the alveolar walls, leading to a type III hypersensitivity reaction. Type III hypersensitivity diseases are caused by the formation of antigen–antibody complexes (see Chapter 10). These antigen–antibody complexes then elicit granulomatous inflammation that leads to lung tissue injury, as evidenced by thickening of alveolar walls; formation of exudate in the bronchiolar lumen; and infiltration by lymphocytes, plasma cells, and eosinophils. Fibrotic lung changes occur in advanced cases. Many individuals develop precipitating antibodies (precipitin) from organic dust exposure, but only a few develop pneumonitis. Genetic predisposition may be involved in an exaggerated response to the offending agent. Experiments in animals show that a delayed hypersensitivity (type IV) reaction to the antigen is also required before pneumonitis can occur.

TABLE 23.3 Causes of Hypersensitivity Pneumonitis

Disease	Antigen	Allergen Source
Farmer's lung	*Thermophila, Actinomyces*	Moldy hay, silage
Bird fancier's lung	Parakeet, pigeon, chicken	Bird excreta, feathers, and animal protein
Bagassosis	Thermophilic bacteria	Moldy sugarcane pulp
Mushroom, cork, maple bark, or malt hypersensitivity; cheese maker's lung, redwood lung	Various fungi	Handling moldy products
Grain handler's lung	Wheat weevil	Insect-infected grain
Pituitary extract hypersensitivity	Heterologous pituitary and serum proteins	—
Fish-meat worker's lung	Protein, fungi	Animal food
Humidifier lung (fever)	Thermophilic bacteria, amoebae, and fungi	Humidifiers and evaporative air coolers

Clinical manifestations. In the acute stage of the disease, symptoms start 4 to 6 hours after exposure and resolve in 18 to 24 hours. General symptoms may include chills, sweating, shivering, myalgias, nausea, lethargy, headache, and malaise. The patient may have a fever. Respiratory symptoms may include dyspnea at rest, dry cough, tachypnea, and chest discomfort. Physical findings may include cyanosis (a late sign) and crackles (rales) in the lung bases.

In the chronic form, progressive diffuse pulmonary fibrosis develops in the upper lobes—the hallmark of the disease. In the intermediate form, the disease may manifest with acute febrile episodes and progressive pulmonary fibrosis with cough, dyspnea, fatigue, and, eventually, cor pulmonale (right-sided heart enlargement attributable to lung disorders).

Diagnosis. During the acute/subacute phase, transient bilateral pulmonary infiltrates or increased bronchial markings with alveolar nodular infiltrates may be found on chest radiographs. In the chronic phase, diffuse reticulonodular infiltrates and fibrosis are present. Skin testing with the causative antigen may produce a red, indurated, hemorrhagic reaction 4 to 12 hours after injection that lasts several days. This reaction suggests precipitin-mediated sensitivity. Skin testing for most precipitating antigens is impractical because most produce irritating reactions before the precipitin reaction occurs, and many individuals without the disease have precipitating antibodies. Common laboratory findings include an increased white blood cell count and a decreased PaO_2. Elevations in erythrocyte sedimentation rate and the level of C-reactive protein are often present. Hypoxemia worsens with exercise. Pulmonary function tests show decreased lung volumes, diffusing capacity, and static compliance.

Treatment. The goal of therapy is to identify the offending agent and prevent further exposure. This may require a change in environment or occupation. Oral corticosteroids may be used to decrease the inflammatory process.

KEY POINTS

- Diffuse interstitial pulmonary fibrosis is a restrictive disorder characterized by thickening of the alveolar interstitium. The disorder is an immune-mediated disorder that follows an initial lung injury.
- Lung tissues are characteristically infiltrated by immune cells (macrophages and lymphocytes). Excess fibrin deposition results in stiff, noncompliant lungs. VC, tidal volume, functional residual capacity (FRC), and diffusion capacity are generally reduced. Respiratory rate increases to compensate for small tidal volumes.
- Treatment centers on administration of drugs to depress immune system activity, such as corticosteroids.
- Sarcoidosis is a restrictive disorder associated with abnormal protein deposits (granulomas) in the lung. Granulomas are fibrotic and are associated with immune cells (histiocytes). The cause is unknown.
- Symptoms include progressive dyspnea, fever, enlarged lymph nodes, and generalized symptoms of inflammation. Pulmonary lymph nodes may be primarily affected, with progression to parenchymal involvement. Pulmonary function test results are consistent with a restrictive disorder, demonstrating reduced lung volumes and increased respiratory rate.
- Treatment centers on alleviation of the symptoms. Corticosteroids may be used to reduce inflammation.
- Hypersensitivity pneumonitis includes a group of inflammatory lung disorders associated with inhalation of organic particles. Antibodies are produced in response to the inhaled particles; then antigen–antibody complexes deposit in the lung, initiating inflammation and granuloma production.
- Hypersensitivity pneumonitis is characterized by general symptoms of inflammation (e.g., fever, chills, malaise), dyspnea, dry cough, and tachypnea. Chronic exposure leads to progressive fibrosis and pulmonary dysfunction characteristic of restrictive parenchymal disease.

Occupational Lung Diseases

Etiology. Occupational lung diseases result from the inhalation of toxic gases or foreign particles. Traditionally, occupational lung diseases included pathologic conditions that were associated with the effects of exposure to inhaled dusts. However, a holistic approach to these diseases requires consideration of atmospheric pollutants, as well as natural genetic resistance and compliance with health maintenance behaviors. The distinction between occupational and environmental respiratory diseases is becoming increasingly difficult. The integration of multiple environmental areas (home, work, and leisure) further compounds the complexity of defining occupational respiratory diseases. Although atmospheric pollutants (toxic gases) are not discussed in detail here, their impact on occupational respiratory diseases must not be minimized. The sources, potential clinical manifestations, and potential pathologic processes associated with common atmospheric pollutants are presented in Table 23.4.

Pneumoconiosis is defined as parenchymal lung disease caused by the inhalation of inorganic dust particles. The greater the exposure to the dust, the worse the pathologic consequences. Anthracosis (coal miner's lung or black lung), silicosis (silica inhalation), asbestosis (asbestos inhalation), and acute beryllium disease are common examples of occupational lung diseases. However, exposure to several other dusts may also impair respiratory function. Included in this list are antimony ore, barium, iron, tin, fuller's earth (clay), kaolin (china clay), and talc. Hairdressers exposed to bleach and hair spray are at risk for developing obstructive lung disease. Asthma caused by occupational exposures is seen in 16.3% of all adult-onset asthma. Many workers are exposed to "pathogenic dust" through the processing, packaging, or manufacturing of a specific product. Predisposing factors such as history of preexisting lung disease, exposure to atmospheric pollutants, duration of dust exposure, amount of dust concentration, and size of dust particles affect the onset and severity of the respiratory impairment.

Pathogenesis. The respiratory tract is protected by two interrelated systems: the mucociliary system and alveolar macrophages. The inhalation of inorganic particles has little effect on the mucociliary system. However, atmospheric pollutants (sulfur oxides, nitrogen oxides, and tobacco smoke) interfere with and can paralyze ciliary action. As a result, the clearance effect is impaired, and inorganic particles cannot be removed. Alveolar macrophages attempt to engulf and remove inorganic dust by one of the following methods: (1) migrating to small airways to use the mucociliary escalator; (2) engulfing dust and exiting through the lymph and/or blood system; (3) passing through bronchial walls, depositing dust particles in extraalveolar tissue; or (4) destroying the particle (silica).

Macrophage impairment is the primary mechanism through which inorganic particles initiate lung diseases. In an attempt to maintain a sterile alveolar environment, macrophages secrete lysozymes to control foreign particle activity. These enzymes, released in response to the particulate stimuli, eventually damage the alveolar walls, which may cause deposition of fibrous materials. Although the type of inorganic

TABLE 23.4 Common Atmospheric Pollutants Contributing to Lung Disease

Pollutant	Source	Clinical Manifestations	Potential Disease Processes
Carbon monoxide	Automobile exhaust (incomplete fossil fuel combustion)	Lethargy, impairs mental skills, cherry-red mucous membranes, headache	Hypoxemia, respiratory failure
Sulfur oxides	Factories (corrosive, poisonous byproducts of combustion of sulfur-containing fuels)	Inflamed mucous membranes, eyes, upper respiratory tract, bronchial mucosa; cough	Pulmonary edema, bronchitis
Photochemical oxidants (ozone, hydrocarbons, or nitrogen oxides)	Byproduct of exposure of hydrocarbons and/or nitrogen oxides (from fossil fuel combustion with high temperatures) to sunlight	Inflammation of eyes, upper respiratory tract; cough	Tracheitis, bronchitis, pulmonary edema
Cigarette smoke	Cigarettes (carbon monoxide, nicotine, "tars")	Impaired exercise tolerance, decreased mental activity, tachycardia, hypertension, sweating	Bronchial carcinoma, chronic bronchitis, emphysema, coronary heart disease
Particulate matter	Factories/power stations; small particles, visible smoke and soot	Cough; dyspnea; itchy, watery eyes; irritated mucous membranes	Bronchitis, tracheitis, asthma

particle inhaled individualizes the pathophysiologic response, the general response is similar in the context of occupational lung diseases. Silica is one of the most toxic particles to alveolar macrophages. Dense deposits of collagen material are formed around the silica particles, resulting in marked fibrotic tissue deposition and restrictive lung disease. Coal dust and asbestos initiate a similar, although less severe, response. The immune system, via T-cell and IgE mediation, leads to the development of sensitization and inflammation. This damage may lead to occupational asthma. The pathologic processes and clinical features for each of the major occupational lung diseases are summarized in Table 23.5.

Clinical manifestations. Pneumoconioses (anthracosis, asbestosis, silicosis) generally produce no symptoms in the early stages. Physical evidence of the disease occurs when the pulmonary circulation is impaired because of increased pulmonary vascular resistance or development of a pulmonary infection. Workers may remain symptom free for up to 10 to 20 years with chronic exposure. Once again, symptom manifestation is dependent on the predisposing factors. As pneumoconioses progress, patients present with a progressive, productive cough and dyspnea, especially with exercise. In addition, patients may complain of progressive weakness and fatigue. Clubbing of fingers may also be present. Late clinical features include chronic hypoxemia, cor pulmonale, and respiratory failure.

Diagnosis. The reliability of pulmonary changes noted on chest radiographs varies with the severity of the disease. When the patient is symptom free, no changes may be noted. However, as the pneumoconioses progress, micronodular mottling and haziness become apparent. In addition, nodules, fibroses, and calcifications resulting from dust particle deposition are noted. Pneumoconioses usually produce one of three radiographic findings: nodular, reticular, or linear. However, because of the insidious progression of occupational lung diseases, radiographs negative for lung disease do not exclude the presence of the disease process. Changes in pulmonary function tests demonstrate predominantly restrictive impairment (see Fig. 22.19) with a component of obstructive functional impairment, depending on the severity and type of dust inhalation.

Finally, hypoxemia is evident from arterial blood gas measurements in the late disease stages. Falling PaO_2 levels are accompanied by decreased $PaCO_2$ levels as the body initially compensates for the hypoxemia with an increased respiratory drive. However, as the disease progresses, both hypoxemia and hypercapnia are evident.

Treatment. The primary goal of therapy is to halt symptom progression. Preventive measures are the key to limiting the onset and severity of occupational lung diseases. Adherence to federal standards for exposure to dust and particulate matter, as well as continuing education of workers and employers, could dramatically affect the incidence of respiratory diseases. The use of respirators and water sprays for miners to decrease airborne particles in mines are two prevention techniques. Early evaluation of a work environment predisposed to occupational lung diseases is where "treatment" must begin. Two primary goals in the management of active occupational lung diseases are to prevent further parenchymal damage and to relieve signs and symptoms, when possible. Ideally, if the problematic dust can be identified, the individual should be removed from the environment. However, if a job change is unrealistic, every possible measure must be implemented to prevent further inhalation of dust particles. Included in this treatment is the evaluation of current health maintenance behaviors. Treatment consists of corticosteroids, inhaled bronchodilators, oxygen therapy, and respiratory treatments (intermittent positive-pressure ventilation, postural drainage, and deep breathing exercises). The effectiveness and utilization of these therapies depend on the patient's condition and the stage of disease. Rarely are those pathologic conditions reversed with medical treatment.

KEY POINTS

- Occupational lung diseases result from chronic inhalation of gases and inert particles. Commonly identified particles include coal, silica, and asbestos. Smoking and environmental pollutants may be contributing factors because they depress the ciliary function necessary to remove inhaled particles.
- The presence of inert particles in the alveoli initiates macrophage activity and inflammation. Inert particles cannot be digested by phagocytes, so they are walled off by deposition of fibrous proteins.
- Manifestations of pneumoconioses are related to the restrictive nature of pulmonary dysfunction. Progressive dyspnea, decreased vital capacity (VC) and functional residual capacity (FRC), and increased respiratory rate are common. Blood gas analyses show progressive hypoxemia; carbon dioxide levels may remain normal or low until late in the disease.
- Treatment includes prevention of further exposure and administration of corticosteroids, bronchodilators, and oxygen therapy.

ATELECTATIC DISORDERS

Acute (Adult) Respiratory Distress Syndrome

Etiology. ARDS is characterized by damage to the alveolar–capillary membrane. In the United States there are more than 150,000 cases per year. Mortality statistics range from 30% to 50%. Clinically, ARDS is

TABLE 23.5 Major Occupational Lung Pneumoconioses

Pneumoconioses	Pathologic Findings	Clinical Features
Anthracosis (coal miner's lung)	*Early:* Collection of coal particles with small amount of dilation of airway	*Early:* Minimal to no symptoms; may be seen with dyspnea with cough but often due to unrelated bronchitis or emphysema
	Late: Progressive, massive fibrosis with condensed areas of black, fibrous tissue	*Late:* Worsening dyspnea on exertion, productive cough, respiratory failure
Silicosis	Dense collagen deposits in respiratory bronchioles and alveoli and along lymphatics	*Early:* No symptoms noted
		Late: Productive cough, dyspnea, especially with exercise; increased risk for tuberculosis
Asbestosis	Fibrous deposits secondary to long, thin fibers, allowing deep lung penetration	Progressive dyspnea on exertion, weakness, clubbing of fingers; pleural thickening with plaque development

associated with a decline in the PaO_2 that is refractory (does not respond) to supplemental oxygen therapy. Damage to the alveolar–capillary membrane causes widespread protein-rich alveolar infiltrates (visible on chest radiographs) and severe dyspnea. Patients who recover from the acute injury can expect to return to relatively normal lung function. Follow-up studies (9 months to 4 years) in ARDS survivors show a mild restrictive pulmonary function accompanied by cough, dyspnea, and excess sputum production. Some individuals continue to have abnormalities in diffusing capacity, oxygenation, and lung mechanics. ARDS is associated with severe trauma (20% cases), sepsis (more than 40% of cases with 90% mortality), aspiration of gastric acid (more than 30% of cases), fat emboli syndrome, and shock from any cause (Box 23.1). The precise mechanism of lung injury is not known, but the common denominator appears to be increased permeability of the pulmonary vasculature and flooding of the alveoli with proteinaceous fluid, leading to the development of protein-rich pulmonary edema (noncardiogenic pulmonary edema). The acute lung injury triggers the immune system to activate the complement system and to initiate neutrophil sequestration in the lung (Fig. 23.2).

BOX 23.1 Major Disorders Associated With ARDS

Shock (any process leading to a low blood flow state)
- Infectious causes
- Sepsis syndrome (primarily from gram-negative bacteria) with or without sustained hypotension (>40% cases)
- Pneumonia (viral, bacterial, fungal, mycobacterial)
- Miliary tuberculosis
- Bronchiolitis obliterans—organizing pneumonia

Trauma: pulmonary contusion

Embolism
- Fat emboli
- Air emboli
- Thrombus formation
- Amniotic fluid embolism

Head injury (increased intracranial pressure)

Aspiration (>30% of cases)
- Gastric contents
- Drowning (fresh/salt water)

Drug overdose
- Heroin
- Methadone
- Propoxyphene
- Barbiturates, salicylates, thiazides, colchicine

Inhaled toxins
- Smoke inhalation
- High concentrations of oxygen (iatrogenic)
- Corrosive chemicals (ammonia, sulfur dioxide, chlorine, nitrogen dioxide)
- Free-base cocaine smoking

Hematologic disorders
- Disseminated intravascular coagulation
- Massive blood transfusion
- Postcardiopulmonary bypass
- Thrombotic thrombocytopenic purpura

Metabolic disorders
- Pancreatitis
- Uremia
- Paraquat (toxic weed killer) ingestion

Burns

Cancer

Anaphylaxis

Eclampsia

Radiation pneumonitis

High-altitude exposure

Pathogenesis. The pathogenic sequence of events in ARDS is shown in Fig. 23.2. The initial injury to the alveolar–capillary membrane may be caused by direct damage, as seen in aspiration of acidic gastric contents, or by indirect damage, as occurs in shock from any cause. Therapeutic interventions (high oxygen and overhydration) may act to compound the effects of the initial lung injury. The resulting injury leads to an increase in alveolar–capillary permeability, which results in interstitial and alveolar edema. The four characteristic pathophysiologic abnormalities of ARDS involve (1) injury to the alveoli from a wide variety of disorders (Box 23.1), (2) changes in alveolar diameter, (3) injury to the pulmonary circulation, and (4) disruptions in oxygen transport and utilization. Common findings in this type of injury include (1) severe hypoxemia caused by intrapulmonary shunting of blood; (2) a decrease in lung compliance; (3) a decrease in FRC; (4) diffuse, fluffy alveolar infiltrates on the chest radiograph; and (5) noncardiogenic pulmonary edema.

The mechanism by which the FRC is decreased appears to be the result of stiff, noncompliant lungs associated with the presence of alveolar edema and exudate that exaggerate surface tension forces. Early alveolar closure and continued closure lead to atelectasis and loss of lung volume. The decrease in lung compliance, often severe in ARDS, is reflected in the high ventilatory pressures required to deliver an adequate volume of air. It is thought that this decrease in lung compliance is due to loss or inactivation of surfactant, with subsequent increased recoil pressure of the lungs. In addition, proteinaceous fluid fills the alveoli and impairs ventilation. Fig. 23.3 shows alveolar damage attributable to dense proteinaceous debris, desquamated cells, and hyaline membranes. The decrease in PaO_2 is a result of perfusion of large numbers of alveoli that are poorly ventilated (areas of low ventilation–perfusion ratio) or not ventilated (areas of shunt).

Clinical manifestations. The clinical features of ARDS usually include a history of a precipitating event that has led to a low blood volume state ("shock" state) 1 or 2 days before the onset of respiratory failure. The patient may complain of sudden marked respiratory distress. Early signs and symptoms include a slight increase in pulse rate, dyspnea, and a low PaO_2. The initial presenting sign may be shallow, rapid breathing. With progression of the disease, the patient demonstrates tachycardia, tachypnea, hypotension, marked restlessness, decreased mental status, and production of frothy secretions. On auscultation of lung fields, crackles and rhonchi are heard. The patient may be using accessory muscles to breathe and demonstrating intercostal and sternal retractions. A late sign is cyanosis.

Diagnosis. The hallmark of ARDS is hypoxemia that is refractory to increasing levels of supplemental oxygen. Uncorrected hypoxemia is associated with hypotension, decreased urine output, respiratory and metabolic acidosis, and eventual cardiopulmonary arrest. Arterial blood gas determinations reveal hypoxia, acidosis, and hypercapnia. The chest radiograph may initially be normal, but progresses to a bilateral diffuse "whiteout" (Fig. 23.4) indicative of diffuse alveolar infiltrates. The infiltrates characteristically spare the costophrenic angles. Blood and urine cultures will help determine whether infection is the etiology. Pulmonary function tests show a marked decreased in FRC, decreased lung volumes, decreased lung compliance, and a ventilation–perfusion ($\dot{V}_A/\dot{Q}$) mismatch with a large right-to-left shunt. Histologic changes

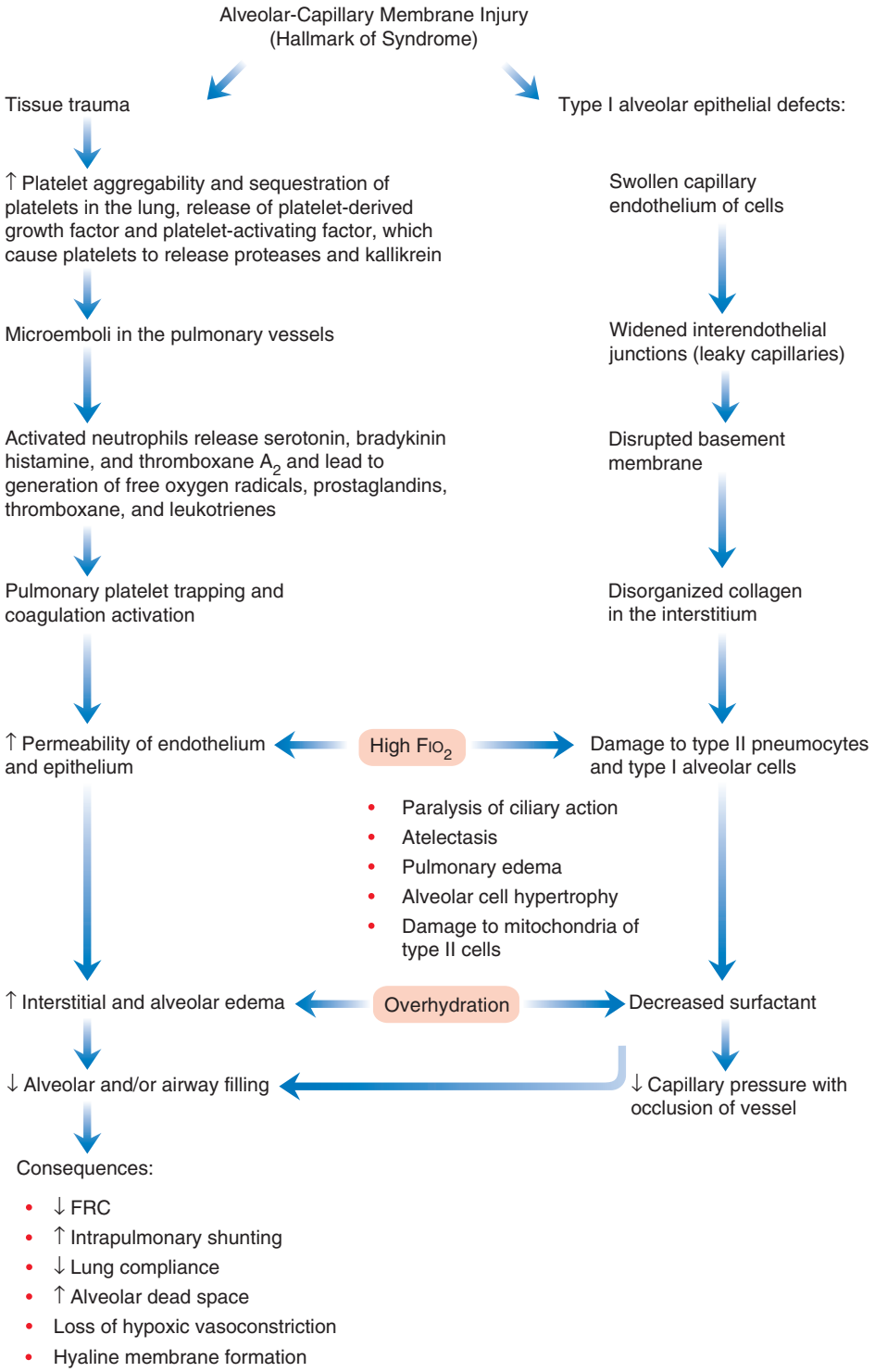


FIG 23.2 Pathogenesis of acute respiratory distress syndrome. *FRC,* Functional residual capacity.

found on open lung biopsy reveal atelectasis, hyaline membranes, cellular debris, and interstitial and alveolar edema (see Fig. 23.3).

Treatment. The management of ARDS entails identifying the underlying cause, addressing the cause (e.g., sepsis), maintaining fluid and electrolyte balance, and providing adequate oxygenation with the use of a volume ventilator utilizing pressure support and positive end-expiratory pressure (PEEP). Patients may require fraction of inspired oxygen (F_IO_2) levels of 1.0 (100%). The goal is to keep the PaO_2 value above 60 mm Hg. Because of increased permeability of the alveolar–capillary membrane, excessive fluid administration can produce or intensify pulmonary edema. High-frequency jet ventilation, inverse ratio ventilation, and inhaled nitric oxide administration have also been used to treat ARDS. ARDS can be prevented experimentally by blocking systemic inflammatory cells. Numerous agents have been investigated but mortality remains high.

KEY POINTS

- Acute respiratory distress syndrome (ARDS) causes profound hypoxemia and a greatly increased work of breathing, often requiring mechanical ventilation and high-level oxygen therapy to maintain the PaO_2 value greater than 60 mm Hg.
- ARDS occurs in association with other pathophysiologic processes, such as trauma, sepsis, or shock. These disorders increase the risk of development of disseminated pulmonary inflammation leading to ARDS. ARDS is associated with a death rate that ranges from 30% to 50%.
- ARDS is a consequence of widespread pulmonary inflammation leading to three major pathophysiologic processes:
  1. Noncardiogenic pulmonary edema associated with "leaky" pulmonary capillaries
  2. Atelectasis associated with lack of surfactant (surfactant normally decreases surface tension in small alveoli and prevents them from collapsing)
  3. Fibrosis (hyaline membranes) associated with inflammatory deposition of proteins
- ARDS is associated with profound alterations in pulmonary function, including decreased VC, decreased FRC, decreased compliance, and decreased tidal volume. Respiratory rate is increased, and symptoms of tissue hypoxia may be apparent.
- Noncardiogenic pulmonary edema is evident as "whiteout" on chest radiograph. Crackles and wheezing may be heard throughout the chest. Profound dyspnea and the use of accessory muscles for breathing are common. Atelectasis and pulmonary edema result in right-to-left pulmonary shunting. Blood gas determinations show hypercarbia and hypoxemia, which do not improve significantly with oxygen therapy.
- Therapy is mostly supportive—to enhance tissue oxygenation until the inflammatory process resolves. Mechanical ventilation with positive end-expiratory pressure (PEEP) and supplemental oxygen is the mainstay of therapy. PEEP is used to increase FRC and prevent alveolar collapse at end expiration. PEEP may also force edema fluid out of the alveoli. High levels of oxygen (>60%) may contribute to ARDS because of absorption atelectasis. The F_IO_2 value should be reduced as soon as possible.

Infant Respiratory Distress Syndrome

Etiology. IRDS, also known as *hyaline membrane disease,* has features similar to those of ARDS. It is a syndrome of premature neonates, characterized by hemorrhagic pulmonary edema, patchy atelectasis, and hyaline (glassy) membranes. Hypoxemia that is refractory to increasing levels of oxygen supplementation is the hallmark of the syndrome. The incidence is 60% in infants less than 30 weeks' gestation without antenatal steroids and 35% in those receiving antenatal steroids. The incidence in

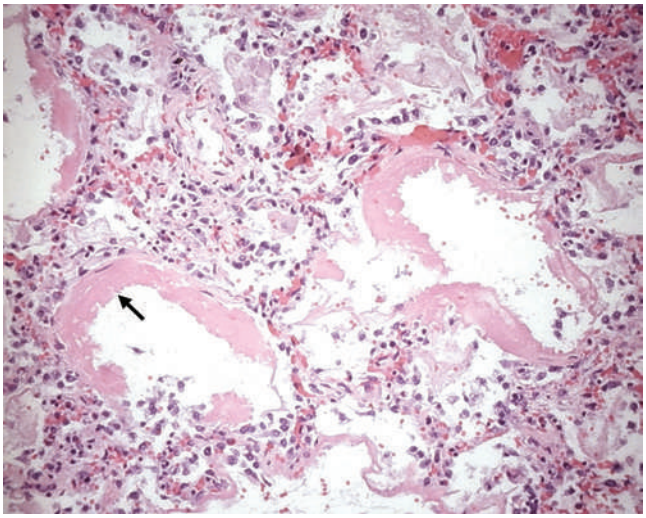


FIG 23.3 Diffuse alveolar damage (acute respiratory distress syndrome) shown in photomicrograph. Some of the alveoli are collapsed; others are distended. Many contain dense proteinaceous debris, desquamated cells, and hyaline membranes *(arrow).* (From Ichikado K: High-resolution computed tomography findings of acute respiratory distress syndrome, acute interstitial pneumonia, and acute exacerbation of idiopathic pulmonary fibrosis. *Semin Ultrasound CT MR* 2014;35(1):39–46.)

FIG 23.4 Chest radiograph of a 28-year-old man who was involved in an automobile accident. The patient presented with multiple bilateral rib fractures and bilateral pneumothorax. Within 24 hours, severe acute respiratory distress syndrome developed (note diffuse "whiteout").

infants older than 34 weeks is 5%. High-risk factors include birth earlier than 25 weeks' gestation; birth at advanced gestational age; poorly controlled diabetes in the mother; deliveries after antepartum hemorrhage; cesarean section without antecedent labor; or the presence of perinatal asphyxia, multiple births, previous infant with respiratory distress syndrome, and Rh factor incompatibility. The increased risk of respiratory morbidity may be due to lack of hormones associated with labor. During normal labor there is a decrease in secretion of fetal lung liquid and an increase in absorption of lung liquid. Labor also stimulates the release of surfactant. During normal labor there is an increase in catecholamine release. Loss of labor-induced catecholamine release may put the neonate at increased risk of respiratory morbidity.

Pathogenesis. The primary cause of IRDS is a lack of pulmonary surfactant, leading to increased alveolar surface tension and decreased lung compliance. Surfactant, a phospholipid, is produced by type II alveolar cells in increasing quantities after 32 weeks' gestation. With IRDS, lung compliance is decreased to one-fifth to one-tenth of normal. The neonate with IRDS must generate high intrathoracic pressures (25 to 30 mm Hg) to maintain patent alveoli. The premature neonate has a soft, compliant chest that is drawn inward with each inspiratory contraction of the diaphragm, making it difficult to maintain the high pressures needed to ensure adequate oxygenation. The end result from increased work of breathing and decreased ventilation is progressive atelectasis, increased pulmonary vascular resistance, profound hypoxemia, and acidosis. Surfactant also functions to maintain pulmonary fluid balance. Alteration of surface tension forces, normally maintained by surfactant, causes further leakage of proteinaceous fluid into the alveoli. This fluid contains fibrin and cellular debris, which causes hyaline membrane formation. Surfactant normally decreases surface tension in the alveolus during expiration, allowing the alveolus to remain partially open, thus maintaining FRC. A secondary cause of IRDS is immaturity of the capillary blood supply, which leads to $\dot{V}_A/\dot{Q}$ mismatch, thus adding to the problems of hypoxemia and metabolic acidosis. In addition, a right-to-left shunt from an open foramen ovale or patent ductus arteriosus may increase the hypoxemia. Histologically, there is progressive damage to the basement membrane and respiratory epithelial cells. With increasing edema and loss of epithelial cells, patchy areas of atelectasis develop. Cellular damage from the disease process, excess fluid administration, and high values of F_IO_2 lead to increased capillary permeability and leakage of high-protein fluid into the alveoli.

Clinical manifestations. The typical neonate presents with shallow respirations; intercostal, subcostal, or sternal retractions; diminished breath sounds; flaring of nares; hypotension; peripheral edema; low body temperature; oliguria; tachypnea (60 to 120 breaths/min); and bradycardia. Late findings include frothy sputum, central cyanosis, and an expiratory grunting sound. Nasal flaring is a physiologic response mechanism used to increase airway diameter in an attempt to overcome airway resistance. An expiratory grunt is a physiologic response mechanism reflecting an attempt to create a physiologic PEEP by exhaling against a partially closed glottis. Paradoxical respirations ("see-saw" movement of the chest wall) may also be noted, indicating increased work of breathing. During the first 48 to 72 hours of life, neonates with IRDS need progressively higher levels of F_IO_2 to maintain adequate (50 to 70 mm Hg) oxygen levels. Inspired oxygen should be adjusted to maintain a saturation of 85% to 94%. As work of breathing increases and oxygen levels decrease, metabolic acidosis may occur.

Diagnosis. Initial arterial blood gas determinations reveal hypoxemia and metabolic acidosis attributable to lactic acid formation by hypoxic tissues. As the disease progresses, hypercapnia and respiratory acidosis develop. Chest radiographs progress from normal, shortly after birth, to a diffuse whiteout or ground-glass appearance indicative of diffuse bilateral atelectasis and alveolar edema. Generalized hypoinflation of the lungs is also seen on chest x-ray. Measurement of the lecithin/sphingomyelin (L/S) ratio and the desaturated phosphatidylcholine concentration in amniotic fluid may be done to determine the ability of the fetus to secrete surfactant. An L/S ratio of ≥2:1 (3:1 in mothers with diabetes) is a reliable indicator of mature lungs. The presence of phosphatidylglycerol in the amniotic fluid is indicative of pulmonary maturity. Administration of glucocorticoids before delivery may stimulate lung maturation and improve the L/S ratio.

Treatment. Prevention of IRDS is aimed at the use of antenatal steroids that significantly increase respiratory compliance (about 23%) and decrease the risk of development of IRDS. The mainstay of therapy is mechanical ventilation with PEEP or continuous positive-airway pressure. Prevention is a primary goal. The therapeutic goal is to maintain adequate oxygen levels between 50 and 70 mm Hg. The lowest F_IO_2 settings should be used to maintain adequate arterial oxygen levels. High F_IO_2 (100%) delivered for extended periods may result in further alveolar damage, primary persistent pulmonary hypertension, and retrolental fibroplasia (failure of the peripheral retina to vascularize, leading to blindness). Exogenous surfactant (bovine, porcine, or synthetic) administration to premature infants has decreased the death rate in IRDS by 50%. Surfactant decreases surface tension, thereby reducing the amount of pressure required to open the alveoli. High-frequency ventilation has proved to be effective in infants with severe IRDS by providing more uniform lung inflation, improving lung mechanics, and enhancing gas exchange. Infants receiving high-frequency ventilation require lower mean airway pressures and have better gas exchange than those ventilated conventionally. General supportive therapy of adequate intravenous nutrition, fluid and electrolyte balance, minimal handling, and a neutral thermal environment should be maintained. Broad-spectrum antibiotics are prescribed for infections after cultures have been done or prophylactically until blood cultures prove negative.

KEY POINTS

- The symptoms of infant respiratory distress syndrome (IRDS) are similar to those of acute (adult) respiratory distress syndrome (ARDS). IRDS occurs most commonly in premature infants born before adequate development of their surfactant-producing pneumocytes (25 weeks' gestation). The maturity of surfactant-producing cells can be estimated from the L/S ratio in amniotic fluid. An L/S ratio of less than 2:1 is associated with a higher risk of IRDS.
- Lack of surfactant causes atelectasis and increased work of breathing as a result of high alveolar surface tension. Leakage of inflammatory exudate into the alveoli results in formation of hyaline membranes.
- Symptoms of IRDS include nasal flaring, expiratory grunt, thoracic retractions, and rapid, shallow respirations. Chest radiographs demonstrate a "whiteout." As in ARDS, blood gas values are poor, indicating severe hypoxemia and acidosis.
- Therapy for IRDS includes supportive measures, such as mechanical ventilation with PEEP or continuous positive-airway pressure, and use of supplemental oxygen, as well as specific measures to increase alveolar surfactant levels.

PLEURAL SPACE DISORDERS

Pneumothorax

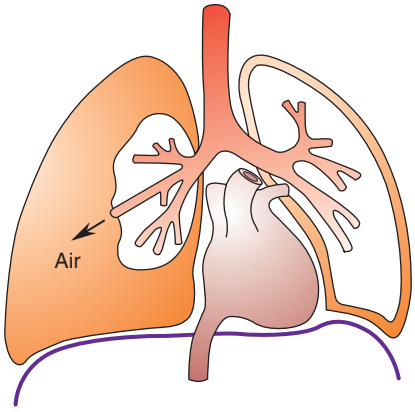
Etiology. Spontaneous pneumothorax (SP) is characterized by the accumulation of air in the pleural space. A primary pneumothorax is classified as spontaneous, occurring mainly in tall, thin men between ages 20 and 40 years without underlying disease factors. Cigarette smoking increases the risk of SP. A secondary pneumothorax occurs as a result of complications from preexisting pulmonary disease (such as asthma, emphysema, cystic fibrosis, infectious disease [pneumonia or TB], or interstitial lung disease). In the United States there are approximately

20,000 new cases annually. SP is six times more common in men than in women. A specific category of secondary pneumothorax associated with menstruation is called *catamenial pneumothorax* (pathogenesis unknown). A catamenial pneumothorax occurs primarily in the right hemothorax and is associated with endometriosis. A third classification (tension pneumothorax) is traumatic in origin, resulting from penetrating or nonpenetrating injury. A tension pneumothorax is a medical emergency. Other examples of traumatic pneumothorax have iatrogenic causes, such as placement of central lines, thoracentesis (6%), percutaneous lung biopsy, and mechanical ventilation.

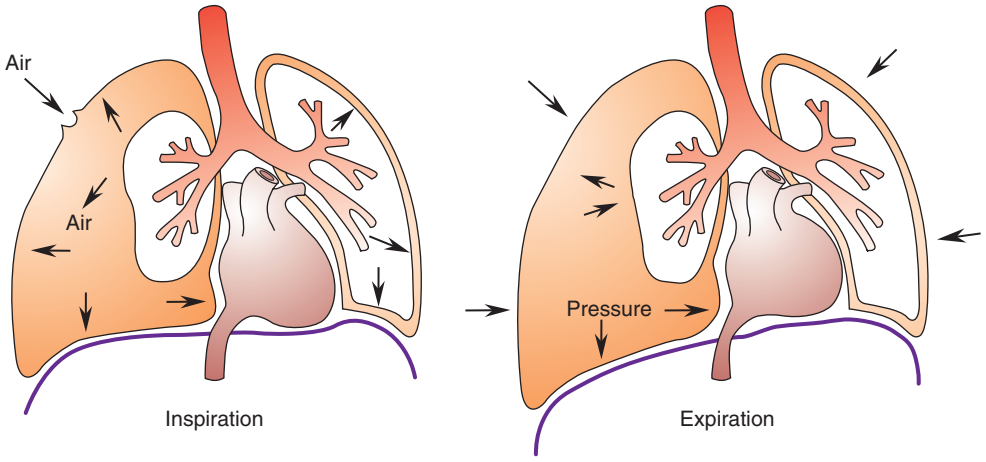
Pathogenesis. Primary SP (Fig. 23.5) results from rupture of small subpleural blebs in the apices. When air enters the pleural space, the lung collapses and the ribcage springs out. The subpleural blebs are believed to occur in the apices as a result of negative mechanical pressures in the upper third of the upright lung field. Secondary pneumothorax occurs as a result of complications from an underlying lung problem and may be due to rupture of a cyst or bleb. Tension pneumothorax (see Fig. 23.5) results from the buildup of air under pressure in the pleural space. Air enters the pleural space during inspiration but cannot escape during expiration. The lung on the ipsilateral (same) side collapses and forces the mediastinum toward the contralateral (opposite) side, thus decreasing venous return and cardiac output (see Fig. 23.5). With an open, "sucking" chest wall wound, air enters during inspiration, but cannot escape during expiration, leading to a shift of the mediastinum (contents of the septum between the two lungs) and the trachea.

Clinical manifestations. The clinical features of pneumothorax include tachycardia, decreased or absent breath sounds on the affected side, hyperresonance, sudden chest pain on the affected side (90%), and dyspnea (80%), Small pneumothoraces (less than 20%) are usually not detectable on physical examination. Tension pneumothorax and a large SP are emergency situations, with patients presenting with severe tachycardia, hypotension, a tracheal shift to the contralateral side, neck vein distention, hyperresonance, and subcutaneous emphysema. Fig. 23.6 shows severe subcutaneous emphysema (air in the tissues attributable to tracheobronchial rupture).

Diagnosis. The typical chest radiograph shows depression of the hemidiaphragm on the side of the pneumothorax and a pleural line with absence of vessel markings peripheral to this line. Chest radiography should be done with the patient standing. Expiratory films show a better demarcation of the pleural line than inspiratory films. Diagnosis may be based on clinical features without radiographic confirmation. However, a chest radiograph is usually obtained. A chest radiograph in tension pneumothorax shows a mediastinal shift (Fig. 23.7). The electrocardiogram may show axis deviations, nonspecific ST-segment changes, and T wave inversion. Arterial blood gas analysis shows a decreased PaO_2 and acute respiratory alkalosis.



SPONTANEOUS PNEUMOTHORAX



TENSION PNEUMOTHORAX

FIG 23.5 *Top,* Spontaneous pneumothorax. *Bottom,* Tension pneumothorax: air builds up under pressure, leading to collapse of the ipsilateral lung and shift of the mediastinum to the contralateral side.

FIG 23.6 Subcutaneous emphysema (air in the tissues) from tracheobronchial disruption and injuries of the esophagus. (From Kirsh MM, Sloan H: *Blunt chest trauma: general principles and management,* Boston, 1977, Little, Brown, p 109.)

FIG 23.7 Upright posteroanterior chest radiograph showing a right-sided tension pneumothorax. Note marked deviation of trachea and cardiac silhouette into the left side of the chest. There is also depression of the right hemidiaphragm. (From Kirsh MM, Sloan H: *Blunt chest trauma: general principles and management,* Boston, 1977, Little, Brown, p 62.)

Treatment. The management of pneumothorax depends on the severity of the problem and the cause of the air leak. If the lung collapse is less than 15% to 25%, the patient may or may not be hospitalized. Typically, the patient is treated symptomatically. A nonhospitalized patient should be monitored closely. The expectation is for complete resolution within several weeks. If the collapse is greater than 15% to 25%, chest tube placement with water seal and suction is recommended. Also, 100% oxygen should be administered to reduce the partial pressure of nitrogen in pleural capillaries, thus quadrupling the rate of pneumothorax absorption. Chemical pleurodesis may be indicated for patients with recurrent SP to promote adhesion of the visceral pleura to the parietal pleura to prevent further ruptures. Patients with a previous pneumothorax should be warned about a possible recurrence at high altitudes, from scuba diving, and from smoking. A thoracotomy may be performed on patients in whom further SP and blebs develop. Surgery permits stapling or laser pleurodesis of ruptured blebs. Smoking cessation should be advised. Approximately 25% to 50% of patients with primary pneumothorax will have a recurrence within 1 year.

Pleural Effusion

Etiology. Pleural effusion is not a disease, but a pathologic collection of fluid or pus in the pleural cavity as a result of another disease process. Normally, about 17 mL of serous fluid is contained in the pleural space. There is a constant movement of pleural fluid from parietal pleural capillaries to the pleural space, which is then reabsorbed into the parietal pleural lymphatics. The pleural membrane is a porous, mesenchymal, serous membrane that allows for movement of interstitial fluid. The fluid has a mucoid characteristic allowing for easy movement of the lungs. The five major types of pleural effusion are: (1) transudates, (2) exudates, (3) empyema attributable to infection in the pleural space, (4) hemothorax or hemorrhagic pleural effusions, and (5) chylothorax or lymphatic pleural effusions.

Transudates have low concentrations of both protein (ratio of pleural fluid protein to serum protein is less than 0.5) and lactate dehydrogenase (LDH; pleural fluid LDH/serum LDH ratio less than 0.6) and have a specific gravity below 1.016. Transudates are frequently associated with severe heart failure or other edematous states, such as cirrhosis with ascites, nephrotic syndrome, and myxedema.

Exudates have high concentrations of both protein (>0.5 mg/dL) and LDH (pleural fluid LDH/serum LDH ratio >0.6). Common causes of exudates are malignancies, infections (especially pneumonia), pulmonary embolism, sarcoidosis, post–myocardial infarction syndrome, and pancreatic disease. Empyema is a high-protein exudative effusion resulting from infection in the pleural space. Hemothorax (the presence of blood in the pleural space) is often the result of chest trauma. Hemorrhagic pleural effusion contains a mixture of blood and pleural fluid. If the hematocrit of the fluid is greater than 50% of the hematocrit of peripheral blood, the fluid collection is called a *hemothorax.* Chylothorax or chylous pleural effusion is an exudative process that develops from trauma as a result of leakage of chyle (lymph fluid) from the thoracic duct or from rheumatoid pleural effusion or tuberculous pleuritis.

Pathogenesis. Pathophysiologic changes associated with the various types of effusions relate to changes in pleural capillary hydrostatic pressure, colloid oncotic pressure, or intrapleural pressure. Transudates can be caused by increased hydrostatic or decreased oncotic pressure. Exudates are associated with increased production of fluid as a result of increased permeability of the pleural membrane (inflammation) or impaired lymphatic drainage. The imbalance in these pressures is associated with fluid formation exceeding fluid removal.

Clinical manifestations. Clinical features vary depending on the cause and size of the effusion. Small effusions may be asymptomatic

(which is common) in patients with less than 290 mL of fluid in the pleural cavity. General features include dyspnea, pleuritic pain that is sharp and worsens with inspiration, dry cough, decreased chest wall movement, absence of breath sounds, dullness to percussion, and decreased tactile fremitus over the affected area. The most useful physical findings are dullness to percussion and tactile fremitus. A massive pleural effusion may lead to a contralateral tracheal shift.

Diagnosis. Thoracentesis should be done to analyze the fluid and to reduce the amount in the pleural cavity. Evaluation of the pleural fluid is done to determine its characteristics, which acts as an additional indicator of its origin. Pleural fluid should be analyzed for chemistry (pH, LDH, and glucose) and presence of pathogenic bacteria. Chest radiography should be done to detect pleural-based densities, infiltrates, signs of congestive heart failure, hilar adenopathy, and loculation of fluid. Once 200 mL of fluid forms, the effusion is visible radiographically. Computed tomography and ultrasonographic tests assist in the diagnosis of complicated effusions and distinguish a mass from a large effusion. Ultrasonography is also useful for thoracentesis guidance.

Treatment. Treatment is directed at the underlying cause of the effusion and relief of symptoms. Closed chest tube drainage in adults or thoracentesis is indicated if the effusion is large. Closed chest drainage in pediatric cases is controversial. A thoracotomy to control bleeding may be required in patients with excessive bleeding (more than 200 mL/hr).

KEY POINTS

- The pleural space is usually a potential space, containing only a small amount of fluid for lubrication. Accumulations of air (pneumothorax), pus (empyema), blood (hemothorax), lymph (chylothorax), or transudate in the pleural space can restrict lung expansion.
- Tension pneumothorax occurs when pleural air progressively accumulates and develops a positive pressure in the pleura. The ipsilateral (same side) lung collapses, and mediastinal structures (trachea, heart) are shifted to the opposite side. Breath sounds are diminished or absent on the affected side.
- Tension pneumothorax and a large, simple (spontaneous) pneumothorax are medical emergencies requiring treatment to remove pleural air and reexpand the lung. This usually requires insertion of a chest tube. Chemical pleurodesis may be done in persons prone to spontaneous pneumothorax (SP).
- A number of disease processes may result in accumulation of fluid in the pleural space. Analysis of the type of fluid (e.g., transudates, exudates, blood, pus) indicates the underlying disease process. General manifestations include dyspnea, cough, pleuritic pain, and diminished breath sounds and dullness to percussion over the effusion.

NEUROMUSCULAR, CHEST WALL, AND OBESITY DISORDERS

Neuromuscular Disorders

Diseases affecting the muscles of respiration or their nerve supply can lead to dyspnea and respiratory failure. Table 23.6 summarizes the features of these disorders.

Poliomyelitis

Poliomyelitis is a viral disease in which the poliovirus attacks motor nerve cells of the spinal cord and brainstem. The incidence of poliomyelitis in the United States is approximately eight cases per year. All of these cases have been related to the oral polio vaccine. The diaphragm and intercostal muscles can be affected, resulting in weakness or paralysis and respiratory failure. At least 95% of infections are asymptomatic. Patients with minor symptoms present with fever, headache, vomiting, diarrhea, constipation, and sore throat. Respiratory muscle function generally recovers, although occasionally patients have chronic respiratory insufficiency from previous disease. As the result of mass vaccination of populations, new cases are quite rare and usually occur in unvaccinated immigrants. Currently, only Pakistan and Afghanistan have new cases of polio.

Amyotrophic Lateral Sclerosis

Amyotrophic lateral sclerosis (ALS) is a degenerative disease of the nervous system that involves both upper and lower motor neurons. ALS occurs in males more than in females (2:1) and has a prevalence of 0.5 to 2 cases in 100,000 persons. Onset is often between the ages of 50 and 70 years. Only 5% of ALS cases are familial. Commonly, muscles innervated by both spinal nerves and cranial nerves are affected. Clinically, progressive muscle weakness and wasting develop, eventually leading to profound weakness of respiratory muscles and death. Although the course of the disease is variable from patient to patient, the natural history is one of irreversibility and progressive deterioration (see Chapter 45).

Muscular Dystrophies

Duchenne muscular dystrophy is a hereditary disease, passed from mothers to sons (X-linked recessive) and occurs in 1 per 3500 male births. The disease is characterized by progressive muscular weakness, initially in the lower extremities, and wasting. In later years (twenties to thirties), respiratory muscles become involved, leading to hypoxia, hypercapnia, and frequent respiratory tract infections (see Chapter 51).

Guillain–Barré Syndrome

Guillain–Barré syndrome, also called *acute polyneuritis,* is a disorder that is presumed to have an immunologic basis and occurs in 0.6 to 1.9 cases per 100,000 persons. Infection involving *Campylobacter jejuni* often precedes the diagnosis. Guillain–Barré syndrome is characterized by demyelination of peripheral nerves. Frequently, patients have a history of recent viral or bacterial illness followed by development of ascending paralysis. Clinically, weakness and paralysis begin symmetrically in the lower extremities and progress or ascend proximally to the upper extremities and trunk. In severe cases, respiratory muscle weakness accompanies limb and trunk symptoms. Generally, the natural history of the disease leads to full recovery in 62% of patients. Mortality is about 5% to 10% worldwide (see Chapter 45).

Myasthenia Gravis

Patients with myasthenia gravis experience weakness and fatigue of voluntary muscles, most frequently those innervated by cranial nerves, but peripheral and respiratory muscles can also be affected. The hallmark of the disorder is weakness made worse by exercise and improved by rest. The incidence in the United States is 10 to 20 cases per year per 1 million persons. Females are more affected than males (3:2). The primary abnormality is found at the neuromuscular junction, where transmission of impulses from nerve to muscle is impaired by a decreased number of receptors on the muscle. Although myasthenia gravis is a chronic illness, the manifestations can often be managed by appropriate therapy, and individual episodes of respiratory failure are potentially reversible. Respiratory failure in this disorder can be due to increasing severity of illness or overmedication (see Chapter 51).

Chest Wall Deformities

Kyphoscoliosis

Etiology. Kyphoscoliosis may develop from an unknown cause (80% to 90% cases) (idiopathic) or may be related to congenital (Pott disease) or neuromuscular disease (muscular dystrophy, Marfan syndrome,

TABLE 23.6 Neuromuscular Disorders Affecting the Respiratory System

Disease	Etiology	Pathophysiology	Clinical Features
Poliomyelitis (myelitis is inflammation of spinal cord)	Develops from an enteral virus acquired by ingestion or respiratory droplet	After a 1- to 3-week period, virus invades intestinal blood supply; once in circulation, virus invades all areas of body; invasion of central nervous system leads to neural damage and initiation of an inflammatory reaction	General symptoms are tremors, muscle weakness; bulbar poliomyelitis affects respiratory muscle nerves, leading to respiratory paralysis; patients usually exhibit shoulder girdle paralysis first, followed by intercostal and diaphragm muscle paralysis; paralysis may be rapid or slowly progressive; also seen are diplopia, facial weakness, dysphagia, difficulty chewing, nasal voice, and loss of gag reflex
Amyotrophic lateral sclerosis	Cause unknown; current theories include autoimmune disease and a slow virus	Affects anterior horn cells of both upper and lower motor neurons	Progressive weakness affecting distal more than proximal muscles; atrophy, fasciculations, and spasticity are noted; involvement of respiratory muscles leads to respiratory dysfunction requiring mechanical ventilation
Muscular dystrophies (most common is Duchenne type)	Hereditary disease (X-linked recessive) passed from mothers to sons	Progressive muscular weakness noted initially in lower extremity muscles; in later years (twenties and thirties) respiratory muscles become involved; patients are at risk for respiratory tract infections	Progressive muscular weakness and wasting; skeletal deformities are also common; involvement of respiratory muscles (diaphragm, intercostals, and accessory muscles) leads to hypoxia and hypercapnia
Guillain–Barré syndrome (acute idiopathic polyneuropathy)	Exact cause unknown, but thought to be an autoimmune disease triggered by a viral infection	Disease usually follows an infection or vaccination; peripheral nerves are affected, leading to neural inflammation, demyelination, and axon destruction	Progressive weakness and loss of motor function beginning in feet and legs and ascending upward; sensory loss may also be noted but is not as dramatic as motor loss; loss of respiratory muscle control leads to respiratory failure, which frequently requires mechanical ventilation; autonomic nervous system symptoms may also be noted (tachycardia, dysrhythmias, hypotension or hypertension, loss of ability to sweat)
Myasthenia gravis	Considered an autoimmune disease with both humoral (B cell) and cell-mediated (T cell) components	Autoantibodies and T cells bind to and damage acetylcholine receptors, leading to decreased functioning of receptors	Common symptoms are diplopia, ptosis, difficulty swallowing, increased weakness with activity, nasal voice, slurred speech, and weakness of proximal extremities; as disease progresses, respiratory muscles become involved, leading to respiratory failure; pneumonia may result from respiratory failure and immobility

neurofibromatosis, Friedreich ataxia, or poliomyelitis). Most idiopathic cases of scoliosis are found in adolescents (11 years or older). The female-to-male ratio is 4 to 5:1.

Pathogenesis. Commonly, a bony deformity of the chest wall occurs as a result of kyphosis (hunchback appearance; posterior curvature deformity) and scoliosis (lateral curvature deformity) (Fig. 23.8). The higher the deformity in the vertebral column, the greater the compromise of respiratory function. Lung volumes are compressed, leading to atelectasis, $\dot{V}_A/\dot{Q}$ mismatch, and hypoxemia.

Clinical manifestations. Common clinical features include dyspnea on exertion; rapid, shallow breathing; and chest wall deformity as evidenced by ribs protruding backward, flaring on the convex side, and being crowded on the concave side. Hypoxemia develops later, and eventually carbon dioxide retention occurs.

Diagnosis. Diagnostic findings include hypercapnia, hypoxemia (due to $\dot{V}_A/\dot{Q}$ mismatch), and decreased lung volumes and lung capacities as evidenced by decreased values on pulmonary function tests. Also noted are increased pulmonary arterial pressures because of the associated pulmonary hypoxemia. Radiographs show accentuated bony curves. Screening for scoliosis and kyphoscoliosis in school-aged children has proved to be an excellent method of early diagnosis of these conditions.

Treatment. Treatment depends on the severity of the deformity and the age of the patient. Kyphosis in elderly persons, especially women, is commonly because of osteoporosis. Curvatures of less than 20 degrees should be monitored on a regular basis. A postural exercise program for mild scoliosis and external braces for moderate scoliosis are recommended. For more advanced cases with curvatures greater than 40 degrees, electrical stimulation of the paraspinal muscles, spinal fusion, and spinal instrumentation (Harrington rod) placement for surgical stabilization are recommended treatments. Curvatures of greater than 60 degrees correlate with poor pulmonary function in later life.

Ankylosing Spondylitis

Etiology. Ankylosing spondylitis occurs in both genders (male/female ratio of 3:1) and is commonly seen in the age range of 15 to 35 years of life. It is characterized by chronic inflammation at the site of ligamentous insertion into the spine or sacroiliac joints. The precise cause is unknown; however, infectious sources are seen in some cases. Ninety percent of patients with the disease have a positive HLA-B27 antigen (Chapter 10). The respiratory system is affected by limited chest expansion and by the formation of pulmonary fibrosis in the upper lobes, which later develops into bronchiectasis and cavitation. Transient acute arthritis of the peripheral joints occurs in about 50% of cases.

Pathogenesis. Ankylosing spondylitis is a progressive inflammatory disease leading to immobility of the vertebral joints and fixation of the ribs. The inflammatory process, with tumor necrosis factor playing a crucial role, affects the articular processes, costovertebral joints, and

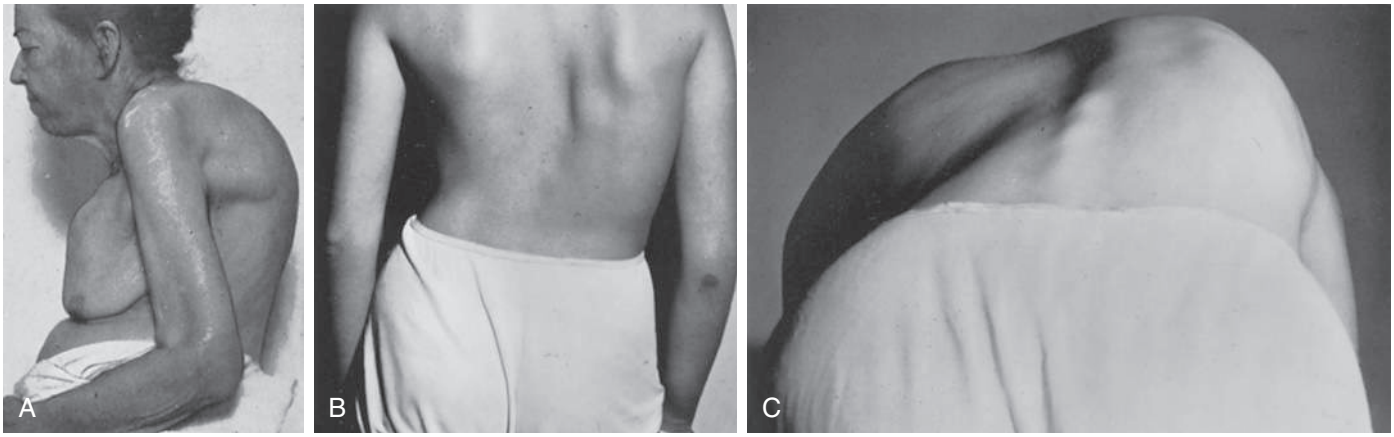


FIG 23.8 Kyphosis (**A**) and scoliosis (**B** and **C**) are structural deformities that can interfere with ventilation. (From Delp MH, Manning RT, editors: *Major's physical diagnosis,* ed 9, Philadelphia, 1981, Saunders.)

sacroiliac joints by inducing a fibrotic response leading to joint calcification, ligament ossification, and skeletal immobility.

Clinical manifestations. Initial symptoms include low to middle back pain and stiffness that is more severe after prolonged rest. With exercise, the pain and stiffness decrease. As the disease process advances, ribcage movement is greatly reduced, leading to restrictive lung dysfunction.

Chest wall muscular atrophy is common and leads to further restriction of ribcage expansion. Breathing is largely accomplished by excursion of the diaphragm as the ribcage becomes immobilized. Associated problems seen with the disease include arthritis, uveitis, spondylitic heart disease, pulmonary fibrosis, and polyarteritis.

Diagnosis. Pulmonary function tests show decreased VC, decreased TLC, and decreased compliance of the respiratory system, mainly the chest wall. Radiographs show destruction of cartilage, erosion of bone, calcification, and bony bridging of joint margins. The earliest radiologic changes are usually seen in the sacroiliac joints. Laboratory findings, although not diagnostic of the disease, include an elevated sedimentation rate in 85% of cases, as well as a decreased red blood cell count and an increased white blood cell count. HLA-B27 antigen is seen in 90% of white patients and 50% of black patients. However, 8% to 10% of the normal population have a positive HLA-B27.

Treatment. General therapy includes development of an exercise program that comprises breathing exercises and mobility exercises with spinal extension training. Pharmacologic management with nonsteroidal antiinflammatory agents (full dose, continuous therapy) provides symptomatic relief of pain and stiffness and promotes function.

Flail Chest

Etiology. Flail chest results from multiple rib fractures as a result of trauma to the chest wall. The ribs are fractured at two distant sites, resulting in an unstable, free-floating chest wall segment that moves paradoxically inward on inspiration and outward on expiration. Bilateral costochondral separation and sternal fractures can also cause a flail segment. Flail chest frequently occurs from the impact of the driver's chest with the steering wheel during an automobile accident.

Pathogenesis. Chest wall instability attributable to fracture at two distant sites on the same rib leads to an impairment of negative intrapleural pressure generation, causing decreased lung expansion on inspiration. The trauma commonly results in lung parenchymal injury, which may lead to pulmonary contusion, decreased lung compliance, and respiratory failure. Interstitial and alveolar hemorrhage leads to further abnormalities.

Clinical manifestations. Patients present after a trauma with paradoxical motion of the chest wall, either unilateral or bilateral. The injury to the chest wall is identified by careful inspection and palpation. Common features are marked shortness of breath, pain on inspiration, hypotension, cyanosis, and hypoxemia. The arterial PO_2 value is often low before clinical symptoms appear. Pneumothorax, hemothorax, and subcutaneous emphysema are common (see Fig. 23.5).

Diagnosis and treatment. Serial blood gas results help determine the treatment regimen. Flail chest with large segments resulting in acute respiratory failure is managed with mechanical ventilation. Mechanical ventilation is achieved by positive pressure, which causes the entire chest, including the flail section, to move as a unit rather than paradoxically. Pain management may be accomplished by continuous epidural anesthesia.

Disorders of Obesity

Etiology. *Obesity* is defined as excessive body fat, with a body mass index (BMI) greater than 30 kg/m^2 based on body weight and height. Overweight is defined as a BMI of 25 to 29.9 kg/m^2. Obesity results from excessive caloric intake and/or reduced caloric expenditure. The National Health and Nutrition Examination Survey (NHANES) reported that 60% of men and 50% of women are overweight. The findings for obesity were 20% of men and 25% of women. A higher prevalence of obesity was found in blacks than in whites and in persons with lower incomes than in those with higher incomes. Obese patients are at risk for a variety of disorders, the most common of which are diabetes mellitus, coronary artery disease, degenerative joint disease, gallstones, certain cancers (colon, rectum, and prostate in men; uterus, biliary tract, breast, and ovary in women), and pulmonary impairment. Approximately 60% of persons with obesity have metabolic syndrome. Persons with a BMI of ≥30 kg/m^2 have an all-cause increase in mortality of 50% to 100% compared with persons with a BMI between 20 and 25 kg/m^2.

Pathogenesis. Endocrine causes of obesity are rare. Only 4% to 6% of human obesity is thought to be due to single gene mutations. Hypothyroidism, the use of corticosteroids, and hypothalamic lesions all can lead to weight gain; however, the major cause of obesity is excess caloric intake in relation to caloric expenditure. Several hormones act on brain receptors to regulate appetite and metabolism. Leptin binds to brain receptors, causing the release of neuropeptides that promote satiety and increase metabolic rate. Ghrelin (a hunger hormone) stimulates appetite. Genetic diseases such as familial partial lipodystrophy, Prader–Willi syndrome, Laurence–Moon syndrome, Bardet–Biedl syndrome, and

Cohen syndrome are associated with obesity. Obesity may be associated with hypoventilation.

The mechanisms of obesity hypoventilation are reduced ventilatory drive and increased work of breathing. Some patients are thought to have an abnormality in the central nervous system. In addition, the increased abdominal size can force the abdominal contents upward into the chest cavity, thus decreasing lung expansion and diaphragmatic shortening. Obesity hypoventilation is also called pickwickian syndrome, named after the obese boy in *Pickwick Papers* written by Charles Dickens. Pickwickian syndrome is associated with hypoventilation and airway obstruction. An additional factor that contributes to the overall clinical picture in many obese patients is upper airway obstruction during sleep, the obstructive form of sleep apnea syndrome. Soft tissue deposits in the neck and tissues surrounding the upper airway predispose the person to episodes of complete upper airway obstruction during sleep. In a large percentage of cases, the daytime somnolence that occurs in patients who have obesity hypoventilation syndrome is related to obstructive sleep apnea.

Clinical manifestations. Obesity hypoventilation is characterized by decreased alveolar ventilation, somnolence, severe hypoxemia, polycythemia, and cor pulmonale. Patients complain of daytime somnolence, erectile dysfunction, shortness of breath, headache, and enuresis.

Diagnosis. The diagnosis of obesity is self-evident on examination. Tests for hypothyroidism, Cushing syndrome, insulinoma, diabetes, and hyperlipidemia may be done to identify comorbid factors. For persons with hypoventilation, arterial blood gas analyses may reveal hypoxemia and hypercapnia. Chest wall compliance, VC, TLC, and expiratory reserve volume are all decreased. Patients may also have an increased red blood cell count and show signs and symptoms of cor pulmonale and pulmonary hypertension.

Treatment. Primary treatment for obesity consists of a weight loss program that includes the family members. Caloric intake that promotes an energy deficit of 800 to 1000 kcal/day is recommended. Aerobic exercise preserves lean body mass and increases energy expenditure. Oxygen delivery through a nasal cannula or mechanical ventilation may be necessary for patients with morbid obesity. Surgical intervention with gastric stapling or gastric bypass to decrease the gastric volume and size has proved successful in some patients. These operations are intended to permanently curtail food intake.

KEY POINTS

- Neuromuscular diseases affect the muscles of respiration, leading to muscular weakness, increased risk of pulmonary infections, and respiratory failure.
- Kyphoscoliosis is a deformity of the bony structure of the chest wall characterized by hunchback and lateral curvature of the spine. The abnormal shape of the chest interferes with the normal mechanics of breathing, resulting in small lung volumes, compression atelectasis, and hypoxemia. Compensatory tachypnea is usually present.
- Ankylosing spondylitis is a progressive inflammatory disease affecting vertebrae and ribs. Chronic inflammation leads to chest wall fibrosis and immobility. Chest wall muscle atrophy and ribcage stiffening result in pulmonary dysfunction characteristic of restrictive disorders.
- Flail chest occurs as a result of blunt trauma to the chest that causes a segment of the ribcage to detach. Flail chest is a life-threatening injury.
- Obesity may interfere with the normal mechanics of breathing because of excessive chest weight and abdominal impingement on the chest cavity. Pickwickian syndrome is a disorder of obesity associated with hypoventilation and upper airway obstruction during sleep.

INFECTION OR INFLAMMATION OF THE LUNG

Pneumonia

Etiology. The term *pneumonia* (from the Greek *pneuma,* which means "breath") refers to an inflammatory reaction in the alveoli and interstitium of the lung, usually caused by an infectious agent. Pneumonia can result from three different sources: (1) aspiration of oropharyngeal secretions composed of normal bacterial flora and/or gastric contents (20% to 35% of all pneumonias); (2) inhalation of contaminants (virus, *Mycoplasma*); or (3) contamination from the systemic circulation.

There are several ways to classify pneumonia. Pneumonias are typically classified as community acquired or hospital acquired. Approximately 15% to 20% of persons presenting with pneumonia require hospitalization. Pneumonia is further classified as bacterial, atypical, and viral. The bacterial pneumonias may be grouped as either gram positive or gram negative, based on the staining characteristics of the organism. *Staphylococcus* and *Streptococcus* (including pneumococci) are the predominant gram-positive organisms. Gram-negative bacteria that may cause pneumonia include *Haemophilus influenzae, Klebsiella* species, *Pseudomonas aeruginosa, Serratia marcescens, Escherichia coli,* and *Proteus* species.

Patients at risk of pneumonia include the elderly; those with a diminished gag reflex; seriously ill, hospitalized patients; hypoxic patients; and immunocompromised patients. Anaerobic bacteria may present clinically as a lung abscess, necrotizing pneumonia, or empyema. These diseases are usually caused by aspiration of normal oral bacteria (such as *Bacteroides* and *Fusobacterium*) into the lung. Mycoplasmal pneumonia is more commonly seen in the summer and fall in young adults. About half of the cases of pneumonia in persons between 5 and 20 years of age can be classified as mycoplasmal pneumonia. Other causes of pneumonia occur less frequently in the general population. Legionnaires' disease, for example, is a severe systemic illness characterized by fever, diarrhea, abdominal pain, liver and kidney failure, and pulmonary infiltrates. The causative organism for legionnaires' disease lives in water and is transmitted by means of potable water, condensers, and cooling towers. The current treatment of choice is administration of a macrolide antibiotic. Patients whose immune systems have been compromised by disease or by drug therapy may be susceptible to the development of opportunistic pneumonia. For example, *Pneumocystis (carinii) jiroveci* pneumonia, an opportunistic fungal infection, is commonly found in patients with cancer or HIV. (See Chapter 12 for further discussion of AIDS.)

Aspergillus, an opportunistic fungus that is widespread in nature, may cause progressive pneumonia. *Aspergillus* is released from the walls of old buildings under reconstruction. Attention should be given when old hospitals are renovated and when susceptible patients are located in a reconstruction area. To assist the reader in differentiating among the various types of pneumonia, Table 23.7 presents the etiologic factors, common clinical features with age-related characteristics, radiologic findings, and antibiotic therapies for 11 forms of the disease. There are many other types of pneumonia that are not listed.

Pathogenesis. Normally, pulmonary defense mechanisms (immune responses, cough reflex, sneezing, mucociliary clearance) protect individuals from pneumonia. Community-acquired pneumonia occurs when defense mechanisms are compromised. A highly virulent organism may also overwhelm a person's defense mechanisms. Community-acquired pneumonias are commonly bacterial in origin. After microbial agents enter the lung, they multiply and trigger pulmonary inflammation. Alveolar air spaces fill with an exudative fluid, and inflammatory cells invade the alveolar septa. Acute bacterial pneumonia may be associated with significant $\dot{V}_A/\dot{Q}$ mismatching and hypoxemia, because inflammatory

TABLE 23.7 Differentiating Features of Types of Pneumonia

Etiologic Organism	Common Clinical Features	Chest Radiograph	Antibiotic Treatment
Staphylococcus aureus; gram-positive cocci in clumps	Follows upper respiratory tract infection; fever, chills, pleuritic chest pain, cough, yellow purulent sputum; seen in patients in chronic care facilities	Consolidation, may have cavitation	Methicillin-susceptible strains: nafcillin or oxacillin with or without rifampin; methicillin-resistant strains: vancomycin with or without rifampin; alternative choice: cephalosporins, clindamycin, vancomycin
Streptococcus pneumoniae (pneumococcus); gram-positive diplococci	More common in alcoholics; also seen with chronic cardiopulmonary disease; fever, chills, pleuritic chest pain, cough, rust-colored sputum	Patchy infiltrates	Procaine penicillin G or aqueous penicillin G, amoxicillin; alternative choice: macrolides, cephalosporins, doxycycline, quinolones; prophylactic vaccine available
Haemophilus influenzae; pleomorphic gram-negative coccobacilli	Upper respiratory tract symptoms, fever, vomiting, irritability, cough, purulent sputum, dyspnea; affects children and older adults; affects people with chronic cardiorespiratory problems	Consolidation	Cefotaxime, ceftriaxone, doxycycline, azithromycin, TMP-SMX; alternative choice: quinolones or clarithromycin
Klebsiella pneumoniae; gram-negative encapsulated rods	Seen frequently in middle-aged men and associated with alcoholism and diabetes mellitus; rust-colored sputum	Consolidation	Aminoglycoside plus third-generation cephalosporin; alternative: aztreonam, imipenem, quinolone
Pseudomonas aeruginosa; gram-negative rods	Chronic obstructive pulmonary disease, cystic fibrosis, and mechanical ventilation; fever, chills, and copious greenish, foul-smelling sputum	Infiltrates, small pleural effusion	Aminoglycoside plus ticarcillin/clavulanate or piperacillin/tazobactam or aztreonam or imipenem
Escherichia coli; gram-negative rods	Complication of gastrointestinal surgery	Infiltrates, may have pleural effusion	Aminoglycoside plus third-generation cephalosporin; alternative: aztreonam, imipenem, quinolone
Virus	Fever, malaise, headache, nonproductive cough	Patchy infiltrates	Amantadine, rimantadine
Legionella species; no bacteria	Acute onset with fever, diarrhea, myalgia, and abdominal pain	Consolidation	Macrolides with or without rifampin; alternative: TMP-SMX, quinolone
Mycoplasma pneumoniae (atypical pneumonia); monocytes and neutrophils; no bacteria	Ages 5–25 years; most common in young adults; associated with otitis media and myringitis; sore throat, headache, myalgia, dry cough, fatigue, low-grade fever	Infiltrates	Erythromycin, doxycycline; alternative: quinolone or other macrolide
Pneumocystis (carinii) jiroveci (fungus)	Immunosuppressed patients (infants, children, and adults); 60% of patients have AIDS	Diffuse infiltrates, or chest x-ray may appear normal	TMP-SMX or pentamidine; isethionate plus prednisone; alternative: dapsone plus TMP-SMX, clindamycin plus primaquine
Anaerobic pneumonia (aspiration pneumonia); mixed flora	Predisposition to aspiration, fever, weight loss, malaise; risk increases with decreased level of consciousness, artificial airway, and sedation; seen in individuals with poor dental hygiene	Infiltrates in dependent lung fields	Penicillin G; alternative choices: clindamycin, metronidazole, cefoxitin

AIDS, Acquired immunodeficiency syndrome; *TMP-SMX,* trimethoprim-sulfamethoxazole.

exudate collects in the alveolar spaces. Alveolar exudate tends to consolidate and becomes difficult to expectorate. Viral pneumonia does not produce exudative fluids. Fig. 23.9 shows the histologic progression of acute bacterial pneumonia. Patients with chronic illnesses and those who are immobile or immunosuppressed or have a decreased level of consciousness are at highest risk for developing pneumonia. Disruption of the body's normal defense mechanisms leads to increased risk. Other patients at risk are those who have undergone thoracic or abdominal surgery or have received a general anesthetic.

Clinical manifestations. Clinically, the pathogenic cause, severity of the disease, and age of the patient may cause variations in the presentation of pneumonia. Some patients present with fever only. Crackles (rales) and bronchial breath sounds may be heard over the affected lung tissue. Patients may present with chills, cough, purulent sputum, and an abnormal chest radiograph. Patients with viral pneumonia may present with an upper respiratory prodrome (fever, coryza [inflammation of the mucous membranes in the nose], cough, hoarseness) accompanied by wheezing and/or rales. Typical features of *Chlamydia* pneumonia are cough, tachypnea, rales, wheezes, and absence of fever. *Mycoplasma* pneumonia is a common cause of pneumonia in older children and adults. Signs and symptoms include fever, cough, headache, and malaise.

Diagnosis. The chest radiograph demonstrates parenchymal infiltrates (white shadows) in the involved area, indicative of inflammatory alveolar processes. In a patient with symptoms and clinical findings of pneumonia, a Gram stain of expectorated sputum from deep in the lungs may be obtained to distinguish bacterial from viral pneumonia and gram-negative from gram-positive organisms. If the patient had been previously healthy, the cause of the majority of these infections would be viral, mycoplasmal, or the gram-positive pneumococcal bacterium. However, if the patient

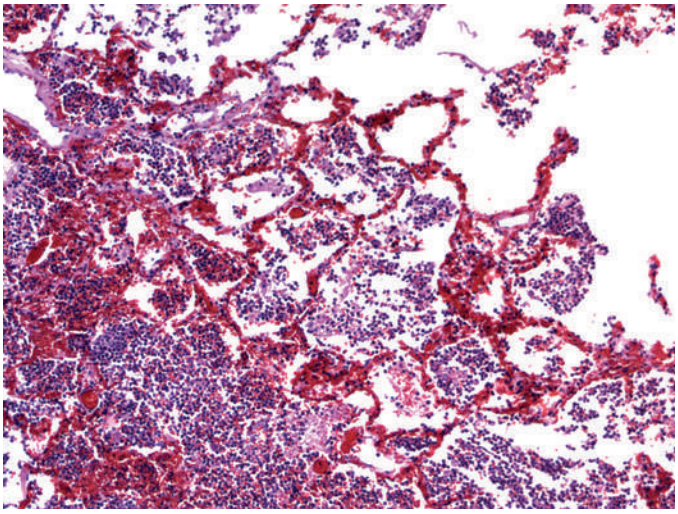


FIG 23.9 Acute pneumococcal pneumonia (microscopic): Distended capillaries and congestion with prominent neutrophilic infiltrate and erythrocytes. (From Procop GW, Pritt BS: *Pathology of infectious diseases,* Philadelphia, 2015, Saunders.)

KEY POINTS

- Pneumonia is an inflammation of the lung that is usually associated with an infectious agent. The most common types of pneumonia are bacterial, mycoplasmal, and viral. A productive cough is the primary differentiating feature between bacterial pneumonia and viral pneumonia, in which coughing is nonproductive.
- Acute bacterial pneumonia may be associated with significant $\dot{V}_A/\dot{Q}$ mismatching and poor blood gas values because inflammatory exudate collects in the alveolar spaces. Alveolar exudate tends to consolidate and becomes difficult to expectorate. Viral pneumonia does not produce exudative fluids.
- Manifestations of bacterial pneumonia may include fever, chills, cough with purulent sputum, crackles, and areas of consolidation on chest radiograph. Dyspnea may be significant.
- The treatment of bacterial pneumonia centers on antibiotic therapy to eliminate the organism and supportive therapy to enhance ventilation and oxygenation. Most cases of viral pneumonia (influenza) are managed symptomatically because no effective antibiotic therapy is available.
- Fungal and protozoal pneumonias are uncommon and tend to occur in immunocompromised individuals.
- SARS and MERS are acute respiratory tract infection caused by a coronavirus. The pneumonia is severe and has a mortality of 10% for SARS and 40% for MERS.

had been hospitalized or has other illnesses such as emphysema, diabetes, or alcoholism, then gram-negative organisms should be suspected. "CURB-65" may be used to determine whether the patient should be hospitalized. "CURB-65" includes (1) *c*onfusion, (2) B*U*N >19.6 mg/dL, (3) *r*espiratory rate >30 breaths/min, (4) systolic *b*lood pressure (BP) <90 mm Hg and diastolic BP ≤60 mm Hg, and (5) age ≥65 years. Blood cultures are positive in approximately 20% of hospitalized patients. Because 48 to 72 hours may be required for culture of the etiologic agent, antibiotic therapy should be started empirically. Once culture and sensitivity results are obtained, antibiotic therapy may be changed. Diagnosis is based on the chest radiograph, white blood cell count (more than 15,000/μL for acute bacterial pneumonia), and sputum culture, coupled with clinical features of fever with recurrent chills, cough, dyspnea, and rales.

Treatment. Once the organism has been cultured, specific antibiotic selection is based on sensitivity of the organism to different antibiotics. Empirical treatment before culture results for community-acquired pneumonia consists of levofloxacin or a macrolide antibiotic. Table 23.7 presents treatment options. The chest x-ray should be repeated 6 to 8 weeks after the infection has been treated.

Severe Acute Respiratory Syndrome

SARS was first reported in February 2003 as a severe form of pneumonia occurring in Asia. There were only eight confirmed cases in the United States during the epidemic. Active global surveillance for SARS in human beings had detected no further confirmed person-to-person transmission of the disease between July 2003 and August 2004. However, two cases of SARS occurred in persons working in laboratories in southern China. At the time of this writing, there are no reported cases. See Table 23.8 for information on SARS.

Middle East Respiratory Syndrome

Middle East respiratory syndrome (MERS) is caused by a coronavirus similar to the one that causes SARS. Early cases were identified in 2012 in Saudi Arabia. It is spread by respiratory droplets. The original source of the virus is camels and camel's milk. See Table 23.8 for further information.

Pulmonary Tuberculosis

Etiology. One-third of the world's population has tuberculosis infection and nearly 2 million people die of TB each year. In the United States there are 3 cases per 100,000 persons. More than 90% of cases involve reactivation of prior infection; the remainder are new infections. The majority of new cases occur in malnourished individuals, those living in overcrowded conditions, immunosuppressed individuals, incarcerated persons, immigrants (36% of new cases in the United States), and elderly persons. Two-thirds of all new cases are found in racial and ethnic minorities. New cases of TB from racial and ethnic minorities are 80%. During the past 45 years, there has been a shift in the care of patients from specialized TB hospitals to outpatient therapy. Hospitalization of patients may be necessary with implementation of isolation precautions for a period of 2 to 4 weeks (longer for multidrug-resistant tuberculosis [MDR-TB]). In some countries, specialized TB hospitals have reopened due to increasing resistance of the organism to treatment and an increasing number of cases. India and China account for the highest number of MDR-TB cases worldwide. Tuberculosis cases should be reported to local and state health departments.

TB is caused by the bacterium *Mycobacterium tuberculosis,* an acid-fast aerobic bacillus. Any organ system can be affected by the disease, but the most common sites are the lungs and the lymph nodes. TB is subdivided into two major classifications: primary (usually clinically and radiographically silent) and reactivating. Primary disease (initial infection) may lie dormant for many years or decades. When the person's immune system becomes impaired, reactivation may occur. HIV, corticosteroid use, silicosis, and diabetes mellitus have been found to be associated with reactivation. Reactivation may occur many years after the primary infection. Distant organ systems may be involved as a result of hematogenous spread during the primary or reactivation phase of infection. In addition, there may be disseminated disease, known as *miliary tuberculosis,* again resulting from hematogenous dissemination of the organisms. Strains of *M. tuberculosis* are becoming resistant to one or more first-line antituberculosis drugs. Entry into the body is by inhalation of small (2 to 10 μm) droplets containing the bacteria.

TABLE 23.8 Differentiating Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS)

	SARS	MERS
Etiology	SARS is spread by person-to-person contact via inhaled particles.	The virus is spread by person-to-person contact via inhaled particles. Original source of MERS is camels and camel's milk.
Pathogenesis	The coronavirus targets epithelial cells in the respiratory tract. The damage leads to extensive hyaline membrane formation, alveolar collapse, desquamation of alveolar cells, and development of fibrous tissue. Vascular injury occurs at the same time. Other organs damaged are the brain, spleen, and kidneys.	The coronavirus targets the dipeptidyl peptidase 4 (DDP4) receptors. The MERS virus establishes an infection in monocytes and macrophages. The release of cytokines from the infected cells leads to severe inflammation and tissue damage. Vascular endothelial cells in the pulmonary interstitium may also be infected, leading to dissemination of the infection.
Clinical manifestations	Fever >100.4°F, chills, headache, muscle aches, nonproductive cough, dyspnea	Fever, productive cough, dyspnea, chills, rigors and diarrhea, muscle aches. The disease progresses to ARDS.
Diagnostics	History of travel to regions identified by WHO as areas with recent local transmission or sexual/casual contact with someone diagnosed with SARS. Positive ELISA, immunofluorescence, or PCR test for SARS. Chest x-ray shows increased opacity bilaterally.	History of contact with an individual with the disease. Positive PCR on two samples from the lower respiratory tract. Chest x-ray shows bilateral patchy infiltrates with more involvement in the lower lobes.
Treatment	Supportive treatment with antipyretics, oxygen, and mechanical ventilation if needed. Requires negative-pressure isolation room.	No specific antiviral treatment is recommended. Treatment with interferon-12b and ribavirin has been tried, but the patients did not survive. Supportive treatment with antipyretics, oxygen, and mechanical ventilation if needed.
Outcomes	No vaccine available. The overall fatality rate is 9.6% from the last epidemic. Some patients have severe long-term sequelae.	No vaccine available. Overall fatality rate is approximately 40%.

ELISA, Enzyme-linked immunosorbent assay; *PCR,* polymerase chain reaction; *WHO,* World Health Organization.

Infected droplets are expelled when an infected person coughs, sneezes, or talks.

Pathogenesis. After infected droplets of *Mycobacterium* are inhaled into the alveoli of the susceptible person, alveolar macrophages ingest and process the microorganisms. *Mycobacterium* are then transported to lymph nodes where spread of the disease is contained. Primary TB (asymptomatic) is an infection that continues for 2 to 12 weeks until T cells contain the infection. T-cell–mediated hypersensitivity is evidenced by a positive skin test reaction to tuberculin purified protein derivative (PPD).

If *Mycobacterium* reach the bloodstream, they disseminate rapidly. Once the infection becomes established, lymphatic and hematogenous dissemination occurs. T cells and macrophages surround the organisms in granulomas that limit multiplication and spread. Dormant organisms (latent TB) persist for years. Reactivation may occur if the patient's immune system becomes impaired. *M. tuberculosis* is a slow-growing aerobic, non–spore-forming, nonmobile bacillus. The pathologic manifestation of pulmonary TB is the Ghon tubercle or complex, which has parenchymal and lymph components. The parenchymal component is composed of a well-circumscribed, necrotic nodule that later becomes fibrotic and calcified. The lymph component is found in the lymph nodes. Primary pulmonary TB is shown in Fig. 23.10. In summary, the organisms either are destroyed or persist and multiply, causing active disease. Reactivation pulmonary TB occurs months to years after the primary TB and involves apical posterior segments of the upper lobes and superior segments of the lower lobes. *Mycobacterium* prefer areas of high concentration of oxygen.

Clinical manifestations. Clinical features of reactivated disease include a history of contact with an infected person, low-grade fever, cough, hemoptysis, night sweats, fatigue, weight loss, malaise, and anorexia. Chronic cough is the most common symptom. As the disease progresses, the patient develops a productive cough with purulent sputum. Physical examination of the lung fields reveals apical crackles (rales) (*M. tuberculosis* organisms prefer lung apices because of the higher concentration of oxygen in this area) or bronchial breath

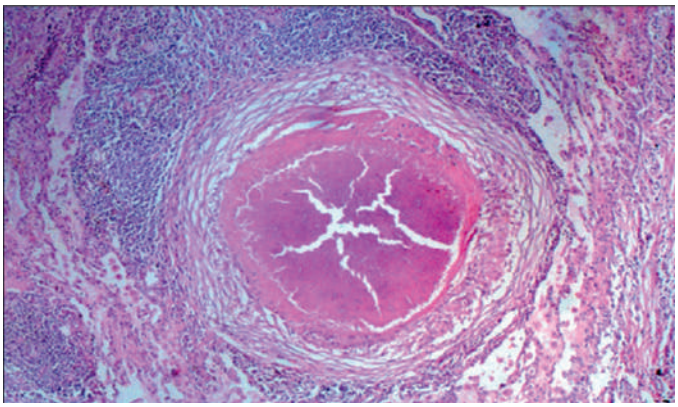


FIG 23.10 Micrograph of tissue specimen from patient with tuberculosis. Low-power micrograph (×100) of a haematoxylin and eosin stained tissue section from an immunocompetent patient with tuberculosis that shows a well-formed tuberculous granuloma with a central area of caseous necrosis surrounded by epithelioid macrophages, giant cells, and T lymphocytes, and surrounding outer fibrosis. (From Lawn SD, Zumla AL: Tuberculosis. *Lancet* 2011;378(9785):57–72.)

sounds over the region of lung consolidation. The patient appears malnourished and chronically ill. Common sites of extrapulmonary TB are the peritoneum, gastrointestinal tract, liver, spleen, bone, joints, lymph nodes, central nervous system, and genitourinary system. (Refer to Chapter 12 for a discussion of TB in HIV-infected patients.) Results of pulmonary function tests are characteristic of restrictive diseases, with decreased lung volumes and decreased compliance.

Diagnosis. Definitive diagnosis is made by results of sputum culture for acid-fast bacilli or by identification of the organism by DNA or RNA amplification techniques. Three consecutive morning sputum specimens are obtained to identify the slow-growing acid-fast bacillus. Expectoration of sputum in the early morning is ideal because the sputum is more concentrated and more plentiful. Cultures require 1 to 3 weeks for determination. Gastric washings or bronchial washings may also be used for diagnostic culturing.

Chest radiographs usually show nodules with infiltrates in the lung apex and posterior segments of the upper lobes. Primary infection shows calcified peripheral lung nodules with calcified hilar nodes in the region of the nodules. Elderly patients may present with lower lobe infiltrates with or without pleural effusion. Fig. 23.11 shows the radiographic appearance of cavitary TB in a 23-year-old man. A miliary pattern (diffuse, small, nodular densities) is seen with dissemination of the organism in miliary TB. Another diagnostic test is the tuberculin (Mantoux test) skin test (5 tuberculin units/0.1 mL of PPD injected intradermally). A second dose is administered a week later. This test does not distinguish between current disease and past infection. A negative result never rules out acute TB. If the induration in a person with HIV infection is 5 mm or greater, if the patient has close contact with individuals with TB, and if the patient has a chest radiograph consistent with TB, the likelihood of active disease is high. An induration of 10 mm or greater is the reaction size for other high-risk individuals, such as intravenous drug abusers, individuals who are debilitated, children younger than 4 years or with immunosuppression, and individuals living in areas with a high incidence of the disease (Asia, Africa, Latin America). An induration of 15 mm or greater is considered positive for TB in all other persons. An interferon gamma release assay (Quantiferon test) may be preferred to detect *M. tuberculosis* in latent TB infections. False-positive PPD results may occur in persons with other mycobacterial infections or if they have received bacille Calmette-Guérin (BCG), a live attenuated strain of *Mycobacterium bovis* that provides active immunity against TB. False-negative results may also occur in patients who are malnourished, elderly, or immunocompromised. Immunocompromised patients may not be able to mount a response (wheal) to injection of the organism.

Treatment. Primary therapy for active TB consists of (1) administering multiple drugs to which the organism is susceptible; (2) adding at least two new agents to the drug regimen when treatment failure is suspected; (3) providing the safest, most effective therapy for the shortest period; and (4) ensuring adherence to therapy by utilizing directly observed therapy. Nonadherence to therapy because of adverse drug reactions is a major cause of treatment failure. HIV-negative individuals with active *M. tuberculosis* and without MDR-TB are typically treated for 6 months, and HIV-positive individuals are treated for 9 months total therapy.

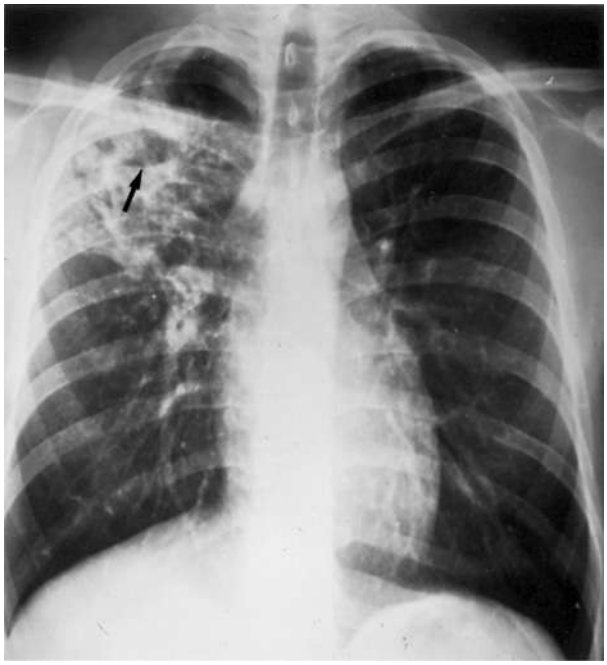


FIG 23.11 Cavitary pulmonary tuberculosis in a 23-year-old man. (From Kersten LD: *Comprehensive respiratory nursing,* Philadelphia, 1989, Saunders, p 146.)

KEY POINTS

- Tuberculosis is caused by inhalation or ingestion of the bacterium *M. tuberculosis.* The organism spreads through the lymph and blood. Bacteria are ingested by macrophages and walled off by inflammatory proteins (granulomas). The organisms may not be killed and can persist in a dormant state for years. These walled-off areas of inflammatory cells and bacteria become fibrotic and calcified, forming Ghon tubercles—the hallmark of TB.
- Symptoms are somewhat nonspecific: low-grade fever, cough, night sweats, fatigue, and weight loss. With progression of the disease, the cough is productive of purulent sputum.
- The diagnosis is based on a positive PPD skin test for TB, positive Quantiferon test, positive sputum cultures, and characteristic nodules on chest radiographs.
- Multiple antimicrobial agents are used for managing TB. Drug therapy continues for 6 to 9 months, depending on HIV status for active disease, and may be used for shorter periods in persons exposed to TB but with no active disease. Referral to a specialist is recommended.

SUMMARY

Restrictive pulmonary disorders are those in which lung expansion is restricted. Restrictions are commonly caused by diseases that affect the lung parenchyma (e.g., diffuse interstitial pulmonary fibrosis), chest wall disorders, neuromuscular disorders, pleural space disorders, pneumonia, and TB. These diseases are characterized by a reduced VC and a small residual lung volume. They differ from obstructive diseases, covered in Chapter 22, in that airway resistance is not increased.

RESOURCES

Disorders of Parenchyma

Beard L: Neonatology. In Engorn B, Flerlage J, editors: *The Harriet Lane handbook*, ed 20, Philadelphia, 2014, Mosby, pp 423–426.

Beasley MB, Travis WD, Rubin E: The respiratory system. In Rubin E, Reisner HM, editors: *Essentials of Rubin's pathology*, ed 6, Philadelphia, 2013, Lippincott Williams & Wilkins, pp 317–350.

Bernstein DI, Kim H: Hypersensitivity pneumonitis. In Bope ET, Kellerman RD, editors: *Conn's current therapy*, Philadelphia, 2015, Elsevier, pp 393–394.

Chakinala M, McEvoy C, Russell TD, et al: Pulmonary diseases. In Godara H, et al, editors: *The Washington manual of medical therapeutics*, ed 34, Philadelphia, 2013, Lippincott Williams & Wilkins, pp 322–376.

Chesnutt MS, Prendergast TJ: Pulmonary disorders. In Papadakis MA, McPhee SI, editors: *Current medical diagnosis and treatment*, ed 54, New York, 2015, Lange/McGraw-Hill, pp 240–319.

Costabel U: Ask the expert—diffuse interstitial lung disease. *Breathe* 4(2):165–172, 2007.

Czaja AJ, Exo J, et al: Critical care. In Hay WW, et al, editors: *Current pediatrics: diagnosis and treatment*, ed 22, New York, 2014, Lange/McGraw-Hill, pp 387–424.

Ferreria A, Garvey C, Conners GL, et al: Pulmonary rehabilitation in interstitial lung disease. *Chest* 135(2):442–477, 2009.

Ferri FF: Acute respiratory distress syndrome. In Ferri FF, editor: *Ferri's clinical advisor: 5 books in 1*, St Louis, 2016, Mosby, pp 73–76.

Ferri FF: Sarcoidosis. In Ferri FF, editor: *Ferri's clinical advisor: 5 books in 1*, St Louis, 2016, Mosby, pp 1097–1098.

Goroll AH, Mulley AG: Management of sarcoidosis: evaluation of interstitial lung disease. In Goroll AH, Mulley AG, editors: *Primary care medicine: office evaluation and management of the adult patient*, ed 7, Philadelphia, 2014, Lippincott Williams & Wilkins, pp 425–428.

Hansen AK, Wisborg K, Uldberg N, Henriksen TB: Risk of respiratory morbidity in term infants delivered by elective caesarean section: cohort study. *BMJ* Online First, 1–7, 2007.

Kramer K: Hypersensitivity pneumonitis. In Ferri FF, editor: *Ferri's clinical advisor: 5 books in 1*, St Louis, 2016, Mosby, pp 663–664.

McEvoy C, Schilling D, Peters D, et al: Respiratory compliance in preterm infants after single rescue of antenatal steroids: a randomized controlled trial. *Am J Obstet Gynecol* 202(6):544.e1–544.e9, 2010.

Medline Plus: *Neonatal respiratory distress syndrome*, Available at: www.nlm.nih.gov/medlineplus/ehex/article1001563.htm. (Accessed September 2015).

Oliver LC: Evaluation and prevention of occupational and environmental respiratory disease. In Goroll AH, Mulley AG, editors: *Primary care medicine: office evaluation and management of the adult patient*, ed 7, Philadelphia, 2014, Lippincott Williams & Wilkins, pp 344–349.

Pagana KD, Pagana TJ, Paganai TN: *Mosby's diagnostic and laboratory test reference*, ed 12, St Louis, 2015, Mosby, pp 49–50.

Pramanik AK: *Respiratory distress syndrome*. emedicine, medscape.com/article/976034-overview#96. Available at: www.netsvic.org.au/nets/handbook/index.cfm. (Accessed February 2012).

Ryu JH, Selman M, Colby TV, King TE: Idiopathic interstitial pneumonias. In Broaddus VC, Mason RJ, Ernst JD, et al, editors: *Murray and Nadel's textbook of respiratory medicine*, ed 6, Philadelphia, 2015, Saunders, pp 1118–1152.

Shih G, Gleit C, Marsh MA: Interstitial lung disease. In Ferri FF, editor: *Ferri's clinical advisor: 5 books in 1*, St Louis, 2016, Mosby, pp 714–717.

Toren K, Blanc PD: Asthma caused by occupational exposures is common—a systematic analysis of estimates of the population—attributable fraction. *BMC Pulm Med* 9:7, 2009.

West JB: *Pulmonary pathophysiology: the essentials*, ed 8, Philadelphia, 2013, Lippincott Williams & Wilkins.

West JB: *Pulmonary physiology and pathophysiology: an integrated, case-based approach*, ed 2, Philadelphia, 2007, Lippincott Williams & Wilkins.

Disorders of Chest Wall and Pleura

Baron RB: Nutrition. In Papadakis MA, editor: *Current medical diagnosis and treatment*, ed 54, New York, 2015, Lange/McGraw-Hill, pp 1245–1263.

Basman CL: Obesity. In Ferri FF, editor: *Ferri's clinical advisor: 5 books in 1*, St Louis, 2016, Mosby, pp 882–885.

Deatrick KB, Long J, Chang AC: Thoracic wall, pleura, mediastinum and lung. In Doherty GM, editor: *Current surgical diagnosis and treatment*, ed 14, New York, 2015, Lange/McGraw-Hill, pp 331–389.

Erickson MA, Capria B, Eilert RE: Orthopedics. In Hay WW, et al, editors: *Current pediatrics: diagnosis and treatment*, ed 22, New York, 2014, Lange/McGraw-Hill, pp 862–883.

Federico MJ, Baker CD, Balasubramaniam V, et al: Disorders of the pleura and pleural cavity. In Hay WW, et al, editors: *Current pediatrics: diagnosis and treatment*, ed 22, New York, 2014, Lange/McGraw-Hill, pp 579–581.

Federico MJ, Baker CD, Balasubramanian V, et al: Respiratory tract and mediastinum. In Hay WW, et al, editors: *Current pediatrics: diagnosis and treatment*, ed 22, New York, 2014, Lange/McGraw-Hill, pp 534–537.

Ferri FF: Scoliosis. In Ferri FF, editor: *Ferri's clinical advisor: 5 books in 1*, St Louis, 2016, Mosby, p 1109.

Figley DJ, Flores R: Pleural effusion and empyema thoracis. In Bope ET, Kellerman RD, editors: *Conn's current therapy*, Philadelphia, 2015, Elsevier, pp 397–399.

Fitzgerald PA: Endrocrine Disorders. In Papadakis MA, McPhee SI, editors: *Current medical diagnosis and treatment*, ed 54, New York, 2015, Lange/McGraw-Hill, pp 1081–1084.

Gordon CE, Feller-Kopman D, Balk EM, Smetana GW: Pneumothorax following thoracentesis. *Arch Intern Med* 170(4):332–339, 2010.

Goroll AH, Mulley AG: Evaluation of pleural effusions. In Goroll AH, Mulley AG, editors: *Primary care medicine: office evaluation and management of the adult patient*, ed 7, Philadelphia, 2014, Walters Kluwer Health, pp 363–369.

Hall JE: *Guyton and Hall: textbook of medical physiology*, ed 13, Philadelphia, 2015, Elsevier.

Harrison T: Amyotrophic lateral sclerosis. In Ferri FF, editor: *Ferri's clinical advisor:5 books in 1*, St Louis, 2016, Mosby, p 106.

Harrison T: Myasthenia Gravis. In Ferri FF, editor: *Ferri's clinical advisor: 5 books in 1*, St. Louis, 2016, Mosby, p 835.

Hellmann DB, Imbodem JB: Arthritis and musculoskeletal disorders. In Papadakis MA, McPhee SJ, editors: *Current medical diagnosis and treatment*, ed 54, New York, 2015, Lange/McGraw-Hill, pp 846–847.

Hemmila MR, Wahl WL: Management of the injured patient. In Doherty GM, editor: *Current surgical diagnosis and treatment*, ed 14, New York, 2015, McGraw-Hill, pp 191–222.

Light RW: Disorders of the pleura. In Kasper DL, Fauci AS, et al, editors: *Harrison's principles of internal medicine*, ed 19, New York, 2015, McGraw-Hill, pp 316–317.

Regnante R, Steiner J: Pneumothorax, spontaneous. In Ferri FF, editor: *Ferri's clinical advisor: 5 books in 1*, St Louis, 2016, Mosby, pp 973–975.

Rosenberg AA, Grover T: The newborn infant. In Hay WW, et al, editors: *Current pediatrics: diagnosis and treatment*, ed 22, New York, 2014, Lange/McGraw-Hill, pp 9–75.

Shandera WX, Zolopa AR: Viral and rickettsial infections. In Papadakis MA, McPhee SJ, editors: *Current medical diagnosis and treatment*, ed 54, New York, 2015, Lange/McGraw-Hill, pp 1339–1410.

Singhal D: Guillain-Barré syndrome. In Ferri FF, editor: *Ferri's clinical advisor: 5 books in 1*, St Louis, 2016, Mosby, pp 559–560.

The National Health and Nutrition Examination Survey (NHANES): *Analytic and reporting guidelines,* Available at: www.cdc.gov/nchs/nhanes.htm. (Accessed September 2015).

The pink book: course textbook, ed 13, 2015. Available at: www.cdc.gov/vaccines/pubs/pinkbook/polio.html#epi. (Accessed September 2015).

Wong CL, Holroyd-Loduc J, Straus SE: Does this patient have a pleural effusion? *JAMA* 301(3):309–317, 2009.

Pulmonary Infections

Centers for Disease Control and Prevention: *Global tuberculosis*, Available at: www.cdc.gov/topic/globaltb/default.html. (Accessed September 2015).

Centers for Disease Control and Prevention: *Severe acute respiratory syndrome,* Available at: www.cdc.gov/sars/about/index.html. (Accessed September 2015).

Chesnutt MS, Pendergast TJ: Pneumonia. In Papadakis MA, McPhee SI, editors: *Current medical diagnosis and treatment*, New York, 2015, Lange/McGraw-Hill, pp 266–274.

Chestnutt MS, Pendergast TJ: Pulmonary tuberculosis. In Papadakis MA, McPhee SI, editors: *Current medical diagnosis and treatment*, New York, 2015, Lange/McGraw-Hill, pp 275–283.

Federico MJ, Baker CD, Balasbramaniam V, et al: Acquired disorders involving the alveoli. In Hay WW, et al, editors: *Current pediatrics: diagnosis and treatment*, ed 22, New York, 2014, Lange/McGraw-Hill.

Ferri FF: Bacterial pneumonia. In Ferri FF, editor: *Ferri's clinical advisor: 5 books in 1*, St Louis, 2016, Mosby, pp 196–198.

Fort GG: Aspiration pneumonia. In *Ferri's Clinical advisor: 5 books in 1*, St. Louis, 2016, Mosby, p 151.

Fort GG: Tuberculosis, pulmonary. In Ferri FF, editor: *Ferri's clinical advisor: 5 books in 1*, St Louis, 2016, Mosby, pp 1257–1259.

Walpert MI: Influenza. In Ferri FF, editor: *Ferri's clinical advisor: 5 books in 1*, St Louis, 2016, Mosby, pp 705–707.

24

Fluid and Electrolyte Homeostasis and Imbalances

Linda Felver

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- What physiologic and pathophysiologic conditions predispose an individual to disturbances in fluid intake?
- How do the compositions of plasma and interstitial fluids differ? How are they similar?
- What regulates water and electrolyte movement between plasma and interstitial fluids? Across cell membranes?
- What are the usual and pathologic routes of fluid exit from the body?
- Under what conditions are extracellular volume deficit and excess likely to occur, and what are the characteristic clinical findings?
- Under what conditions are hyponatremia (water excess) and hypernatremia (water deficit) likely to occur, and what are the characteristic clinical findings?
- What capillary-level mechanisms cause edema?
- What physiologic and pathophysiologic conditions can lead to alterations in electrolyte intake, absorption, distribution, or excretion? How do these differ between specific electrolytes?
- What are the characteristic clinical findings of plasma excesses and deficits of potassium, calcium, magnesium, and phosphate ions?

CHAPTER OUTLINE

Body Fluid Homeostasis, 522
Fluid Intake and Absorption, 522
Fluid Distribution, 523
Fluid Excretion, 524
Fluid Loss Through Abnormal Routes, 524
Fluid Imbalances, 525
Extracellular Fluid Volume, 525
Volume Deficit, 525
Volume Excess, 526
Body Fluid Concentration, 526
Hyponatremia, 526
Hypernatremia, 527
Both Volume and Concentration, 528
Clinical Dehydration, 528
Interstitial Fluid Volume, 528
Edema, 528
Principles of Electrolyte Homeostasis, 529
Electrolyte Intake and Absorption, 529
Electrolyte Distribution, 531
Electrolyte Excretion, 531
Electrolyte Loss Through Abnormal Routes, 531
Electrolyte Imbalances, 531
Plasma Potassium, 531
Hypokalemia, 532
Hyperkalemia, 532
Plasma Calcium, 533
Hypocalcemia, 533
Hypercalcemia, 533
Plasma Magnesium, 534
Hypomagnesemia, 534
Hypermagnesemia, 534
Plasma Phosphate, 535
Hypophosphatemia, 535
Hyperphosphatemia, 536

The fluid in the body flows in arteries, veins, and lymph vessels; it is secreted into specialized compartments as diverse as joints, cerebral ventricles, and the intestinal lumen; it both surrounds and permeates the cells. Body fluid serves as a lubricant and as a solvent for the chemical reactions that we call *metabolism;* it transports oxygen, nutrients, chemical messengers, and waste products to their destinations; it plays an important role in the regulation of body temperature. Because the fluid within the body is so widespread and serves so many functions, it is not surprising that abnormalities in the volume, concentration, or electrolyte composition of body fluid can cause clinical problems.

Disorders of fluid or electrolyte homeostasis arise from many different pathophysiologic conditions. In severe cases, these disorders cause death. Although these disorders develop from many specific causes in different patient populations, these specific causes fall into general categories that arise from the principles of normal fluid and electrolyte homeostasis. This chapter first presents the principles of normal fluid homeostasis and then, building on that foundation, continues with a discussion of fluid imbalances. Similarly, it explains the principles of electrolyte homeostasis before presenting plasma electrolyte imbalances.

BODY FLUID HOMEOSTASIS

The term *body fluid*, as used in this chapter, pertains to water within the body and the particles dissolved in it. Body fluid is contained in two major compartments: extracellular (outside the cells) and intracellular (inside the cells). In all age groups except infants, approximately two thirds of body fluid is intracellular. The other one third of body fluid is extracellular. Infants have more extracellular fluid than intracellular fluid; this proportion reverses within a few months as the infant grows. The extracellular fluid lies between the cells (interstitial compartment), in the blood vessels (vascular compartment), in dense connective tissue and bone, and in several minor compartments that are collectively termed the *transcellular fluids* (e.g., synovial, cerebrospinal, and gastrointestinal fluids). The major body fluid compartments are depicted in Fig. 24.1.

The fluids in the various body compartments have different compositions, although their total particle concentration is equal. The intracellular fluid is relatively rich in potassium and magnesium ions, inorganic and organic phosphates, and proteins. It is relatively low in sodium and chloride ions. In contrast, the extracellular fluid in the vascular and interstitial compartments is relatively rich in sodium, chloride, and bicarbonate ions and relatively low in potassium, magnesium, and phosphate ions. The vascular portion of the extracellular fluid contains many proteins, whereas the interstitial and transcellular portions of the extracellular fluid contain very few proteins. Most transcellular fluids are secreted by epithelial cells; their composition varies according to their function.

Total body water is the total amount of water in all fluid compartments. The percentage of body weight that is water varies according to a person's age and proportion of body fat (Fig. 24.2). A full-term newborn infant is about 75% water by weight. (Preterm infants have an even higher percentage of water.) This percentage decreases with age. In a standard adult man, body water is about 60% of body weight. The percentage is less (about 50%) in women because they have a greater proportion of body fat than men of the same weight. In obese adults, with a much larger proportion of body fat, less of the body weight is water. With normal aging there is a relative increase in body fat so that in older men, 50% of the body weight is typically composed of water; in older women, it is even less.

One liter of water weighs 1 kg (2.2 lb). Thus a lean, middle-aged, healthy adult man who weighs 70 kg (154 lb) has approximately 42 L of body water. Of this amount, approximately 25 L is intracellular water. The approximately 17 L of extracellular water is distributed as 3 L of plasma water, 8 L of interstitial and lymph water, 5 L of water trapped in dense connective tissue and bone, and 1 L of transcellular water.

Fluid homeostasis is a dynamic process. This process may be viewed as the net result of four subprocesses: *fluid intake, fluid absorption, fluid distribution,* and *fluid excretion.* In some individuals who have pathophysiologic conditions, loss of fluid through abnormal routes also occurs. The interplay of these subprocesses is fluid homeostasis (Fig. 24.3).

Fluid Intake and Absorption

Fluid intake is entry of fluid into the body by any route. People normally ingest fluids orally, both by drinking and by eating (water contained in food). They also synthesize a small amount of water through cellular metabolism of the foods they eat. Fluid intake by drinking is influenced by habit, social factors, and thirst. Physiologic triggers of thirst include increased osmolality (concentratedness) of extracellular

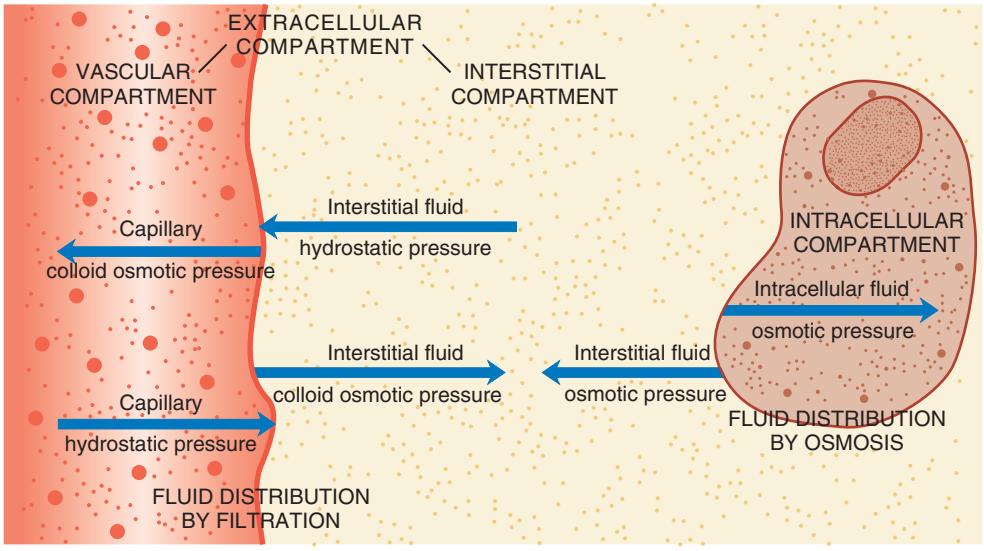


FIG 24.1 Factors that influence body fluid distribution. Fluid distribution between the vascular and interstitial compartments is the net result of filtration across permeable capillaries. The distribution of fluid between the interstitial and intracellular compartments occurs by osmosis rather than by filtration.

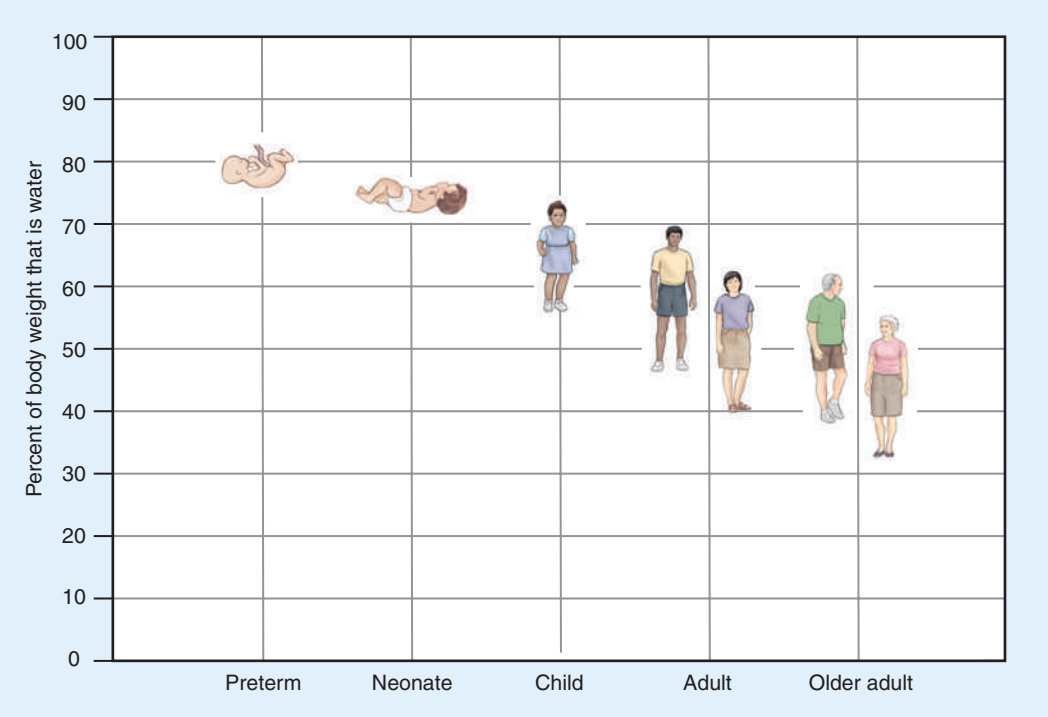


FIG 24.2 Percentage of total body water by age. The percentage of body weight that is water is high in infancy and decreases with increasing age.

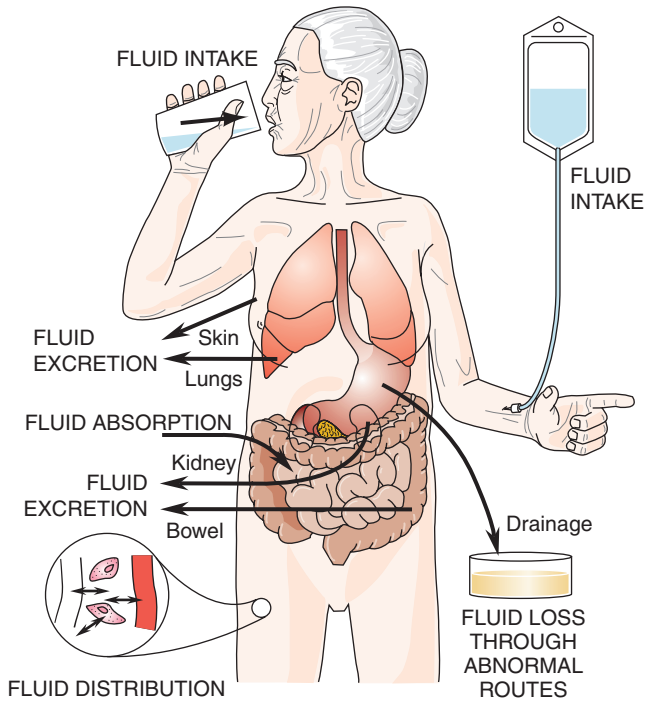


FIG 24.3 Fluid homeostasis. Fluid homeostasis is the interplay of fluid intake and absorption, fluid distribution, fluid excretion, and fluid loss through abnormal routes.

fluid (osmoreceptor-mediated thirst), decreased circulating blood volume (baroreceptor-mediated and angiotensin II–mediated thirst), and dryness of the mucous membranes of the mouth and possibly other visceral signals. In older adults, cerebral osmoreceptor-mediated thirst diminishes; thus older adults who do not have a habit of drinking fluids throughout the day may not have sufficient fluid intake to meet their needs.

Additional routes of fluid intake that may occur in patients who have various pathophysiologic conditions include intravenous intake; intake tubes into the gastrointestinal tract, other body cavities, subcutaneous tissue, or bone marrow; rectal intake (such as tap water enema); and, occasionally, intake through the lungs (such as near-drowning). Health care professionals often control fluid intake by many of these routes.

Unless fluid intake occurs intravenously, the fluid must be absorbed before it reaches the vascular compartment. Fluid absorption from the gastrointestinal tract partially depends on osmotic forces generated by absorption of electrolytes and other particles.

Fluid Distribution

Much of the fluid that reaches the vascular compartment then distributes into other fluid compartments. *Fluid distribution between the vascular and interstitial compartments is the net result of* filtration *across permeable capillaries.* At the capillary level, two forces tend to move fluid from the capillaries into the interstitial compartment: capillary hydrostatic pressure (the outward push of vascular fluid against the capillary walls) and interstitial fluid colloid osmotic pressure (the inward-pulling force of particles in the interstitial fluid). Concurrently, two forces tend to move fluid from the interstitial compartment into the capillaries: capillary colloid osmotic pressure (the inward-pulling force of particles in vascular fluid) and interstitial fluid hydrostatic pressure (the outward push of interstitial fluid against the outside of the capillary walls).

The distribution of fluid between the vascular and interstitial compartments is analogous to two groups of people pushing on opposite sides of a swinging door—the strongest "push" will determine in which direction the door will swing. Thus at any one point along a capillary, the direction and amount of fluid flow between the vascular and interstitial compartments are determined by the net result of opposing forces. These forces are illustrated in Fig. 24.1.

In contrast, *the distribution of fluid between the interstitial and intracellular compartments occurs by* osmosis**,** *rather than by filtration.* Cell membranes contain aquaporins that allow water to pass through.

Cell membranes are permeable to water but not to electrolytes, many of which require specialized transport mechanisms to cross a cell membrane. Thus water can move freely through a cell membrane, but electrolytes and other particles cannot. When there is a difference in particle concentration (osmolality) inside and outside cells because the particles cannot move freely, the water crosses the membrane rapidly to equalize the osmolality. Osmosis occurs until the intracellular osmotic pressure and the interstitial fluid osmotic pressure become equal, as in Fig. 24.1.

The direction of movement of water by osmosis is determined by the particle concentrations on the two sides of the semipermeable cell membrane. If, on the one hand, the particle concentration (osmolality) of the interstitial fluid becomes higher than the particle concentration inside cells, water will move by osmosis from the cells to the interstitial fluid to equalize the osmolality in the two compartments. If, on the other hand, the osmolality of the interstitial fluid becomes lower than the osmolality of the intracellular fluid, then water will move from the interstitial compartment to the intracellular compartment to equalize the osmolality. In this way, changes in the osmolality of the interstitial and intracellular compartments control the distribution of water between them.

Distribution of fluid between the intracellular and transcellular compartments is controlled by processes within the epithelial cells that secrete these fluids.

Fluid Excretion

The fourth component of fluid homeostasis is fluid excretion. Fluid excretion normally occurs through the urinary tract, bowels, lungs, and skin. Fluid is excreted through the skin as visible sweat (which may or may not occur) and as insensible perspiration (which always occurs). Another obligatory route of excretion of water is through the lungs as a person exhales. Fecal excretion of fluid occurs with normal bowel function and increases dramatically in a person who has diarrhea. In most circumstances, the largest volume of fluid is excreted in the urine.

The amount of fluid excreted in the urine is controlled primarily by the hormones antidiuretic hormone (ADH), aldosterone, and natriuretic peptides (NPs; e.g., A-type natriuretic peptide [ANP]), and to a lesser degree by minor hormones such as renal prostaglandins and by the renal sympathetic nerves. ADH is synthesized by cells in the supraoptic and paraventricular nuclei of the hypothalamus. The axons of these cells extend down the median eminence of the pituitary stalk. The release of ADH thus occurs from the posterior pituitary gland. Factors that increase release of ADH into the blood include increased osmolality (concentratedness) of the extracellular fluid, decreased circulating fluid volume, pain, nausea, and physiologic and psychological stressors. The hormone circulates to the distal tubules and collecting ducts in the kidneys where, consistent with its name, ADH causes reabsorption of water that dilutes the blood and other body fluids. Reabsorption of water decreases the urine volume and makes the urine concentrated, thus decreasing fluid excretion. Factors that decrease ADH release (such as decreased osmolality of the extracellular fluid and ethanol intake) allow a large, dilute urine volume.

Aldosterone is another hormone that influences urine volume. Aldosterone is synthesized and secreted by cells in the adrenal cortex. The major stimuli for its release are angiotensin II (from the renin–angiotensin system, which is activated by decreased circulating blood volume) and an increased concentration of potassium ions in the plasma. Aldosterone causes the renal tubules to reabsorb sodium and water (saline), which expands the extracellular fluid volume. This renal action decreases fluid excretion, although by a different mechanism than ADH. When more aldosterone is secreted, the urine volume is smaller; decreased secretion of aldosterone causes a larger urine volume.

A comparison of ADH and aldosterone is useful to remember their actions. ADH is the "tap water" hormone. It causes the kidneys to reabsorb plain water. Renal reabsorption of water caused by ADH makes a smaller volume of more concentrated urine and dilutes body fluids. Aldosterone is the salt water hormone. It causes the kidneys to reabsorb sodium and water. Renal reabsorption of sodium and water caused by aldosterone makes a smaller volume of urine and expands extracellular fluid volume.

ANP normally is stored in granules in the cardiac atrial cells and released when the atria are stretched. B-type natriuretic peptide (BNP) is synthesized and released from ventricular cells when ventricular diastolic pressure increases abnormally, as in heart failure. ANP and BNP cause natriuresis (sodium excretion in the urine), which is accompanied by water excretion. Thus these NPs promote fluid excretion in the urine. When the vascular volume increases, the heart is stretched and more NPs are released to cause renal excretion of the excess fluid. When the vascular volume is decreased, the heart is less stretched; therefore fewer NPs are released and the kidneys excrete less fluid. NPs oppose the action of aldosterone, but they are not as strong as aldosterone.

The urine volume that an individual produces also is highly dependent on having adequate blood pressure to perfuse the kidneys and on the glomerular filtration rate. Thus renal excretion of fluid is the end result of several factors, including hormones that respond to different stimuli and have different actions on the renal tubules.

Fluid Loss Through Abnormal Routes

People who have pathophysiologic conditions often experience loss of fluid through abnormal routes. Examples of these routes are emesis; tubes in the gastrointestinal tract or other body cavities; hemorrhage; drainage from fistulas, wounds, or open areas of skin; and paracentesis. Fluid lost through abnormal routes may be a significant factor in disturbing fluid homeostasis.

If the body's physiologic mechanisms are functioning well, the processes of fluid homeostasis maintain normal body fluid status. If fluid intake is large, fluid excretion increases by the mechanisms described previously that increase urine volume (large volume of dilute urine). If fluid intake is diminished or if fluid is lost through abnormal routes, fluid excretion decreases (small volume of concentrated urine), and thirst may cause an increase in fluid intake.

If pathophysiologic processes interfere with normal fluid homeostasis or if the normal processes become overwhelmed, then fluid imbalances result. For example, a person who has a pathophysiologic process that prevents the kidneys from excreting much fluid may accumulate too much fluid unless the fluid intake is reduced. The opposite problem will occur in a person whose fluid intake is too small to replace a large amount of fluid excreted or lost through abnormal routes.

KEY POINTS

- Fluid homeostasis is a dynamic process that includes fluid intake and absorption, fluid distribution across body fluid compartments, fluid excretion, and, in some individuals, fluid loss through abnormal routes.
- Habit and thirst are important regulators of fluid intake. Individuals who are unable to control their own fluid intake (such as those receiving fluids intravenously and immobile or unconscious patients) are at high risk for fluid imbalance.
- Fluid (water and small particles) moves back and forth between the vascular and interstitial areas by filtration at the capillaries. Capillary hydrostatic pressure is the primary force promoting fluid movement from the capillaries to the interstitial fluid. Plasma colloid osmotic pressure is the primary force that causes interstitial fluid to move back into the capillaries.

- Water moves in and out of cells by osmosis. Electrolytes do not move freely across cell membranes but are transported by membrane channels and carriers.
- Fluid excretion may be visible (urine, feces, sweat) or invisible (respiration and insensible perspiration). Fluid loss may occur through abnormal routes such as emesis and wound drainage. To maintain fluid balance, fluid intake must counterbalance fluid excretion and loss through abnormal routes.
- Healthy kidneys adjust fluid excretion in response to blood pressure and several hormones. Aldosterone induces the kidneys to conserve saline (salt and water), which expands the extracellular fluid volume; NPs (ANP and BNP) promote saline excretion. Antidiuretic hormone (ADH) causes the kidneys to retain water, thus concentrating the urine and diluting the body fluids. Urine volume and concentration are important indicators of body fluid balance.

FLUID IMBALANCES

If fluid homeostasis is disturbed by pathophysiologic processes or other factors (such as medications), fluid imbalances may occur. Fluid imbalances fall into two major categories: imbalances of extracellular fluid volume (saline imbalances) and imbalances of body fluid concentration (water imbalances).

Extracellular Fluid Volume

In some circumstances, individuals have too much or too little extracellular fluid. These disorders are called extracellular fluid volume (ECV) imbalances because they involve a change in the *amount* (volume) of the extracellular fluid. These disorders also are termed *saline imbalances* because they are disorders of isotonic salt water. (Isotonic saline is salt water in the same concentration as the normal plasma concentration.) In an ECV imbalance, the *concentration* of the extracellular fluid is normal; there is simply too much or too little of it. Some individuals have an ECV imbalance and an imbalance of body fluid concentration at the same time. In this case both the volume and serum sodium concentration of the extracellular fluid are abnormal. This section discusses only the isotonic volume imbalances; the concentration imbalances are discussed separately because they may occur separately.

Volume Deficit

ECV deficit is caused by removal of a sodium-containing fluid from the body. It is a decrease in saline (isotonic salt water) in the same concentration as the normal extracellular fluid, which is why the condition sometimes is termed *saline deficit*. In an uncomplicated ECV deficit, the serum sodium concentration is normal. The *concentration* of the extracellular fluid is normal; the *amount* of the extracellular fluid is abnormally decreased (Fig. 24.4).

Etiology. Specific causes of ECV deficit are listed in Box 24.1. All causes involve removal of a sodium-containing fluid from the extracellular compartment. The sodium-containing fluid usually is removed from the body; however, it may be sequestered in a "third space" in the body that is outside the extracellular compartment. For example, ascites (fluid in the peritoneal cavity) that develops rapidly may deplete the ECV. Another example is fluid that accumulates rapidly in the bowel during an acute intestinal obstruction. Although the fluid in these examples remains in the body, it no longer is part of the extracellular fluid, and signs and symptoms of ECV deficit occur.

Clinical manifestations. Signs and symptoms of ECV deficit are the result of decreased fluid volume in the vascular and interstitial areas. These clinical manifestations include sudden weight loss, postural blood pressure decrease with concurrent increased heart rate, flat neck veins (or veins collapsing with inspiration) when a patient is supine, prolonged small-vein filling time, prolonged capillary refill time, lightheadedness, dizziness, syncope, and oliguria. If the kidneys are responding normally, the small volume of urine will be concentrated (and thus quite yellow). An ECV deficit that develops slowly also may be manifested by decreased skin turgor (skin tenting when it is pinched up over the sternum), dryness of oral mucous membranes between cheek and gum, hard stools, soft sunken eyeballs, longitudinal furrows in the tongue, and absence of tears and sweat. An infant who develops ECV deficit has a sunken fontanel; neck veins are not reliably assessed in infants.

Sudden weight loss is a sensitive measure of ECV deficit. One liter of saline weighs 1 kg; therefore a person who loses 1 kg in 24 hours has excreted 1 L of fluid or lost it through an abnormal route. It is not possible to lose a kilogram of fat overnight; a sudden weight loss of this magnitude results only from fluid loss, if the body weight is measured

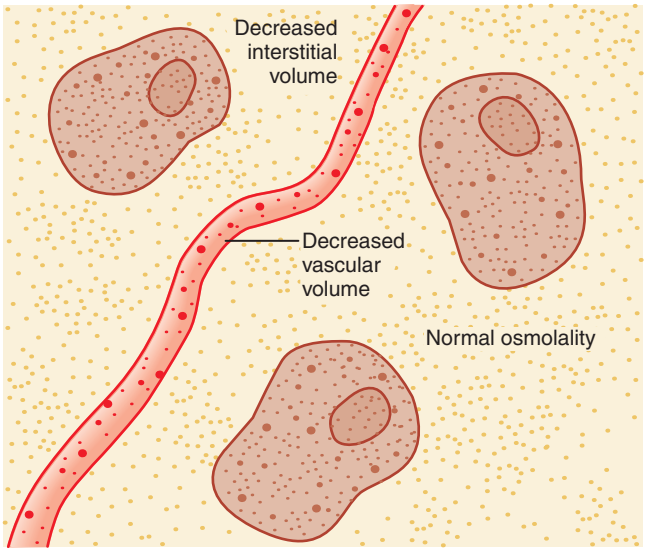


FIG 24.4 Extracellular fluid volume deficit. Decreased volume of extracellular fluid in vascular and interstitial compartments is characteristic of extracellular fluid volume deficit. Intracellular volume is unchanged.

BOX 24.1 Causes of Extracellular Fluid Volume Deficit

Gastrointestinal Excretion or Loss of Sodium-Containing Fluid
Emesis
Diarrhea (includes laxative overuse or abuse)
Gastric suction or intestinal decompression
Fistula drainage

Renal Excretion of Sodium-Containing Fluid
Adrenal insufficiency
Salt-wasting renal disorders
Extensive diuretic use
Bed rest

Other Loss of a Sodium-Containing Fluid
Hemorrhage
Massive diaphoresis
Third-space fluid accumulation
Paracentesis and similar procedures
Burns

accurately. An ECV deficit may occur without a weight loss if fluid is sequestered in a third space somewhere in the body, as with ascites or intestinal obstruction.

A postural blood pressure decrease with concurrent increased heart rate that is measured when a previously supine person stands or sits with legs dependent is a good indicator of fluid volume depletion in the vascular compartment. Severe ECV deficit may lead to hypovolemic shock, which can be fatal if not treated effectively with fluid replacement.

Volume Excess

ECV excess is essentially the opposite of an ECV deficit. It is the condition in which the amount of extracellular fluid is abnormally increased. Both the vascular and the interstitial areas have too much isotonic fluid (Fig. 24.5). In an uncomplicated ECV excess, the *concentration* of the extracellular fluid is normal, but an excessive *amount* of that fluid is present.

Etiology. ECV excess is caused by addition or retention of saline (salt water in the same concentration as normal plasma). For this reason, it sometimes is termed *saline excess*. As mentioned previously, the hormone aldosterone causes the kidneys to retain saline. ECV excess therefore may be caused by conditions that involve excessive aldosterone secretion. For example, increased aldosterone secretion is a compensatory mechanism that commonly accompanies chronic heart failure and eventually leads to ECV excess. Additional causes of ECV excess are presented in Box 24.2.

Clinical manifestations. Signs and symptoms of ECV excess are sudden weight gain, edema, and manifestations of circulatory overload: bounding pulse, neck vein distention in a person in the upright position, crackles in the dependent portions of the lungs, dyspnea, orthopnea, and even the frothy sputum of pulmonary edema. An infant who develops ECV excess has a bulging fontanel; assessment of neck veins is not effective in infants.

Sudden weight gain is a sensitive measure of ECV excess. It is impossible to gain a kilogram of fat overnight; such a sudden weight gain is an accumulation of saline. People who eat salty food in a restaurant weigh more the next day because the water they drank combined with the salt in the food to make isotonic saline. The isotonic saline expands the extracellular fluid, causing a mild saline excess until it is excreted by the kidneys. This is the reason that low-sodium diets are prescribed for people who have pathophysiologic processes that cause saline excess (e.g., compensated heart failure).

Body Fluid Concentration

In contrast to the ECV disorders just discussed, imbalances of body fluid concentration are disorders of the *concentration* rather than of the *amount* of extracellular fluid. Body fluid concentration disorders also are called water imbalances. The serum sodium concentration reflects the osmolality (concentratedness) of the blood. Imbalances of body fluid concentration are recognized by abnormal serum sodium concentration. The normal serum sodium concentration is 135 to 145 mEq/L (may vary slightly with different laboratories). Many individuals develop imbalances of both ECV and serum sodium concentration at the same time. Isolated imbalances of serum sodium concentration may also occur. This section discusses the concentration imbalances separately.

Hyponatremia

Natrium is the Latin word for sodium. A serum sodium concentration below the lower limit of normal indicates hyponatremia. When

BOX 24.2 Causes of Extracellular Fluid Volume Excess

Excessive Intravenous Infusion of Sodium-Containing Isotonic Solutions
- Normal saline (0.9% sodium chloride)
- Ringer infusion
- Lactated Ringer infusion

Renal Retention of Sodium and Water
- Hyperaldosteronism
- Chronic heart failure
- Cirrhosis
- Acute glomerulonephritis
- Chronic end-stage renal disease
- Cushing disease
- Corticosteroid therapy

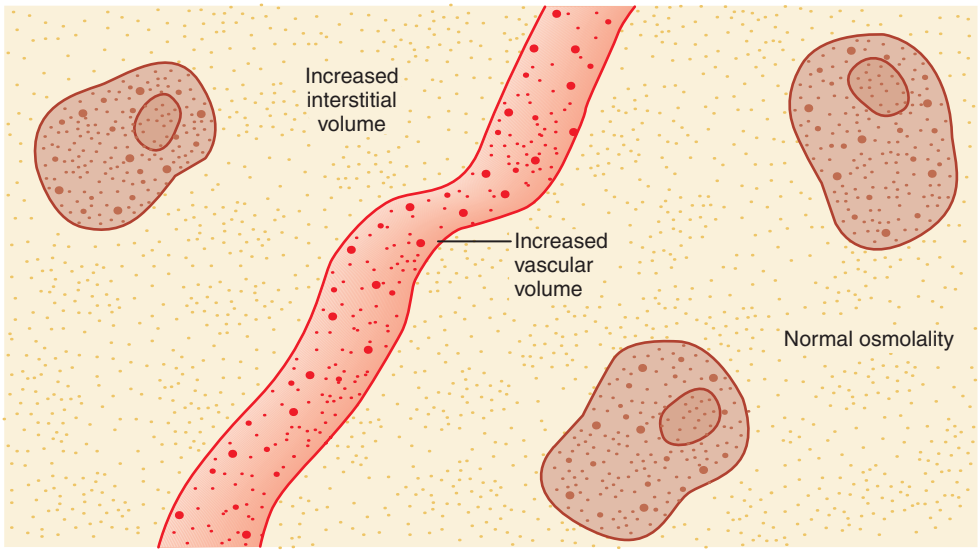


FIG 24.5 Extracellular fluid volume excess. Increased volume of extracellular fluid in vascular and interstitial compartments is characteristic of extracellular fluid volume excess. Intracellular volume is unchanged.

hyponatremia is present, the extracellular fluid contains relatively too much water for the amount of sodium ions present; it is more dilute than normal.

Etiology. Hyponatremia is caused by factors that produce a relative excess of water in proportion to salt in the extracellular fluid. Because the serum sodium concentration reflects the osmolality of the blood, the reduced serum sodium concentration of hyponatremia indicates that the extracellular fluid has a reduced osmolality; it is too dilute. Hyponatremia also is called *hypotonic syndrome, hypoosmolality,* and *water intoxication.* All of these terms reflect the abnormally dilute concentration of the extracellular fluid that results when the normal proportion of salt to water in the extracellular fluid is disrupted by gaining more water than salt or losing relatively more salt than water.

A *gain of relatively more water than salt* will cause hyponatremia that is known as *dilutional hyponatremia.* As mentioned previously, the hormone ADH causes the kidneys to retain water (not sodium and water) in the body. This hormone is part of the system that normally regulates the osmolality of extracellular fluid. However, circumstances that cause prolonged or excessive release of ADH cause the kidneys to retain too much water, which effectively dilutes the blood; hyponatremia is the result. ADH secretion is excessive in the syndrome of inappropriate secretion of ADH. ADH also may be produced ectopically. For example, small cell (oat cell) carcinoma is a type of lung tumor that frequently synthesizes and releases ADH. This ectopic production of ADH from a tumor is not subject to the feedback inhibition of normal ADH release, so inappropriate amounts are released. With continually high levels of ADH being produced by the tumor, the kidneys retain excessive amounts of water—a gain of water relative to salt. Pain, nausea, and other physical and psychological stressors also increase ADH release, which can be significant in hospitalized patients, especially in the postoperative period. Although it is difficult to drink enough water to cause hyponatremia, water intake that exceeds renal excretory capacity is seen in some situations. For example, the hyponatremia of beer potomania arises when people habitually drink large amounts of beer daily with very little food intake. Factors that cause hyponatremia by gain of water relative to salt are presented in Box 24.3.

Hyponatremia also may be caused by a *loss of relatively more salt than water,* in which case it is known as *depletional hyponatremia.* If salt is removed from the body while water remains, then the extracellular fluid once again will become too dilute; hyponatremia results. The action of thiazide diuretics can cause this to occur, especially in older women. Factors that cause hyponatremia by loss of salt relative to water also are presented in Box 24.3. Although Box 24.3 separates causes of hyponatremia into two categories, some types are due to simultaneous gain of water and loss of salt. For example, hyponatremia in marathon runners and triathletes occurs from loss of salt through heavy sweating and gain of water from excessive water intake plus inappropriate renal water reabsorption caused by increased ADH secretion.

Clinical manifestations. Clinical manifestations of hyponatremia are nonspecific manifestations of central nervous system dysfunction. They vary from malaise, anorexia, nausea, vomiting, and headache to confusion, lethargy, seizures, and coma. Profound hyponatremia causes fatal cerebral herniation. The signs and symptoms are caused by swelling of neurons and glial cells as a result of the decreased osmolality of extracellular fluid. When the extracellular fluid becomes too dilute, the intracellular fluid initially is more concentrated. Therefore water moves into cells by osmosis (Fig. 24.6). The severity of the signs and symptoms depends on how rapidly hyponatremia develops as well as on the absolute value of the serum sodium concentration. A rapid decrease in osmolality produces more severe manifestations than a slow decline, other factors being equal.

Hypernatremia

Hypernatremia is a serum sodium concentration above the upper limit of normal (145 mEq/L). When hypernatremia is present, the extracellular fluid contains relatively too little water for the amount of sodium ions present; it is too concentrated. Hypernatremia also is called *water deficit, hypertonic syndrome,* and *hyperosmolality.* These terms all reflect the relative deficit of water to salt in the extracellular fluid that occurs in hypernatremia.

Etiology. Hypernatremia is caused by a *gain of relatively more salt* than water or by a *loss of relatively more water than salt.* Both of these processes cause the body fluids to become too concentrated. Patients who receive concentrated tube feedings without enough water, especially older adults, are at high risk for hypernatremia because they gain relatively more solute than water, which causes an obligatory loss of relatively more water than salt in the urine. Hypernatremia can be prevented in these individuals by administering water between feedings. Other specific

BOX 24.3 Causes of Hyponatremia

Gain of Relatively More Water Than Salt (Dilutional Hyponatremia)

Excessive antidiuretic hormone
Excessive intravenous infusion of 5% dextrose in water (D_5W)
Hypotonic irrigating solutions
Tap water enemas
Psychogenic polydipsia (compulsive water drinking)
Forced excessive water ingestion (child abuse or club initiation)
Excessive beer ingestion (beer potomania)
Near-drowning in fresh water
Selective serotonin reuptake inhibitors (SSRIs)

Loss of Relatively More Salt Than Water (Depletional Hyponatremia)

Diuretics, especially thiazides
Salt-wasting renal disease
Replacement of water, but not salt, lost through emesis, diarrhea, gastric suction, diaphoresis, or burns

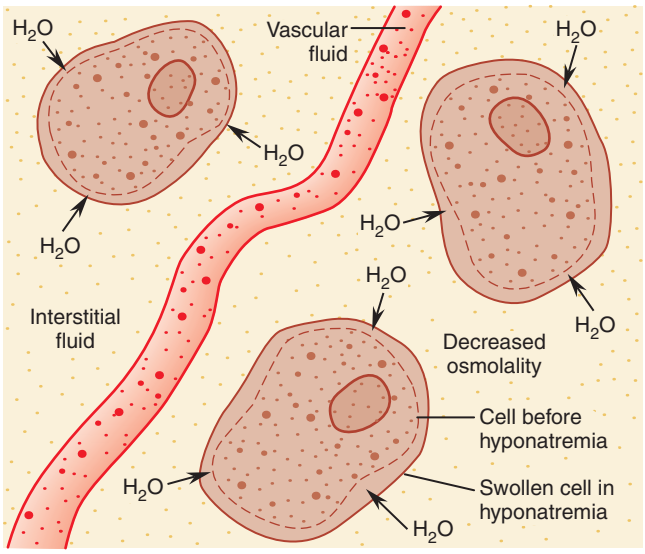


FIG 24.6 Cell swelling in hyponatremia. Decreased osmolality (concentration) of extracellular fluid in hyponatremia causes water to move into cells by osmosis.

factors that cause hypernatremia are presented in Box 24.4 under the two major categories.

Clinical manifestations. Signs and symptoms of hypernatremia are similar to those of hyponatremia in that they are nonspecific manifestations of central nervous system dysfunction. In hypernatremia, the increased osmolality of the extracellular fluid causes neurons and glial cells to shrivel because water moves from the cells to the interstitial fluid by osmosis (Fig. 24.7). The dysfunction ranges from confusion and lethargy to seizures and coma. Thirst and oliguria (except for hypernatremia of renal origin) are common. Severe hypernatremia may cause death.

Both Volume and Concentration

Clinical Dehydration

Clinical dehydration is a combination of two fluid disorders: ECV deficit and hypernatremia. A person who has clinical dehydration has too small a volume of fluid in the extracellular compartment (vascular and interstitial) and the body fluids are too concentrated (Fig. 24.8).

Etiology. Clinical dehydration occurs commonly in individuals who have vomiting and diarrhea and do not know how (or are unable) to replace the salt and the water that is exiting the body. Fluid excreted in diarrhea and lost by vomiting, plus the normal daily respiratory, skin, and urine excretion, is the equivalent of hypotonic sodium–containing fluid (isotonic saline with extra water added). Removal of the saline portion of this fluid from the body causes ECV deficit, and removal of the extra water from the body causes hypernatremia. The combination of these two imbalances is clinical dehydration.

Clinical manifestations. Signs and symptoms of clinical dehydration are the combination of the signs and symptoms of the two separate disorders. Therefore a person who is clinically dehydrated will have clinical manifestations as listed in Box 24.5. Infants and older adults are at highest risk for clinical dehydration, although it can occur at any age.

BOX 24.4 Causes of Hypernatremia

Gain of Relatively More Salt Than Water

Tube feeding
Intravenous infusion of hypertonic saline
Near-drowning in salt water
Overuse of salt tablets
Food intake with reduced fluid intake
Difficulty swallowing fluids
No access to water
Inability to respond to thirst

Loss of Relatively More Water Than Salt

Diabetes insipidus (deficient antidiuretic hormone)
Tube feeding (causes obligate water loss in urine)
Osmotic diuresis
Prolonged emesis, diarrhea, or diaphoresis without water replacement

Interstitial Fluid Volume

Edema

Edema is an excess of fluid in the interstitial compartment. It may be a manifestation of ECV excess, or it may arise from other mechanisms. Forces that determine the distribution of fluid between the vascular and interstitial compartments are described previously in this chapter (see the "Fluid Distribution" section). An increase in the forces that tend to move fluid from the capillaries into the interstitial compartment or a decrease in forces that tend to move fluid from the interstitial compartment into the capillaries will cause edema by altering normal fluid distribution between the vascular and interstitial compartments. Thus edema may arise from increased capillary hydrostatic pressure, increased interstitial fluid colloid osmotic pressure, blockage of lymphatic drainage, or decreased capillary colloid osmotic pressure (Fig. 24.9). Edema may be localized or generalized (existing in many areas of the body simultaneously).

Increased capillary hydrostatic pressure is caused by increased ECV, by the increased local capillary flow that accompanies inflammation, and by venous congestion. *Increased interstitial fluid colloid osmotic pressure* occurs when inflammation increases vascular permeability and proteins leak into the interstitial fluid. Lymphatic drainage normally removes minute amounts of protein that enter the interstitial fluid. *Blockage of lymphatic drainage* (e.g., by a tumor, parasites, fibrosis from radiation therapy, or surgical removal of lymph nodes) also causes edema when the interstitial accumulation of protein increases interstitial fluid colloid osmotic pressure. This type of edema is called *lymphedema*

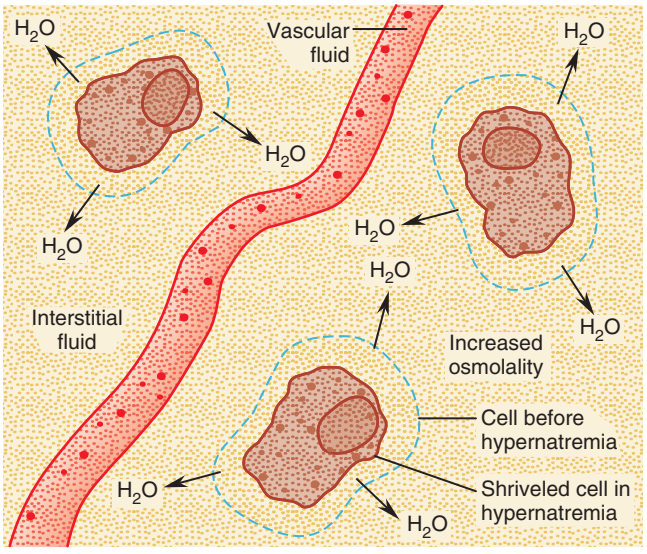


FIG 24.7 Cell shriveling in hypernatremia. Increased osmolality (concentration) of extracellular fluid in hypernatremia causes water to move from cells by osmosis.

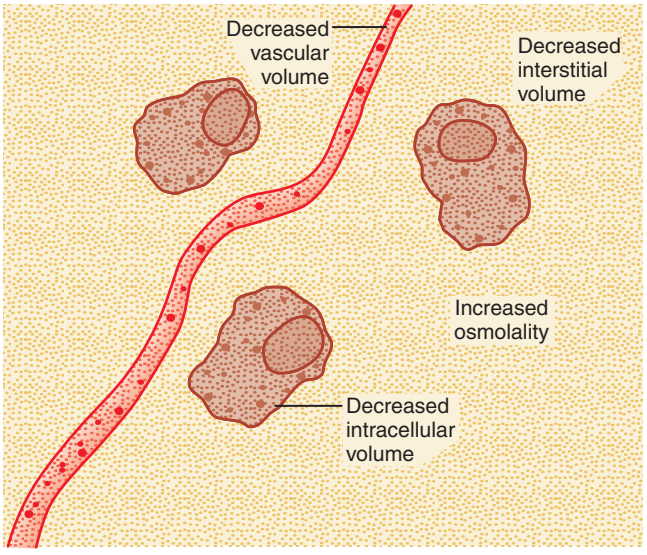


FIG 24.8 Clinical dehydration. Decreased volume of extracellular fluid in vascular and interstitial compartments plus cell shriveling from increased osmolality of extracellular fluid are combined in clinical dehydration.

TABLE 24.1 Normal Serum Electrolyte Concentrations

Electrolyte	Normal Concentration Range
Calcium (total)	9–11 mg/dL (4.5–5.5 mEq/L)
Magnesium	1.5–2.5 mEq/L
Phosphate	2.5–4.5 mg/dL (adults and older children)
	4.5–6.5 mg/dL (children)
	4.3–9.3 mg/dL (neonates)
Potassium	3.5–5.0 mEq/L
	3.9–5.9 mEq/L (neonates)
Sodium	135–145 mEq/L
	135–162 mEq/L (neonates)

BOX 24.5 Signs and Symptoms of Clinical Dehydration

- Sudden weight loss
- Postural blood pressure decrease with concurrent increased heart rate
- Lightheadedness, dizziness, or syncope upon standing
- Flat neck veins when supine or neck veins that collapse during inspiration (older children and adults)
- Sunken fontanel (infants)
- Rapid, thready pulse
- Prolonged small-vein filling time
- Prolonged capillary refill time
- Oliguria
- Decreased skin turgor
- Dryness of oral mucous membranes
- Absence of sweat and tears
- Hard stools
- Soft, sunken eyeballs
- Longitudinal furrows in the tongue
- Thirst
- Increased serum sodium concentration
- Confusion, lethargy
- Coma
- Hypovolemic shock

and may be persistent. Edema caused by increased interstitial fluid colloid osmotic pressure or blockage of lymphatic drainage frequently is localized. *Decreased capillary colloid osmotic pressure* occurs when the concentration of plasma proteins is decreased, as in malnutrition or liver disease (decreased protein synthesis). Edema from this cause usually is generalized.

In summary, edema represents increased interstitial fluid volume, a condition that may be local or generalized. Edema may be a sign of ECV excess (which causes increased capillary hydrostatic pressure), or it may be caused by other factors that alter the distribution of fluid between the vascular and interstitial compartments.

KEY POINTS

- Extracellular fluid volume (ECV) deficit (saline deficit) occurs when sodium-containing fluids are lost from the body (e.g., emesis). It is an abnormally reduced volume of the vascular and interstitial fluids. ECV deficit is characterized by normal serum sodium concentration and manifestations of volume deficit (weight loss, poor skin turgor, postural hypotension, oliguria).
- ECV excess (saline excess) commonly is due to processes that cause the kidneys to retain sodium and water. It is an abnormally increased volume of the vascular and interstitial fluids. ECV excess is characterized by a normal serum sodium concentration and manifestations of volume excess (weight gain, peripheral edema, distended neck veins, dyspnea).
- Hyponatremia is associated with excessive antidiuretic hormone (ADH) secretion or hypotonic fluid intake. It is characterized by a low serum sodium concentration, which indicates that body fluids are abnormally dilute. Clinical manifestations (confusion, lethargy, seizure, coma) occur because of brain cell swelling.
- Hypernatremia is associated with inadequate water intake or excessive water excretion or loss. It is characterized by a high serum sodium level, which indicates that body fluids are too concentrated. Clinical manifestations (confusion, lethargy, seizure, coma) occur because of brain cell shriveling.
- Clinical dehydration occurs commonly in individuals who have gastroenteritis or other conditions that remove hypotonic sodium–containing fluids from the body. It is the combination of ECV deficit and hypernatremia. The clinical manifestations are those of both fluid disorders.
- Edema occurs when there is too much fluid in the interstitial compartment. It may be localized or generalized. The causes of edema at the capillary level are increased capillary hydrostatic pressure, increased interstitial fluid colloid osmotic pressure, blockage of lymphatic drainage, and decreased capillary colloid osmotic pressure.

PRINCIPLES OF ELECTROLYTE HOMEOSTASIS

Electrolytes are ionized salts dissolved in water. The most clinically important electrolytes are the ions sodium, potassium, calcium, magnesium, chloride, bicarbonate, and phosphate. Although sodium ions are electrolytes, serum sodium imbalances are osmolality (concentration) imbalances, as explained previously in this chapter. This section discusses homeostasis and imbalances of potassium, calcium, magnesium, and phosphate ions. Bicarbonate is discussed in Chapter 25 because it is important in acid–base balance and imbalances.

The concentration of an electrolyte in the plasma is different from its concentration inside cells. For normal body function, the electrolyte concentration must be normal in both areas. In clinical situations, the plasma (or serum) concentration of an electrolyte is measured. Normal serum electrolyte concentrations are listed in Table 24.1. The concentration of an electrolyte in the plasma is the net result of four processes: *electrolyte intake, electrolyte absorption, electrolyte distribution,* and *electrolyte excretion.* These processes work together in a dynamic fashion to maintain electrolyte concentrations within their normal limits (Fig. 24.10). Thus if intake of a specific electrolyte increases, excretion of that electrolyte also may increase and normalize the plasma levels. Similarly, if electrolyte intake decreases dramatically, electrolytes may be redistributed into the plasma to maintain the normal plasma concentration.

Electrolyte Intake and Absorption

Electrolyte intake normally occurs orally through food and drink. It is important to remember that oral medications (e.g., magnesium antacids) also may be an important source of electrolyte intake. Intravenous fluids and nutritional solutions are common sources of parenteral intake of electrolytes. Blood transfusions may provide significant amounts of electrolytes. Less common, but important if it occurs, is intramuscular injection of the electrolyte magnesium.

Some patients have electrolyte intake through tubes into body cavities. The most obvious examples are nasogastric and gastrointestinal feeding tubes, but more unusual situations may cause significant electrolyte intake in specific individuals (e.g., irrigation of the renal pelvis with magnesium-rich solutions). Rarely, electrolyte intake may occur through such unusual routes as the lungs (e.g., near-drowning in salt water,

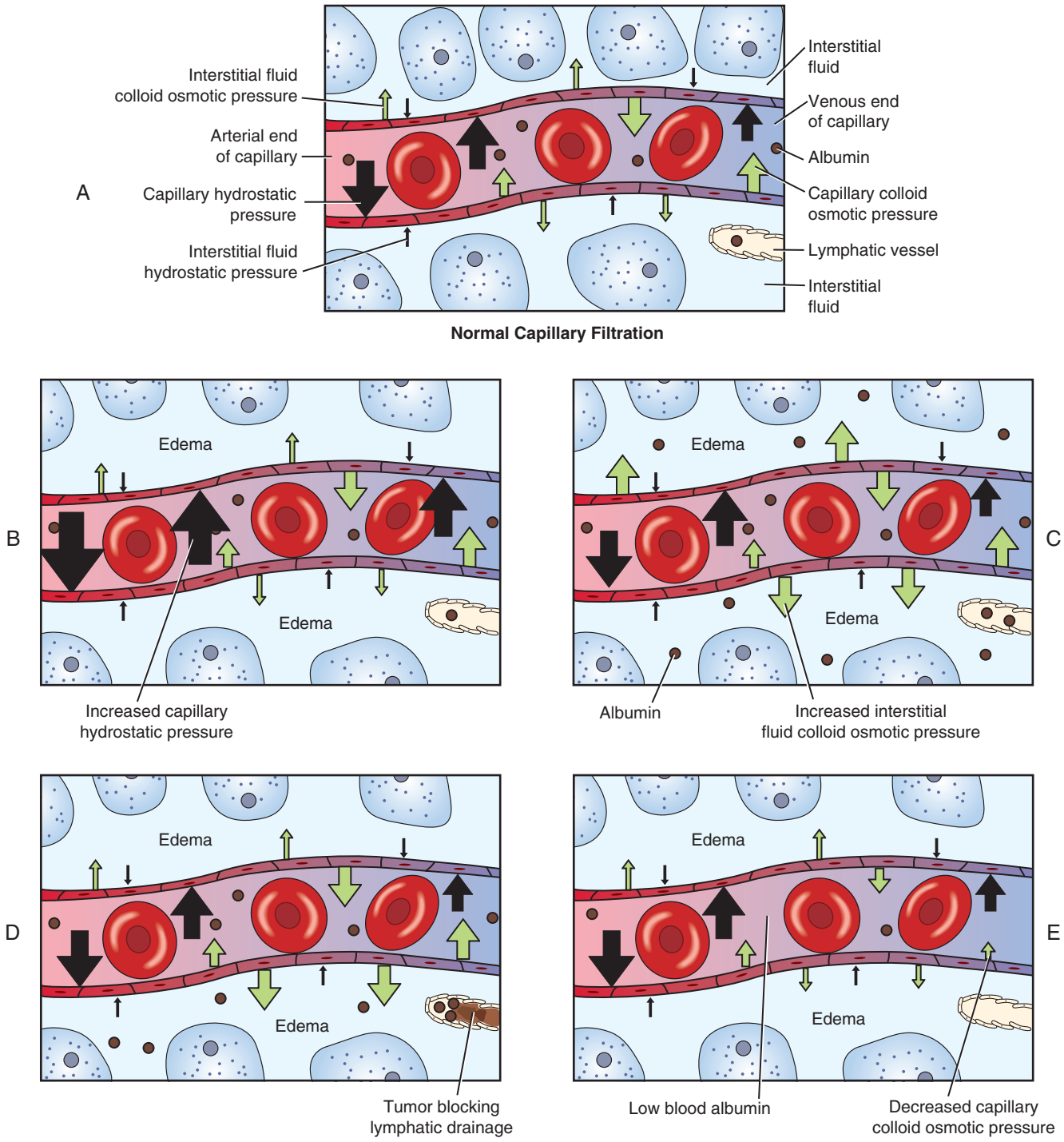


FIG 24.9 Causes of edema. **A,** Normal capillary filtration without edema. **B,** Edema caused by increased capillary hydrostatic pressure. **C,** Edema caused by increased interstitial fluid colloid osmotic pressure from increased capillary permeability. **D,** Edema caused by blocked lymphatic drainage. **E,** Edema caused by decreased capillary colloid osmotic pressure from hypoalbuminemia.

which is rich in magnesium) or the skin (e.g., through application of ointments to large areas of broken or burned skin). Electrolyte intake is controlled by the individual and by health care providers.

If electrolyte intake occurs orally, the electrolyte must be absorbed before it is physiologically useful. Absorption of some electrolytes, such as potassium ions, depends on concentration gradients. Absorption of other electrolytes, such as calcium, depends on the availability of binding proteins, which is influenced by the activity of vitamin D. The contents of the gastrointestinal tract may influence electrolyte absorption. Many agents bind electrolytes and prevent them from being absorbed. For example, undigested fat in the intestines binds calcium and magnesium ions contained in food and prevents them from being absorbed. The pH of intestinal contents also influences the absorption of certain electrolytes, especially calcium ions. Medications often alter electrolyte absorption. Surgical removal of portions of the gastrointestinal tract can decrease electrolyte absorption.

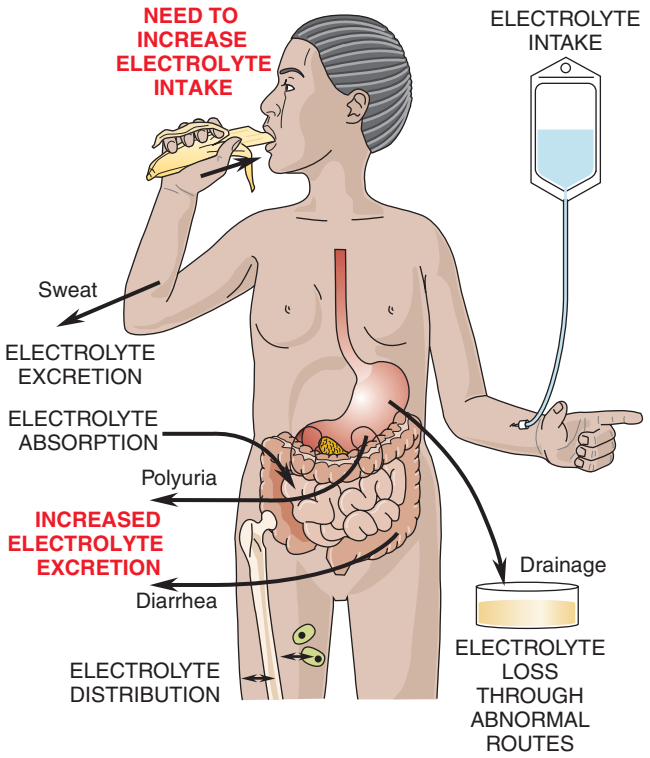


FIG 24.10 Electrolyte homeostasis. Electrolyte homeostasis is the interplay of electrolyte intake and absorption, electrolyte distribution, electrolyte excretion, and electrolyte loss through abnormal routes. If electrolyte excretion or loss through abnormal routes increases, electrolyte intake also must increase to prevent electrolyte imbalance.

Electrolyte Distribution

Every fluid compartment contains electrolytes. However, the electrolyte composition differs in these various compartments. The concentrations of potassium, magnesium, and phosphate ions are higher inside cells than in the fluid outside the cells. Although total calcium ion content concentration is higher inside cells, much of the intracellular calcium is bound to other molecules; the concentration of physiologically active ionized calcium ions is higher in the extracellular fluid. The bones serve as an important reservoir of calcium, magnesium, and phosphate ions. The cells and the bones are often called the *electrolyte pools.*

Distribution of electrolytes between the extracellular fluid and the electrolyte pools is influenced primarily by hormones such as epinephrine (potassium ions), insulin (potassium and phosphate ions), and parathyroid hormone (PTH) (calcium ions). Certain medications also influence electrolyte distribution. Significant movement of electrolytes between the cells and the extracellular fluid may occur within minutes. In the absence of changes in electrolyte intake and excretion, a shift of electrolytes from the extracellular fluid into the electrolyte pools will decrease the plasma electrolyte concentration. Conversely, a shift of electrolyte from an electrolyte pool into the extracellular fluid will increase the plasma electrolyte concentration.

Electrolyte Excretion

Electrolyte excretion occurs through urine, feces, and sweat. Urinary excretion of some electrolytes is influenced by hormones (e.g., aldosterone increases potassium ion excretion), although factors such as the flow rate of renal tubular fluid are also influential. Many different medications alter the rate of urinary excretion of electrolytes. For example, commonly used drugs that increase urinary excretion of potassium include corticosteroids, such as prednisone, and potassium-wasting diuretics, such as furosemide and hydrochlorothiazide. Drugs that increase urinary magnesium excretion include diuretics and aminoglycoside antibiotics, such as gentamicin. Conversely, thiazide diuretics, such as hydrochlorothiazide, decrease urinary excretion of calcium.

Fecal excretion of electrolytes is influenced by the type of feces produced. Diarrhea increases the excretion of potassium and magnesium ions in particular. The composition of the feces also influences the amount of electrolyte excretion. Undigested fat in the intestines binds calcium and magnesium ions that are secreted into the gastrointestinal tract and prevents them from being reabsorbed. Thus these electrolytes are excreted in the feces.

Electrolyte Loss Through Abnormal Routes

When electrolytes exit the body through routes other than the normal urine, feces, and sweat, this may be termed *electrolyte loss through abnormal routes.* This factor alters electrolyte homeostasis in patients who have diverse pathophysiologic conditions. Examples of electrolyte loss through abnormal routes are emesis, nasogastric suction, paracentesis, hemodialysis, wound drainage, and fistula drainage. Loss of electrolytes through abnormal routes may be uncontrollable or may result from therapeutic procedures.

Electrolyte homeostasis is a dynamic interplay between the processes of electrolyte intake, electrolyte absorption, electrolyte distribution, and electrolyte excretion. In some people, electrolyte loss through abnormal routes becomes an important factor that requires adjustment of electrolyte intake and/or electrolyte excretion to prevent development of electrolyte imbalances. Individuals who have acute or chronic illnesses have many factors that tend to cause electrolyte imbalances by disrupting or interfering with electrolyte intake, absorption, distribution, or excretion. As a result, they may develop single or multiple electrolyte imbalances.

KEY POINTS

- The electrolyte composition of the body is maintained by a careful balance of electrolyte intake, absorption, distribution, and excretion. Electrolyte imbalances result from disruption of one or more of these processes or from electrolyte loss through abnormal routes.
- The plasma concentration of an electrolyte may not reflect the intracellular concentration. Cells contain higher concentrations of potassium, magnesium, and phosphate ions, whereas the extracellular fluid contains higher concentrations of sodium, chloride, calcium, and bicarbonate ions.

ELECTROLYTE IMBALANCES

Electrolyte imbalances are widespread in many pathophysiologic conditions. An electrolyte imbalance may be a total body imbalance, or it may be an imbalance in the distribution of electrolytes within compartments, with the total body amount remaining normal. Based on the principles of electrolyte homeostasis explained in the previous section of this chapter, an excess of electrolytes in the extracellular fluid may be caused by increased electrolyte intake or absorption, shift of electrolytes from an electrolyte pool into the extracellular fluid, and decreased electrolyte excretion, either singly or in combination. Conversely, a deficit of electrolytes in the extracellular fluid may be caused by decreased electrolyte intake or absorption, shift of electrolytes from the extracellular fluid to an electrolyte pool, increased electrolyte excretion, loss of electrolytes through abnormal routes, or some combination of these factors.

Plasma Potassium

The normal concentration of potassium ions in plasma, is 3.5 to 5.0 mEq/L (may vary slightly with different laboratories), except in

neonates, in whom it may be higher. Most of the potassium ions in the body are inside cells; the standard serum potassium measurement gives only the concentration of the small portion of potassium ions in the extracellular fluid. Because a number of factors cause potassium ions to move into or out of body cells, concentration of potassium in the plasma and total body potassium content are not necessarily correlated. Whether or not they are accompanied by total body potassium imbalances, plasma potassium imbalances may cause clinically significant signs and symptoms.

Hypokalemia

Hypokalemia denotes a decreased potassium ion concentration in the extracellular fluid. A decrease in the plasma potassium concentration does not necessarily denote a decrease in total body potassium. Thus hypokalemia may coexist with a total body potassium deficit, a total body potassium excess, or a normal total body potassium ion concentration.

Etiology. Hypokalemia is caused by factors that decrease potassium intake, shift potassium from the extracellular fluid into the cells, increase potassium excretion through the normal routes, and cause potassium loss from the body by some abnormal route. Potassium-wasting diuretics and corticosteroids such as prednisone are well-known causes of hypokalemia from increased renal potassium excretion. The hormone aldosterone increases potassium excretion in urine; hypokalemia is associated with pathophysiologic conditions such as compensated heart failure and cirrhosis that are accompanied by increased aldosterone levels. Black licorice contains a substance that increases renal potassium excretion. Many traditional Chinese medicines and other herbal preparations contain black licorice, and excessive ingestion of these agents or black licorice candy leads to hypokalemia. In many cases, several factors lead to hypokalemia. For example, people who have eating disorders may eat very little (decreased potassium intake) and abuse diuretics and/or laxatives (increased potassium excretion). Specific causes of hypokalemia are listed in Box 24.6.

Clinical manifestations. The resting membrane potential of muscle cells is determined by the ratio of intracellular to extracellular potassium ion concentration. For this reason, potassium imbalances cause altered function of muscles (skeletal, smooth, and cardiac). In hypokalemia, both smooth and skeletal muscle cells are hyperpolarized (more electrical charge than usual across the cell membrane). Therefore these muscles are less reactive to stimuli. The resulting clinical manifestations include abdominal distention, diminished bowel sounds, paralytic ileus, postural hypotension, skeletal muscle weakness, and flaccid paralysis. The skeletal muscle weakness of hypokalemia is bilateral weakness that typically begins in the lower extremities and ascends. It may involve the respiratory muscles, causing respiratory paralysis more commonly than does hyperkalemia.

Many types of cardiac dysrhythmias arise from hypokalemia. Cardiac muscle cells usually become hyperpolarized with hypokalemia. However, with very low plasma potassium concentrations, hypopolarization of cardiac muscle occurs, most likely because of decreased potassium conductance. Hypokalemia also increases the rate of diastolic depolarization, which may give rise to ectopic beats, decreases conduction velocity in the atrioventricular node, prolongs cardiac action potentials by decreasing the rate of repolarization, shortens the absolute refractory period, and prolongs the relative refractory period.

Hypokalemia also may cause polyuria by interfering with the action of ADH at the renal tubules. The plasma potassium concentration at which the various clinical manifestations of hypokalemia appear depends on individual responsiveness and the presence of other concurrent electrolyte and acid–base disorders. Chronic hypokalemia can cause rhabdomyolysis (skeletal muscle breakdown), selective myocardial cell necrosis, and nephropathy.

BOX 24.6 Causes of Hypokalemia

Decreased Potassium Intake
Anorexia
NPO (nothing by mouth) orders and intravenous solutions without potassium
Fasting
Unbalanced diet

Shift of Potassium From Extracellular Fluid to Cells
Alkalosis
Excess insulin (e.g., during total parenteral nutrition)
Excess β-adrenergic stimulation
Hypokalemic familial periodic paralysis

Increased Potassium Excretion Through Normal Routes
Renal Route
Potassium-wasting diuretics
Corticosteroid therapy
Cushing disease
Hyperaldosteronism
Excessive ingestion of black licorice (glycyrrhizin)
Hypomagnesemia
Parenteral piperacillin or similar agents
Amphotericin B, cisplatin, cyclosporine, and many other drugs

Fecal Route
Diarrhea (includes laxative overuse or abuse)

Skin Route
Excessive diaphoresis

Loss of Potassium Through Abnormal Routes
Emesis
Gastric suction
Fistula drainage

Hyperkalemia

If the serum potassium concentration rises above 5.0 mEq/L (the upper limit of normal), hyperkalemia is present. Hyperkalemia denotes an elevation of potassium ion concentration in the extracellular fluid. As mentioned previously, most of the potassium ions in the body are inside cells, and many factors cause potassium ions to move into or out of the cells. Thus total body potassium content may be increased, normal, or decreased in hyperkalemia, depending on its cause.

Etiology. Hyperkalemia is caused by factors that increase potassium intake, shift potassium from the cells into the extracellular fluid, and decrease potassium excretion. For example, massive blood transfusion can cause hyperkalemia by increased potassium intake because the transfused fluid surrounding the red blood cells (RBCs) is high in potassium that was released from the RBCs during storage. Large numbers of potassium ions shifting from cells into the extracellular fluid after a crushing injury or massive cell death from cytotoxic chemotherapy will cause hyperkalemia. People who take two or more drugs that can increase plasma potassium concentration need monitoring for hyperkalemia. Several factors together can cause hyperkalemia, such as when an individual who develops decreased potassium excretion because of oliguric chronic kidney disease continues to have a normal dietary potassium intake. Specific causes of hyperkalemia are summarized by category in Box 24.7.

Clinical manifestations. As might be expected from the role of potassium ions in establishment of the resting membrane potential of muscle cells, hyperkalemia causes muscle dysfunction. As hyperkalemia

BOX 24.7 Causes of Hyperkalemia

Increased Potassium Intake

Excessive or too-rapid intravenous potassium infusion
Insufficiently mixed intravenous potassium infusion
Large transfusion of stored blood
Massive doses of potassium penicillin G

Shift of Potassium From Cells to Extracellular Fluid

Acidosis caused by nonorganic acids
Insufficient insulin
Crushing injury
Cytotoxic drugs (tumor lysis syndrome)
Hyperkalemic periodic paralysis
β-Adrenergic blockade with prolonged strenuous exercise

Decreased Potassium Excretion

Oliguria (such as in hypovolemia, acute kidney injury, or chronic end-stage renal disease)
Potassium-sparing diuretics
Adrenal insufficiency
Renin-deficient states
Drugs that reduce aldosterone effects (direct renin inhibitors, angiotensin-converting enzyme [ACE] inhibitors, angiotensin II receptor antagonists, and selective aldosterone blockers)
Nephrotoxic drugs

develops, smooth muscle and skeletal muscle cells become hypopolarized. The main clinical manifestation at this stage is mild intestinal cramping and diarrhea, which occurs only in some individuals. As hyperkalemia worsens, skeletal muscle cells become hypopolarized to the extent that their resting membrane potentials lie above their threshold potential; once they have discharged, they are unable to contract again. This situation causes the typical skeletal muscle weakness and flaccid paralysis of hyperkalemia. The skeletal muscle weakness is an ascending weakness that appears first in the lower extremities. Both hypokalemia and hyperkalemia cause skeletal muscle weakness and/or paralysis, but the underlying alterations in the resting membrane potentials are different.

Cardiac muscle undergoes the same changes in resting membrane potential as skeletal muscle in hyperkalemia. In addition, hyperkalemia decreases the duration and rate of rise of cardiac action potentials and decreases conduction velocity in the heart. These pathophysiologic mechanisms underlie the cardiac dysrhythmias of hyperkalemia. Severe hyperkalemia causes cardiac arrest.

The plasma potassium concentration at which each of these clinical manifestations occurs varies, depending on the rapidity of rise of the potassium concentration, the causes of the hyperkalemia, and other concurrent electrolyte or acid–base imbalances. Patients who have chronic end-stage renal disease often undergo potassium adaptation and have relatively mild symptoms at high plasma potassium concentrations that would be disabling in other persons. The mechanisms of potassium adaptation include increased aldosterone levels that increase potassium excretion by the colon and shift potassium ions from extracellular fluid into cells, helping normalize resting membrane potentials.

Plasma Calcium

Calcium in the plasma is present in three forms: some calcium ions are bound to plasma proteins (such as albumin), some are bound to small organic ions (such as citrate), and the rest are unbound. Only the free ionized calcium is physiologically active. Two laboratory measurements are available for calcium: total serum calcium and ionized calcium. The *total serum calcium* measurement includes all of the calcium (bound plus unbound). The normal range of total serum calcium concentration in adults is 9 to 11 mg/dL or 4.5 to 5.5 mEq/L (may vary slightly with different laboratories). Unless a calcium value specifies ionized calcium, it is total calcium. The *ionized calcium* measurement includes only the unbound ionized form. The normal range of ionized calcium in adults is 4.0 to 5.0 mg/dL, about half of the total calcium (varies with different laboratories). Clinically significant calcium imbalances are caused by alterations in the plasma concentration of unbound ionized calcium.

Hypocalcemia

Hypocalcemia occurs if the serum calcium concentration drops below the lower limit of normal. If the fraction of unbound ionized calcium in the blood is decreased by more calcium binding to plasma proteins or other organic ions such as citrate, the total serum calcium concentration (the usual laboratory measurement) may be normal, but *ionized* hypocalcemia is present and may cause signs and symptoms. Ionized hypocalcemia is common with massive transfusion of blood or fresh frozen plasma because citrate is part of the anticoagulant used to preserve both of these solutions.

Etiology. Hypocalcemia is caused by factors that decrease calcium intake or absorption, decrease the physiologic availability of calcium, and increase calcium excretion. For example, hypocalcemia in pancreatitis arises from impaired fat digestion caused by lack of pancreatic lipase in the intestines. Both dietary calcium and calcium ions secreted into the intestine from the extracellular fluid bind to undigested fat in the intestine and are excreted in the feces. Thus both decreased calcium absorption and increased calcium excretion play a part in hypocalcemia associated with pancreatitis. In addition, calcium ions can bind to necrotic tissue in the pancreas, decreasing their physiologic availability. Ionized hypocalcemia is common in intensive care unit patients, again due to multiple factors. PTH increases plasma calcium concentration; thus hypocalcemia occurs in people who have hypoparathyroidism from parathyroid injury during thyroid surgery or other causes. Box 24.8 lists specific causes of hypocalcemia organized according to the general etiologic factors.

Clinical manifestations. Calcium ions play an important role in determining the speed of ion fluxes through nerve and muscle cell membranes. Thus calcium imbalances alter normal neuromuscular excitability. Clinical manifestations of hypocalcemia are those of increased neuromuscular excitability: positive Trousseau sign, positive Chvostek sign, paresthesias, muscle twitching and cramping, hyperactive reflexes, carpal spasm, pedal spasm, tetany, laryngospasm, seizures, and cardiac dysrhythmias. The increased neuromuscular excitability of hypocalcemia is caused by a decrease in the threshold potential of excitable cells so that action potentials are generated more easily. Cardiac effects of hypocalcemia arise from the prolonged plateau phase of the cardiac action potential, impaired atrioventricular and intraventricular conduction, and impaired myocardial contractility, which can cause heart failure.

Positive Trousseau sign is occurrence of a carpal spasm after occlusion of arterial blood flow to the hand for approximately 3 minutes. Positive Chvostek sign is spasm of muscles in the cheek and corner of the mouth produced by tapping the facial nerve in front of the ear. Positive Trousseau and Chvostek signs are general indicators of increased neuromuscular excitability from any cause, so they must be interpreted in the context of other clinical manifestations and specific risk factors for hypocalcemia. Chvostek sign may be positive in neonates without electrolyte imbalances.

Hypercalcemia

Hypercalcemia occurs when the serum calcium concentration rises above the upper limit of normal (11 mg/dL or 5.5 mEq/L). It

BOX 24.8 Causes of Hypocalcemia

Decreased Calcium Intake or Absorption
Diet with insufficient calcium and vitamin D
Chronic kidney disease (deficient activated vitamin D)
Excessive dietary phytates or oxalates
Steatorrhea
Pancreatitis
Chronic diarrhea (includes laxative overuse or abuse)
Malabsorption syndromes

Decreased Physiologic Availability of Calcium
Hypoparathyroidism
Excessive phosphate intake
Tumor lysis syndrome (high phosphate)
Hypomagnesemia
Alkalosis
Large transfusion of citrated blood or fresh frozen plasma
Rapid infusion of plasma expanders that bind calcium
Elevated plasma free fatty acids
Chronic kidney disease

Increased Calcium Excretion Through Normal Routes
Steatorrhea
Pancreatitis

BOX 24.9 Causes of Hypercalcemia

Increased Calcium Intake or Absorption
Milk-alkali syndrome
Vitamin D overdose (includes shark cartilage supplements)

Shift of Calcium From Bone to Extracellular Fluid
Hyperparathyroidism
Immobilization
Paget disease
Bone tumors
Multiple myeloma
Leukemia
Nonosseous malignancies that produce bone-resorbing factors

Decreased Calcium Excretion
Thiazide diuretics
Familial hypocalciuric hypercalcemia

indicates an elevation of the calcium concentration of the extracellular fluid.

Etiology. Hypercalcemia is caused by factors that increase calcium intake or absorption, cause a shift of calcium from bone to extracellular fluid, and decrease calcium excretion. Because PTH shifts calcium out of bone, hyperparathyroidism causes hypercalcemia. Many malignant tumors produce chemicals that circulate in the blood and shift calcium from bones into extracellular fluid. These bone-resorbing factors include PTH-related peptide and prostaglandins. In addition, circulating factors in malignancy may decrease renal excretion of calcium ions, which also contributes to hypercalcemia. Specific causes of hypercalcemia are listed by category in Box 24.9.

Clinical manifestations. Hypercalcemia causes decreased neuromuscular excitability. Clinical manifestations of hypercalcemia include anorexia, nausea, emesis, constipation, fatigue, polyuria, muscle weakness, diminished reflexes, headache, confusion, lethargy, personality change, and cardiac dysrhythmias. The decreased neuromuscular excitability is caused by elevation of the threshold potential of excitable cells. Cardiac effects of hypercalcemia include shortened plateau phase of the action potential, increased rate of diastolic depolarization of sinus node cells, and delayed atrioventricular conduction. Renal calculi may occur as a result of the high calcium concentration of the urine. Hypercalcemia caused by bone resorption may lead to pathologic fractures.

Plasma Magnesium

The normal serum magnesium concentration is 1.5 to 2.5 mEq/L (may vary slightly with different laboratories). Similar to calcium, magnesium ions also are present in the blood as bound (physiologically inactive) and unbound ionized (physiologically active) forms. Measurement of ionized magnesium levels is available in some research settings. Plasma magnesium concentration imbalances may occur concurrent with or in the absence of total body magnesium content imbalances.

Hypomagnesemia

If the serum magnesium concentration decreases below the lower limit of normal (1.5 mEq/L), hypomagnesemia is present. Hypomagnesemia indicates a decreased magnesium concentration of the extracellular fluid and does not necessarily indicate a total body magnesium deficit (although the two may occur concurrently).

Etiology. Causes of hypomagnesemia are decreased magnesium intake or absorption, decreased physiologic availability of magnesium, increased magnesium excretion, and loss of magnesium by an abnormal route. Chronic alcoholism is a major risk factor for hypomagnesemia because it is associated with decreased magnesium intake, decreased physiologic availability of magnesium, increased urinary and fecal magnesium excretion, and magnesium loss through emesis. Hypomagnesemia often causes hypokalemia by increasing urinary excretion of potassium. Hypomagnesemia also may cause hypocalcemia by inhibiting secretion of PTH. In both cases, correction of hypomagnesemia is necessary before the other electrolyte imbalance can be corrected. Specific causes of hypomagnesemia are listed in Box 24.10.

Clinical manifestations. Magnesium ions in the extracellular fluid normally depress the release of acetylcholine at neuromuscular junctions. If too few magnesium ions are present, excessive amounts of acetylcholine are released (Fig. 24.11). Therefore the clinical manifestations of hypomagnesemia are those of increased neuromuscular excitability. Such manifestations may include insomnia, hyperactive reflexes, muscle cramps, muscle twitching, grimacing, positive Chvostek sign, positive Trousseau sign, nystagmus, dysphagia, ataxia, tetany, and seizures. Cardiac dysrhythmias also occur.

Hypomagnesemia causes decreased activity of the enzyme that drives the Na^+–K^+ pump in cell membranes so that intracellular potassium concentration decreases in the myocardium. Increased spontaneous firing in the sinus node, shortening of the absolute refractory period, and lengthening of the relative refractory period contribute to cardiac dysrhythmias in hypomagnesemia.

Hypermagnesemia

If the serum magnesium concentration rises above the upper limit of normal (2.5 mEq/L), hypermagnesemia is present. Hypermagnesemia indicates an excess of magnesium in the extracellular fluid.

Etiology. The major causes of hypermagnesemia are increased magnesium intake and decreased magnesium excretion. Shift of magnesium from bones to extracellular fluid is seen transiently in some stages of hyperparathyroidism. Hypermagnesemia from excessive intake of magnesium in laxatives and antacids occurs in people of any age who have unrecognized renal impairment or receive high doses. Older

adults are at high risk from these magnesium-containing medications. Individuals who have oliguria, as in chronic end-stage renal disease, are another high-risk group for development of hypermagnesemia. Specific causes of hypermagnesemia are summarized in Box 24.11.

Clinical manifestations. Too many magnesium ions in the extracellular fluid depress neuromuscular function by decreasing the release of acetylcholine at neuromuscular junctions (see Fig. 24.11). Thus manifestations of hypermagnesemia include decreased deep tendon reflexes, lethargy, hypotension, flushing, diaphoresis, drowsiness, flaccid paralysis, respiratory depression, bradycardia, cardiac dysrhythmias, and even cardiac arrest. Mechanisms that cause the cardiac effects of hypermagnesemia include decreased cardiac conduction and depression of membrane excitability.

BOX 24.10 Causes of Hypomagnesemia

Decreased Magnesium Intake or Absorption
Chronic alcoholism
Malnutrition
Prolonged intravenous therapy without magnesium supplementation
Ileal resection
Chronic diarrhea (includes laxative overuse or abuse)
Malabsorption syndromes
Steatorrhea
Pancreatitis

Decreased Physiologic Availability of Magnesium
Elevated plasma free fatty acids

Increased Magnesium Excretion Through Normal Routes
Renal Route
Diabetic ketoacidosis
Chronic alcoholism
Hyperaldosteronism
Diuretic therapy
Aminoglycoside (e.g., gentamicin) toxicity
Amphotericin B, cisplatin, and many other drugs

Fecal Route
Steatorrhea
Pancreatitis

Magnesium Loss Through Abnormal Routes
Emesis
Gastric suction
Fistula drainage

Plasma Phosphate

The normal range of phosphate concentration in adult plasma is 2.5 to 4.5 mg/dL (may vary slightly with different laboratories). Symptomatic phosphate imbalances are less common than other electrolyte imbalances, but, like other electrolyte imbalances, they may be fatal if untreated.

Hypophosphatemia

Hypophosphatemia is present when the phosphate concentration in the plasma decreases below the lower limit of normal (2.5 mg/dL). People who have moderate hypophosphatemia may or may not have symptoms, but people with severe hypophosphatemia (plasma phosphate concentration less than 1.5 mg/dL) have severe clinical manifestations that are fatal unless treated successfully.

Etiology. Hypophosphatemia is caused by factors that decrease phosphate intake, shift phosphate from extracellular fluid to cells, increase phosphate excretion, and cause loss of phosphate through abnormal routes. Frequently, many factors combine to produce severe symptomatic hypophosphatemia. Any factor that causes a rapid increase in cellular

BOX 24.11 Causes of Hypermagnesemia

Increased Magnesium Intake or Absorption
Ingestion or aspiration of seawater
Excessive ingestion of magnesium-containing medications (e.g., laxatives, antacids)
Excessive intravenous infusion of magnesium

Decreased Magnesium Excretion
Oliguric renal disease
Adrenal insufficiency

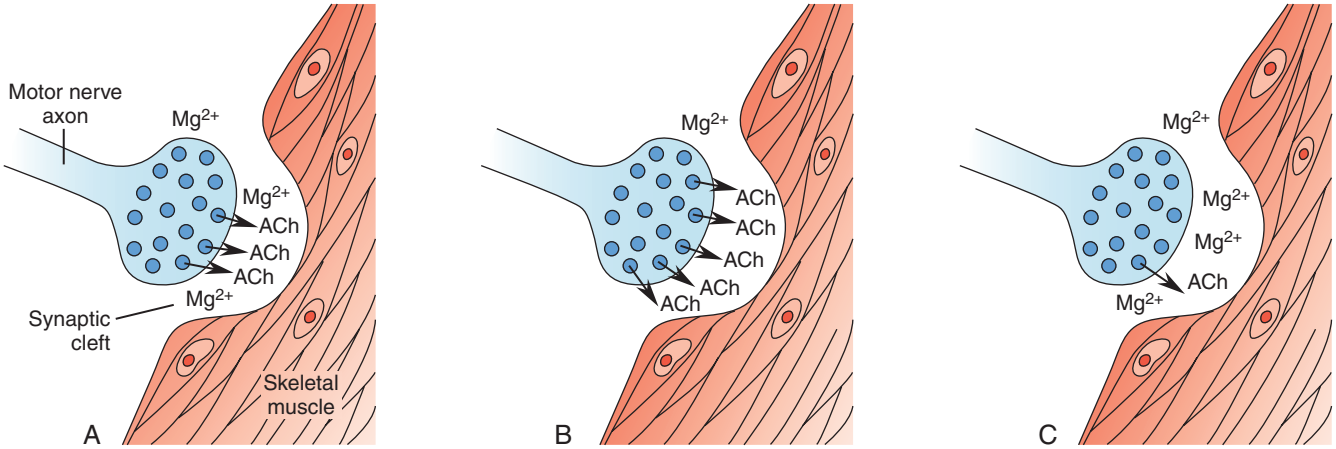


FIG 24.11 Acetylcholine (ACh) release at neuromuscular junctions is altered in magnesium imbalances. **A,** Normal magnesium concentration suppresses the release of ACh at neuromuscular junctions to normal levels. **B,** In hypomagnesemia, more ACh is released at neuromuscular junctions, causing increased neuromuscular excitability. **C,** In hypermagnesemia, less ACh is released at neuromuscular junctions, causing decreased neuromuscular excitability.

BOX 24.12 Causes of Hypophosphatemia

Decreased Phosphate Intake or Absorption
Chronic alcoholism
Chronic diarrhea
Malabsorption syndromes
Excessive or long-term use of antacids that bind phosphate

Shift of Phosphate From Extracellular Fluid to Cells
Refeeding after starvation (includes anorexia nervosa)
Total parenteral nutrition
Hyperventilation (respiratory alkalosis)
Insulin
Epinephrine
Intravenous glucose, fructose, bicarbonate, or lactate

Increased Phosphate Excretion Through the Normal Renal Route
Alcohol withdrawal
Diuretic phase after extensive burns
Diabetic ketoacidosis
Diuretic therapy

Phosphate Loss Through Abnormal Routes
Emesis
Hemodialysis

BOX 24.13 Causes of Hyperphosphatemia

Increased Phosphate Intake or Absorption
Overzealous phosphate therapy
Excessive use of phosphate-containing enemas or laxatives

Shift of Phosphate From Cells to Extracellular Fluid
Tumor lysis syndrome
Crushing injury
Rhabdomyolysis

Decreased Phosphate Excretion
End-stage renal disease
Oliguric acute kidney injury
Adrenal insufficiency

metabolism will cause phosphate to shift from extracellular fluid into cells. Patients who are severely malnourished (such as cancer patients with advanced disease or people who have severe anorexia nervosa) are at high risk for severe symptomatic hypophosphatemia after nutritional replacement is started because of their increased cellular metabolism and previously depleted phosphate stores. Specific factors that cause hypophosphatemia are summarized in Box 24.12.

Clinical manifestations. Phosphate is an important component of adenosine triphosphate (ATP), the major source of energy for many cellular processes. Signs and symptoms of symptomatic hypophosphatemia are due, in part, to decreased amounts of ATP within the cells. Another contributing mechanism is tissue hypoxia caused by decreased amounts of 2,3-bisphosphoglycerate (2,3-BPG) in the RBCs. Signs and symptoms include anorexia, malaise, paresthesias, hemolysis, diminished reflexes, muscle aches, muscle weakness, respiratory failure caused by weak diaphragm muscles, confusion, stupor, seizures, coma, cardiac dysrhythmias, and impaired cardiac function. The impaired cardiac function of severe hypophosphatemia arises from decreased cardiac contractility concurrent with increased left ventricular end-diastolic pressure and may cause heart failure.

Hyperphosphatemia

Hyperphosphatemia is an increase of the serum phosphate concentration above the upper limit of normal (4.5 mg/dL).

Etiology. Hyperphosphatemia may be caused by increased phosphate intake, shift of phosphate from cells or bones to extracellular fluid, and decreased phosphate excretion. Examples of specific causes in these categories are listed in Box 24.13. Hyperphosphatemia is common in people who have oliguric renal disease, either acute or chronic. In chronic kidney disease, renal phosphate excretion is severely decreased by substantial nephron loss and elevated fibroblast growth factor 23 (FGF-23), which inhibits phosphate excretion. In addition, intestinal absorption of dietary phosphate continues and elevated parathyroid hormone in chronic kidney disease shifts phosphate from bones into extracellular fluid.

Clinical manifestations. Clinical manifestations of hyperphosphatemia depend on the effect of the elevated phosphate ion concentration on calcium ions. Typically, hyperphosphatemia causes hypocalcemia. The signs and symptoms are thus the manifestations of increased neuromuscular excitability that were presented in the discussion of hypocalcemia. However, in some patients, especially those who have chronic kidney disease, hyperphosphatemia causes deposition of calcium phosphate salts in the soft tissues of the body. These patients develop signs and symptoms such as aching and stiffness of joints, itching (pruritus), and conjunctivitis, depending on the areas in which these salts precipitate.

KEY POINTS

- Plasma electrolyte deficits are caused by factors that decrease electrolyte intake or absorption, shift electrolytes from the extracellular fluid to an electrolyte pool, increase electrolyte excretion, and cause loss of electrolytes through abnormal routes.
- Plasma electrolyte excesses are caused by factors that increase electrolyte intake or absorption, shift electrolytes from an electrolyte pool to the extracellular fluid, and decrease electrolyte excretion.
- Abnormalities in plasma electrolyte concentrations may profoundly affect cellular function. Excitable cells, such as nerve and muscle, are particularly sensitive to electrolyte imbalances.
- Manifestations of potassium imbalances are due to changes in resting membrane potentials. Hypokalemia causes hyperpolarization; hyperkalemia causes hypopolarization. Both hyperkalemia and hypokalemia cause skeletal muscle weakness, flaccid paralysis, and cardiac dysrhythmias.
- Manifestations of calcium imbalances are caused by changes in the threshold potential of nerve and muscle cells. Hypocalcemia decreases the threshold potential, causing hyperexcitability (twitching, tetany); hypercalcemia increases the threshold potential, causing neuromuscular depression (hyporeflexia).
- Manifestations of magnesium imbalances are similar to those of calcium imbalances. Magnesium ions normally inhibit release of acetylcholine at neuromuscular junctions. Hypomagnesemia increases neuromuscular excitability (hyperreflexia and twitching), and hypermagnesemia depresses neuromuscular excitability (hyporeflexia and flaccid paralysis).
- Symptomatic hypophosphatemia is characterized by manifestations of generalized cellular adenosine triphosphate (ATP) deficiency. Hyperphosphatemia may cause hypocalcemia, with resulting increased neuromuscular excitability, or it may be associated with precipitation of calcium phosphate into soft tissues of the body.

SUMMARY

This chapter presents the principles of fluid and electrolyte homeostasis and imbalances. The following boxes summarize pediatric and geriatric considerations, respectively. Fluid and electrolyte homeostasis involves the continuous interplay of intake, absorption, distribution, and excretion of fluid and electrolytes. Loss of fluid and electrolytes through abnormal routes may also occur. When the normal mechanisms are impaired or overwhelmed, fluid and electrolyte imbalances occur. Fluid imbalances may involve the volume or the concentration of body fluid. Plasma electrolyte imbalances may be deficits or excesses and may not reflect total body electrolyte deficits or excesses. Signs and symptoms of fluid and electrolyte imbalances are summarized in Table 24.2. The pathophysiology of specific fluid and electrolyte imbalances can be derived from a working knowledge of normal fluid and electrolyte homeostasis.

PEDIATRIC CONSIDERATIONS

Fluid and Electrolyte Homeostasis and Imbalances

- Infants have more extracellular fluid than intracellular fluid; this proportion reverses by a few months of age.
- About 75% of the body weight of a term infant is water; this percentage is even higher in preterm infants. The percentage of body weight that is water decreases as the child ages.
- In the first few days after birth, an infant loses fluid equal to 5% to 10% of its body weight; this is a normal process during adjustment to extrauterine life.
- Neonates have a high metabolic rate and thus a high turnover rate of water.
- Infants have increased insensible water excretion caused by proportionately large body surface area, proportionately large respiratory mucosa surface area, vasomotor immaturity, and increased skin permeability. Preterm infants have even greater insensible water excretion through the skin because of flaccid extended posture (and thus greater exposed body surface area) and greater vasomotor immaturity.
- Use of phototherapy and radiant heat warmers increases insensible water excretion.
- Glomerular filtration rate is lower in infants than in adults.
- The kidneys of infants have limited ability to concentrate urine or to dilute it; thus infants are unable to excrete a large load of water effectively or to conserve fluid when needed.
- Infants communicate thirst by crying, which may not be understood by their caregivers.
- Assessment of extracellular volume imbalances in infants should focus on the tension of the fontanel rather than the degree of filling of neck veins.
- Infants whose caregivers use powdered formula are at high risk for hypernatremia if the formula is reconstituted with extra powder to "strengthen" the baby.
- Laboratory normal ranges of electrolytes generally are wider for infants than for older children and adults.
- Neonatal hypocalcemia may occur in infants who needed resuscitation at birth or have high-risk conditions.
- Preterm infants may have reduced body calcium stores because fetal calcium stores are built during the last trimester of pregnancy; these infants have increased incidence of neonatal hypocalcemia.
- Assessment of increased neuromuscular excitability (hypocalcemia and hypomagnesemia) in infants should not include Chvostek sign; this sign is often positive in normal neonates. Increased neuromuscular excitability in infants includes jitteriness, hyperactive reflexes, and a high-pitched cry.
- Neonates whose mothers were given magnesium sulfate for eclampsia in the 24 hours before birth may be born with hypermagnesemia. Hypermagnesemic infants lie in a flaccid, extended posture.

GERIATRIC CONSIDERATIONS

Fluid and Electrolyte Homeostasis and Imbalances

- Older adults have less body water than do middle-aged adults because body composition changes with increasing age (decreased muscle mass, increased fat in internal organs). About 50% of the body weight of a lean older man is water, and about 45% of the body weight of a lean older woman is water; the percentage is lower in obese older adults.
- Glomerular filtration rate is lower in older adults than in middle-aged adults.
- The kidneys of older adults are less able to concentrate urine and thus less able to conserve fluid when needed. This decreased ability to concentrate urine also is responsible for nocturia, because a larger-than-normal volume of urine is produced at night.
- Older adults have a reduced thirst response when the osmolality of body fluids increases; thus they may not be aware that they are becoming dehydrated.
- Older adults are a high-risk group for clinical dehydration.
- Decreased skin turgor is not reliable as a sign of extracellular fluid volume depletion in older adults because of age-related changes in collagen and elastin. Decreased skin turgor (skin tenting when pinched) may occur in older adults who have normal fluid volume.
- Older adults who receive tube feedings are at higher risk for hypernatremia than are middle-aged adults.
- Older adults probably absorb more magnesium from antacids and cathartics than do middle-aged adults. With age-related changes in renal excretion, older adults who use oral magnesium laxatives or antacids regularly are at high risk for hypermagnesemia.

TABLE 24.2 Summary of Signs and Symptoms of Fluid and Electrolyte Imbalances

Imbalance	Heart	Blood Vessels	Interstitial Area	CNS	Lungs
↓ Extracellular volume	Tachycardia	Postural blood pressure decrease with concurrent heart rate increase, flat neck veins, ↑ small-vein filling time, thready pulse	↓ Skin turgor; soft, sunken eyeballs; longitudinal furrows in tongue	Lightheadedness, syncope	
↑ Extracellular volume		Distended neck veins, bounding pulse	Edema		Crackles, dyspnea, orthopnea, frothy sputum
↓ Na^+				Confusion, lethargy, coma, seizures	
↑ Na^+				Confusion, lethargy, coma, seizures	
↓ K^+	Dysrhythmias	Postural hypotension			
↑ K^+	Dysrhythmias, cardiac arrest				
↓ Ca^{++}	Dysrhythmias, impaired myocardial contractility			Seizures	
↑ Ca^{++}	Dysrhythmias			Confusion, lethargy, personality change	
↓ Mg^{++}	Dysrhythmias			Insomnia, seizures	
↑ Mg^{++}	Bradycardia, dysrhythmias, cardiac arrest	Hypotension, flushing		Drowsiness, lethargy	Respiratory depression
↓ Pi	Impaired cardiac function, decreased cardiac output			Confusion, stupor, coma, seizures	Respiratory failure
↑ Pi (may cause ↓ Ca^{++})					

CNS, Central nervous system; *ECV,* extracellular fluid volume; *Pi,* inorganic phosphate.

Skeletal Muscle	Neuromuscular Excitability	Gastrointestinal Tract	Kidneys	Other
		Dry oral mucous membranes, hard stools	Oliguria	Sudden weight loss, sunken fontanel (infants), no tears or sweat, thirst with severe ↓ ECV
		Hepatomegaly		Sudden weight gain, bulging fontanel (infants)
		Anorexia, nausea, emesis		Malaise, headache
			Oliguria	Thirst
Ascending weakness, flaccid paralysis		Abdominal distention, bloating, ↓ bowel sounds, constipation, paralytic ileus	Polyuria	
Ascending weakness, flaccid paralysis		Transient mild cramping, diarrhea		
Twitching, cramping, carpal spasm, pedal spasm	Increased excitability, Trousseau sign, Chvostek sign, paresthesias, hyperactive reflexes, tetany			Laryngospasm
Weakness	Decreased excitability, depressed reflexes	Anorexia, nausea, emesis, constipation	Polyuria	Fatigue, headache
Twitching, cramping, grimacing, tremors	Increased excitability, Trousseau sign, Chvostek sign, hyperactive reflexes, tetany	Dysphagia		Nystagmus, ataxia
Flaccid paralysis	Depressed reflexes			Diaphoresis
Aching, weakness	Paresthesias, depressed reflexes	Anorexia		Malaise, hemolysis
			If Ca^{++} remains high, damage from deposition of crystals	If Ca^{++} remains high, pruritus, conjunctivitis, arthritis

RESOURCES

Fluid, Water, and Sodium

Bauersachs J, Jaisser F, Toto R: Mineralocorticoid receptor activation and mineralocorticoid receptor antagonist treatment in cardiac and renal diseases. *Hypertension* 65(2):257–263, 2015.

Biglia N, Librino A, Ottino MC, et al: Lower limb lymphedema and neurological complications after lymphadenectomy for gynecological cancer. *Int J Gynecol Cancer* 25:521–525, 2015.

Burchum JR, Rosenthal LD: *Lehne's pharmacology for nursing care*, ed 9, St. Louis, MO, 2016, Saunders Elsevier.

Danz M, Pottgen K, Tonjes PM, et al: Hyponatremia among triathletes in the ironman european championship. *New Engl J Med* 374:997–998, 2016.

Felver L: Fluid and electrolyte balance. In Giddens J, editor: *Concepts for nursing practice*, ed 2, Philadelphia, 2017, Mosby Elsevier, pp 62–74.

Hall JE: *Guyton and Hall textbook of medical physiology*, ed 13, Philadelphia, 2016, Elsevier.

Longo D, Jameson J, Loscalzo J, et al: *Harrison's principles of internal medicine*, ed 19, New York, 2015, McGraw-Hill.

Morley JE: Dehydration, hypernatremia, and hyponatremia. *Clin Geriatr Med* 31:398–399, 2015.

Pallavi R: An unsuspected cause of hyponatremia: beer potomania. *J Am Geriatr Soc* 63:1714–1715, 2015.

Palmer BF, Clegg DJ: An emerging role of natriuretic peptides: igniting the fat furnace to fuel and warm the heart. *Mayo Clinic Proceedings* 90:1666–1678, 2015.

Paskett ED: Symptoms: lymphedema. *Advances Experimental Med Biol* 862:101–113, 2015.

Rodenburg EM, Hoorn EJ, Ruiter R, et al: Thiazide-associated hyponatremia: a population-based study. *Am J Kidney Dis* 62(1):67–72, 2013.

Spasovsky G, Vanholder R, Allolio B, et al: Clinical practice guideline on diagnosis and treatment of hyponatraemia. *Eur J Endocrinol* 170:G1–G47, 2014.

Stanhewicz AE, Kenney WL: Determinants of water and sodium intake and output. *Nutr Rev* 73(Suppl 2):73–82, 2015.

Sterns R: Disorders of plasma sodium—causes, consequences, and correction. *New Engl J Med* 372:55–65, 2015.

Verbrugge FH, Steels P, Grieten L, et al: Hyponatremia in acute decompensated heart failure: depletion versus dilution. *J Am Coll Cardiol* 65:480–492, 2015.

Potassium

Ayach T, Nappo RW, Paugh-Miller JL, Ross EA: Postoperative hyperkalemia. *Eur J Int Med* 26:106–111, 2015.

Balhara KS, Highet B, Omron R: Hypokalemia causing rhabdomyolysis in a patient with short bowel syndrome. *J Emer Med* 48(4):e97–e99, 2015.

Delhey PR, Holzapfel BM, Kanz K, et al: Hypokalemia mimicking a herniated vertebral disc. *Spine J* 15:e53–e56, 2015.

Gumz ML, Rabinowitz L, Wingo CS: An integrated view of potassium homeostasis. *New Engl J Med* 373:60–72, 2015.

Hampton JR: *The ECG in clinical practice*, ed 6, New York, 2013, Churchill Livingstone Elsevier.

Kawji MM, Glancy DL: Electrocardiogram in a 35-year-old man. *Am J Cardiol* 116:162, 2015.

Michel A, Martín-Pérez M, Ruigómez A, García Rodríguez LA: Risk factors for hyperkalaemia in a cohort of patients with newly diagnosed heart failure: a nested case–control study in UK general practice. *Eur J Heart Failure* 17:205–213, 2015.

Palmer BF, Clegg DJ: JAMA diagnostic test interpretation: hyperkalemia. *JAMA* 314:2405–2406, 2015.

Tanel RE: ECGs in the ED. *Ped Emer Care* 31(1):81–82, 2015.

Vaidya GN, Acevedo R: Tumor lysis syndrome in metastatic breast cancer after a single dose of paclitaxel. *Am J Emer Med* 33(2):308.e1–308.e2, 2015.

Westmoreland P, Krantz MJ, Mehler PS: Medical complications of anorexia nervosa and bulimia. *Am J Med* 129:30–37, 2016.

Calcium

Ahmad S, Kurangati G, Steencamp D: Hypercalcemic crisis: a clinical review. *Am J Med* 128:239–245, 2015.

Han P, Trinidad BJ, Shi J: Hypocalcemia-induced seizure: demystifying the calcium paradox. *ASN Neuro* 7(2):1–9, 2015.

Joosen DA, van de Laar RJ, Koopmans RP, et al: Acute dyspnea caused by hypocalcemia-related laryngospasm. *J Emer Med* 48(1):29–30, 2015.

Kelly A, Levine MA: Hypocalcemia in the critically ill patient. *J Intensive Care Med* 28:166–177, 2013.

Lee DL, Hinson AM, Siegel ER, et al: Comparison of intraoperative versus postoperative parathyroid hormone levels to predict hypocalcemia earlier after total thyroidectomy. *Otolaryngol Head Neck Surg* 153:343–349, 2015.

Minisola S, Pepe J, Piemonte S, et al: The diagnosis and management of hypercalcaemia. *BMJ* 350:h2723, 2015.

Romero AB, Beerman L, Arora G: Electrolyte screening in the evaluation of prolonged QTc interval. *Cardiol in the Young* 25:398–399, 2015.

Wagner L, Tata AL, Fink JC: Patient safety issues in CKD: core curriculum 2015. *Am J Kidney Dis* 66:159–169, 2015.

Magnesium and Phosphate

de Baaij JHF, Hoenderop JGJ, Bindels RJM: Magnesium in man: Implications for health and disease. *Physiol Rev* 95:1–46, 2015.

Manghat P, Sodi R, Swaminathan R: Phosphate homeostasis and disorders. *Ann Clin Biochem* 51:631–656, 2014.

Mukai A, Yamamoto S, Matsumura K: Hypocalcemia secondary to hypomagnesemia in a patient with Crohn's disease. *Clin J Gastroenterol* 8:22–25, 2015.

Parli SE, Ruf KM, Magnuson B: Pathophysiology, treatment, and prevention of fluid and electrolyte abnormalities during refeeding syndrome. *J Infusion Nurs* 37(3):197–202, 2014.

Paulin GP, Sandip SS, Hiren SI, et al: A study on changes in serum GGT and magnesium level in alcoholic liver disease. *Intl J Med Res Health Sci* 3(1):212–215, 2014.

25

Acid–Base Homeostasis and Imbalances

Linda Felver

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- What is the role of the bicarbonate buffer system in regulating pH of the blood?
- What is the role of the respiratory system in regulating carbonic acid (carbon dioxide)?
- What is the role of the kidneys in regulating bicarbonate ion and acids other than carbonic acid?
- How do the lungs compensate for acid–base imbalances caused by altered levels of metabolic acids?
- How do the kidneys compensate for acid–base imbalances caused by altered levels of carbonic acid?
- How are arterial blood gas values used to categorize an acid–base disorder as acidosis or alkalosis, respiratory or metabolic, compensated or uncompensated?
- What pathophysiologic conditions predispose an individual to each of the four primary acid–base imbalances?

CHAPTER OUTLINE

Acid–Base Homeostasis, 541
Buffers, 542
Respiratory Contribution, 542
Renal Contribution, 543

Acid–Base Imbalances, 545
Metabolic Acidosis, 545
Respiratory Acidosis, 546
Metabolic Alkalosis, 547
Respiratory Alkalosis, 548
Mixed Acid–Base Imbalances, 549

When the pH of body fluids becomes abnormal, cellular function is impaired. The pH of a fluid reflects its degree of acidity or alkalinity. Technically, pH is the negative logarithm of the hydrogen ion (H^+) concentration. The normal hydrogen ion concentration of the blood is about 40 nmol/L (40×10^{-9} mol/L)—a very small number. The pH (negative logarithm) of this number is 7.40, which is easier to use in clinical settings. An alteration in pH is a change in the hydrogen ion concentration. A high pH indicates few hydrogen ions, meaning that the solution is alkaline (basic). A low pH indicates a lot of hydrogen ions, meaning that the solution is acidic.

An acid releases hydrogen ions. The more hydrogen ions are present, the more acidic the solution. The normal pH of adult blood ranges from 7.35 to 7.45 (may vary slightly with different laboratories). The range is somewhat wider in infants and children. Table 25.1 lists normal laboratory values for pH and other acid–base parameters. If the blood and other body fluids become too acidic (reflected by pH decreased below the lower limit of the normal range), dysfunction occurs; if the pH of the blood falls below 6.9, death is likely to occur. Similarly, if the body fluids become too alkaline, as reflected by pH increased above the upper limit of the normal range, dysfunction also occurs. If the pH of the blood rises above 7.8, death is likely.

Normal cellular metabolism continually releases acids (carbonic and metabolic) that must be excreted from the body to prevent body fluids from becoming too acidic. This chapter discusses the normal mechanisms of acid–base homeostasis and the acid–base imbalances that arise when these homeostatic mechanisms become dysfunctional or overwhelmed.

ACID–BASE HOMEOSTASIS

Three major mechanisms regulate the acid–base status of the body: buffers, the respiratory system, and the renal system. Laboratory measurements such as arterial blood gas values are useful indicators of the acid–base status of extracellular fluids. The partial pressure of carbon dioxide in arterial blood ($PaCO_2$) is an indicator of the respiratory component of acid–base balance. The plasma bicarbonate ion (HCO_3^-) concentration is an indicator of the renal (metabolic) component of acid–base balance. The pH of the blood indicates the net effect of normal acid–base regulation, any acid–base imbalance, and the body's compensatory responses. It is important to remember that the pH measured clinically is that of the blood and may not reflect the pH inside cells or in cerebrospinal fluid.

TABLE 25.1 Normal Laboratory Values for Acid–Base Parameters

Characteristic	Normal Range
$PaCO_2$ (arterial blood)	36–44 mm Hg (adults) 30–34 mm Hg (infants)
HCO_3^- (serum)	22–26 mEq/L (adults) 19–23 mEq/L (infants)
pH (arterial blood)	7.35–7.45 (adults) 7.11–7.36 (neonates) 7.36–7.41 (infants)

Buffers

Buffers are chemicals that help control the pH of body fluids. Each buffer system consists of a weak acid, which releases hydrogen ions when the fluid is too alkaline, and a base, which takes up hydrogen ions when the fluid is too acidic. In this way, potential changes in pH are adjusted immediately by the action of buffers. All body fluids contain buffers. Chief among them are bicarbonate buffers (in the extracellular fluid), phosphate buffers (in intracellular fluid and renal tubular fluid that becomes the urine), hemoglobin buffers (inside erythrocytes), and protein buffers (in intracellular fluid and the blood). These buffers are the first line of defense against pH imbalances.

The bicarbonate buffer system is the most important buffer in the extracellular fluid. Bicarbonate ion (HCO_3^-) is the base portion, and carbonic acid (H_2CO_3) is the weak acid portion. These two components of the bicarbonate buffer system are in chemical equilibrium in the extracellular fluid. If too much metabolic acid (e.g., lactic acid) is present, the bicarbonate ions take up hydrogen ions (H^+) released by the metabolic acid and become carbonic acid. Through the action of the enzyme carbonic anhydrase, the carbonic acid then is excreted through the respiratory system in the form of carbon dioxide and water. Thus the excess acid is neutralized when bicarbonate ions are used in the buffering process.

$$HCO_3^- + H^+ \rightleftarrows H_2CO_3 \xrightarrow[\text{carbonic anhydrase}]{} CO_2 + H_2O$$

Conversely, if too little metabolic acid is present in the extracellular fluid, the carbonic acid portion of the bicarbonate buffer system releases hydrogen ions. This action helps keep the pH from becoming too high or at least minimizes the increase.

$$H_2CO_3 \rightleftarrows HCO_3^- + H^+$$

The pH of any fluid is determined by the relative amounts of acids and bases contained in it. For the pH of the blood to be within the normal range, the ratio of bicarbonate ions to carbonic acid must be 20:1, which means that 20 bicarbonate ions must be present for every carbonic acid molecule. This relationship is explained formally by the Henderson–Hasselbalch equation, which is a mathematical description of the pH of a buffered solution, here written specifically for the bicarbonate buffer system:

$$pH = pK_a + \log\frac{[HCO_3^-]}{[H_2CO_3]}$$

Square brackets, used throughout this chapter, are a standard notation for concentration. pK_a is the dissociation constant for any particular acid; it equals 6.1 for carbonic acid. If the normal 20:1 ratio of bicarbonate ions to carbonic acid is present, the pH will be 7.4.

$$pH = 6.1 + \log 20$$

$$pH = 6.1 + 1.3$$

$$pH = 7.4$$

The 20:1 ratio of bicarbonate ions to carbonic acid necessary for a normal pH is an important concept in understanding the compensatory mechanisms for acid–base imbalances that are discussed later in this chapter.

Respiratory Contribution

The respiratory system is the second defense against acid–base disorders. The cells continuously produce carbon dioxide (CO_2). Together, CO_2 and water (H_2O) make carbonic acid (H_2CO_3). The lungs excrete carbon dioxide and water from the body. Therefore during the process of exhalation the lungs effectively excrete carbonic acid. The respiratory system adjusts the amount of carbonic acid that remains in the body by altering rate and depth of respiration.

Rate and depth of respiration are influenced strongly by chemoreceptors that sense the $PaCO_2$, PaO_2, and pH of the blood. If too much carbonic acid begins to accumulate in the blood of a healthy person, the rate and depth of respiration increase and excess carbonic acid is removed. This response corrects the imbalance and restores blood chemistry to normal. If, on the other hand, too little carbonic acid is present in the blood, the rate and depth of respiration decrease to retain carbonic acid until it once more is present in normal amounts. Again, the imbalance is corrected and the blood chemistry returns to normal. Thus the body's correction of a carbonic acid excess or deficit requires normal function of all components of the respiratory system, including the chemoreceptors, respiratory neurons in the brainstem, motor nerves to respiratory muscles, diaphragm and other respiratory muscles, chest wall, and, of course, the airways, lungs, and pulmonary circulation.

The $PaCO_2$ indicates how effectively the respiratory system is excreting the carbonic acid that the cells are producing. If the $PaCO_2$ is elevated above the upper limit of the normal range, carbonic acid has accumulated in the blood. In other words, the respiratory rate and depth have been insufficient or lung disease has prevented sufficient carbonic acid (carbon dioxide and water) excretion. Similarly, if the $PaCO_2$ is decreased below the lower limit of the normal range, the lungs have excreted more carbonic acid than usual. In other words, the respiratory rate and depth have been excessive.

Carbonic acid is known as a volatile acid because it can be excreted as gases (CO_2 and H_2O). It is the only volatile acid in the body. Other acids that accumulate in the body, such as lactic acid and acetoacetic acid, are nonvolatile. They are organic acids that have no gaseous form. The lungs can excrete only carbonic acid; they cannot excrete nonvolatile acids that may accumulate in the body. If a nonvolatile acid (such as lactic acid) accumulates in the blood, the rate and depth of respiration will increase because the excess hydrogen ions stimulate the chemoreceptors. This hyperventilation does not excrete lactic acid (which would correct the problem), but it does remove carbonic acid from the blood. Removing carbonic acid from the blood when another acid is present in excess helps keep the pH from dropping too low. However, this response makes other values abnormal. The respiratory response to an imbalance of any acid except carbonic acid is called *compensation.* A compensatory response does not correct a pH disorder, but it does compensate for it by adjusting the pH back toward normal, even though other blood chemistry values are made abnormal in the process.

The compensatory response to a deficit of any acid except carbonic acid is hypoventilation. By decreasing rate and depth of respiration, the body retains carbonic acid. This carbonic acid accumulation helps

TABLE 25.2 Respiratory Responses to Changes in Carbonic and Metabolic Acids

Stimulus	Respiratory Response	Result
Increased $PaCO_2$, decreased pH	Hyperventilation	Correction of imbalance
Decreased $PaCO_2$, increased pH	Hypoventilation	Correction of imbalance
Decreased pH from excess of metabolic acids	Hyperventilation	Compensation for imbalance
Increased pH from deficit of metabolic acids	Hypoventilation	Compensation for imbalance

keep blood pH from rising to a fatal level when another acid is deficient in the body. Respiratory compensation for an imbalance of metabolic acid begins in minutes but may require several hours for full effectiveness. Respiratory responses to changes in carbonic and metabolic acids are summarized in Table 25.2.

Renal Contribution

The third defense against acid–base disorders is the kidneys. The kidneys can excrete any acid from the body except carbonic acid (which is excreted by the lungs). These acids that are not carbonic acid are called *metabolic acids* because cells continuously produce them during normal metabolism. The kidneys normally excrete metabolic acids. If a metabolic acid begins to accumulate in the blood, the kidneys increase their acid excretion mechanisms to correct the problem. If a metabolic acid is deficient in the blood, the kidneys slow their acid excretion mechanisms to allow acid to accumulate to normal levels. The body's ability to correct an excess or deficit of a metabolic acid depends on normal function of the renal system. Infants excrete more bicarbonate in their urine than do older children or adults; their kidneys are less effective in excreting acid. The renal response to a large acid load also is less efficient in older adults.

The kidneys have several mechanisms that accomplish acid excretion. Understanding these mechanisms requires a knowledge of basic renal physiology. Briefly, at the glomerulus, fluid filtered from the blood enters the glomerular (Bowman) capsule, which is the beginning of the nephron. The cells that line the lumen of the renal tubule modify the fluid inside the nephron (renal tubular fluid). Renal tubular fluid that passes through the entire nephron becomes the urine. Renal tubular epithelial cells have different membrane structures on opposite sides of the cells. The luminal membrane (next to the renal tubular fluid) contains different transporter proteins than the basolateral membrane (next to the interstitial fluid). This structure allows these cells to secrete certain substances into the renal tubular fluid and move other substances into the interstitial fluid.

In the proximal tubules, renal tubular epithelial cells excrete metabolic acid by secreting both the anion portion of the metabolic acid (e.g., lactate) and the hydrogen ions into the tubule lumen. For every hydrogen ion (H^+) that is secreted into the renal tubular fluid, one bicarbonate ion (HCO_3^-) is moved into the interstitial fluid. The fluid filtered from the blood at the glomerulus contains many bicarbonate ions, and most or all of that bicarbonate is reabsorbed (returned to the blood) during secretion of hydrogen ions. Renal tubular cells are able to secrete additional hydrogen ions into the tubular fluid to excrete large amounts of hydrogen ions from metabolic acid.

Once the H^+ are in the renal tubular fluid, most of them combine with other chemicals: bicarbonate ions, which were filtered at the glomerulus, as described previously, or urine buffers, such as phosphate, which were filtered at the glomerulus. Some H^+ are secreted with ammonia (NH_3), most of which is produced by renal tubular cells in the distal nephron. Net H^+ excretion occurs after HCO_3^- has been reabsorbed in the amount that was filtered at the glomerulus. Thus net H^+ excretion is accomplished in the form of buffered H^+ (called *titratable acidity*) and H^+ attached to ammonia (ammonium ions, NH_4^+). Fig. 25.1 illustrates the major processes in the proximal tubule. Some of these processes operate also in the thick ascending limb of the loop of Henle and the distal nephron, where the intracellular chemistry differs slightly but the overall processes are the same. In addition, the distal nephron is the major site of the ammonia mechanism.

When the kidneys need to excrete more hydrogen ions, renal tubular cells increase their production of ammonia (NH_3), primarily in the distal tubules and collecting ducts. Ammonia and H^+ are transported into the renal tubular fluid where they become ammonium ions (NH_4^+). Ammonium ions do not cross easily from the renal tubular fluid back to the blood. Only free hydrogen ions contribute to the acidity of the urine, not those that are part of ammonium ions. Consequently, increased production of ammonia is an effective way of excreting more hydrogen ions in the renal tubular fluid without making the urine too acidic.

The concentration of HCO_3^- in plasma reflects the effectiveness of renal regulation of metabolic acids. If metabolic acids are accumulating in the blood, they will be buffered by HCO_3^- and the HCO_3^- concentration will drop below normal. Thus a decreased concentration of HCO_3^- in plasma indicates a relative excess of metabolic acids. An increased HCO_3^- concentration in the plasma indicates a relative deficit of metabolic acids (in other words, a relative excess of base).

Although the kidneys are unable to excrete carbonic acid, they can compensate for carbonic acid imbalances by adjusting the excretion of metabolic acids. For example, if carbonic acid accumulates in the blood, the kidneys can increase the excretion of metabolic acids, reabsorbing more HCO_3^- in the process. This compensatory action helps keep the pH of the blood from becoming too abnormal. Similarly, if a deficit of carbonic acid in the blood is prolonged, the kidneys will decrease the excretion of metabolic acids, reabsorbing less HCO_3^- in the process. As these metabolic acids accumulate in the blood, they will compensate for the lack of carbonic acid and return the pH of the blood toward normal. The body's compensatory response to an imbalance of one kind of acid thus returns the pH of the blood toward normal by creating an imbalance of another kind of acid. The renal compensatory response to an imbalance of carbonic acid requires several days to be fully operative. Renal responses to changes in metabolic and carbonic acids are summarized in Table 25.3.

TABLE 25.3 Renal Responses to Changes in Metabolic and Carbonic Acids

Stimulus	Renal Response	Result
Decreased pH from excess of metabolic acids	Secrete more H^+ into renal tubules Make more ammonia	Correction of imbalance
Increased pH from deficit of metabolic acids	Secrete fewer H^+ into renal tubules Excrete HCO_3^- Make less ammonia	Correction of imbalance
Decreased pH from excess of carbonic acid	Secrete more H^+ into renal tubules Make more ammonia	Compensation for imbalance
Increased pH from deficit of carbonic acid	Secrete fewer H^+ into renal tubules Excrete HCO_3^- Make less ammonia	Compensation for imbalance

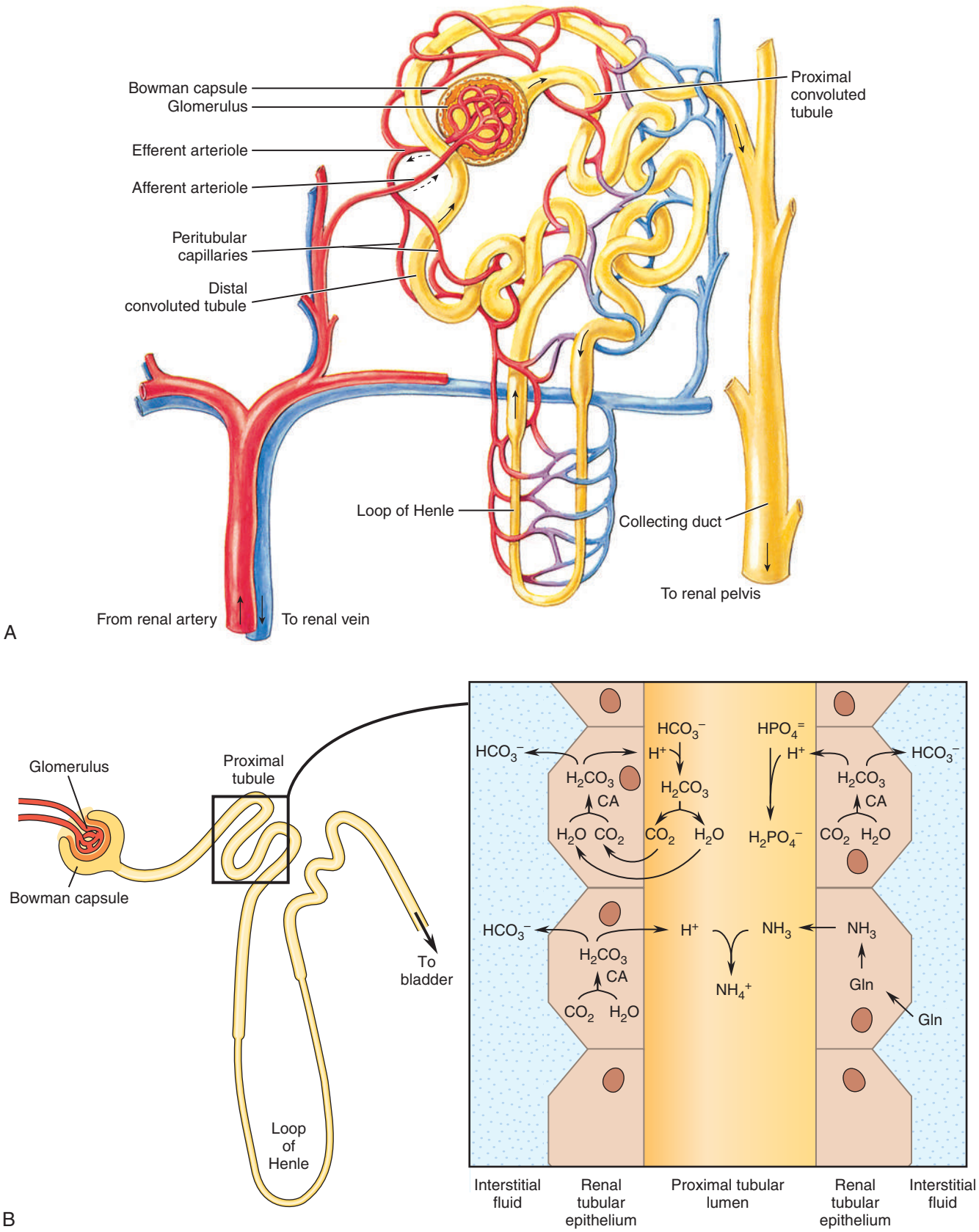


FIG 25.1 **A,** Diagram of a nephron. **B,** Renal proximal tubule mechanisms for excretion of metabolic acid. Hydrogen ions secreted into the renal tubular lumen combine with filtered bicarbonate (which then is converted to CO_2 and reabsorbed) or buffers (phosphate buffer illustrated here) that remain in the tubular fluid. *CA,* Carbonic anhydrase, an enzyme. (A, From Solomon EP: *Introduction to human anatomy and physiology,* ed 3, St Louis, MO, 2009, Mosby, p 266.)

KEY POINTS

- Normal cellular metabolism produces both carbon dioxide and metabolic acids. Carbon dioxide (CO_2) combines with water (H_2O) to form carbonic acid (H_2CO_3). Both carbonic and metabolic acids must be excreted to maintain acid–base homeostasis.
- Buffers are chemicals (a weak acid plus its base) that prevent large changes in pH by releasing or taking up hydrogen ions (H^+). The bicarbonate buffer system is the most important buffer in the extracellular fluid. The normal ratio of bicarbonate to carbonic acid is 20:1. Any deviation from this ratio alters the pH of the blood.
- The lungs excrete carbon dioxide and water (carbonic acid). Rate and depth of respiration normally are adjusted by chemoreceptors in response to acid–base and oxygen status. Increased ventilation (hyperventilation) decreases the amount of carbon dioxide in blood and thus reduces the amount of carbonic acid. Decreased ventilation (hypoventilation) allows carbon dioxide to accumulate and thus increases the amount of carbonic acid in the blood.
- The kidneys excrete metabolic acids. They can secrete H^+ into the renal tubular fluid and retain HCO_3^- in the body or may allow some HCO_3^- to be excreted, depending on homeostatic demands. Most H^+ in the urine is buffered (titratable acidity) or in the form of ammonium ions. The concentration of HCO_3^- in plasma reflects the relative amount of metabolic acid in the blood.
- The lungs compensate for acid–base imbalances resulting from altered levels of metabolic acids; the kidneys compensate for acid–base imbalances resulting from altered levels of carbonic acid. With compensation, the pH returns toward normal but $Paco_2$ and HCO_3^- levels are abnormal.

ACID–BASE IMBALANCES

The four primary acid–base disorders are metabolic acidosis, respiratory acidosis, metabolic alkalosis, and respiratory alkalosis. *Acidosis* is the presence of a condition that tends to decrease the pH of the blood below normal (make the blood relatively more acidic). If blood pH actually is decreased, *acidemia* also is present. *Alkalosis* is the presence of any factor that tends to increase the pH of the blood above normal (make the blood relatively more alkaline). The term *alkalemia* denotes an increased blood pH. The pathophysiology of the four primary acid–base disorders can be reasoned logically from the principles of acid–base homeostasis.

Metabolic Acidosis

Etiology. Metabolic acidosis is a condition that tends to cause a relative excess of any acid except carbonic acid. Metabolic acidosis may be caused by an increase in acid (not carbonic), by a decrease in base, or by a combination of the two. These mechanisms decrease the normal 20:1 ratio of HCO_3^- to H_2CO_3.

An increase of any acid except carbonic acid will decrease the normal ratio of bicarbonate to carbonic acid because the bicarbonate ions are used up in buffering the excess acid. For example, when caloric intake is insufficient, as with prolonged fasting, the body begins to use its fat stores for energy. If too little glucose is ingested, the fat is metabolized incompletely and ketoacids accumulate in the blood. This condition is termed *starvation ketoacidosis.* A similar condition, *diabetic ketoacidosis,* occurs in diabetes when insufficient insulin is available to move glucose into cells. Again, fat is metabolized incompletely, producing more ketoacids than the kidneys can excrete.

BOX 25.1 Common Causes of Metabolic Acidosis

Increase in Metabolic Acid

Ketoacidosis (diabetes mellitus, starvation, alcoholism)
Severe hyperthyroidism
Burns
Circulatory shock
Tissue anoxia (lactic acidosis)
Oliguric acute kidney injury or end-stage chronic kidney disease
Excessive intake of acids or acid precursors (e.g., salicylates, methanol, ethylene glycol)

Decrease in Base (Bicarbonate)

Severe or prolonged diarrhea
Gastrointestinal fistula that drains intestinal or pancreatic secretions
Intestinal decompression
Renal tubular acidosis

Bicarbonate ions are a type of base. Any condition that causes excessive removal of bicarbonate ions from the body may cause metabolic acidosis. For example, the intestinal fluid is rich in bicarbonate ions, which originate from pancreatic secretions. Diarrhea causes removal of this base from the body and thus contributes to the development of metabolic acidosis.

Other causes of metabolic acidosis are listed in Box 25.1 under the two general mechanisms discussed: increase in metabolic acid (any acid except carbonic acid) and decrease in base (bicarbonate). Either mechanism tends to make the blood overly acidic. The pathophysiology of diarrhea and other disorders that may cause metabolic acidosis is discussed in other chapters of this text.

Clinical manifestations. Signs and symptoms of metabolic acidosis include headache, abdominal pain, and central nervous system depression (confusion, lethargy, stupor, coma).

Central nervous system depression that occurs in patients with metabolic acidosis is due primarily to the decreased pH of the cerebrospinal and interstitial fluid in the brain. When the pH of the interstitial fluid falls, intracellular pH decreases, the protein structure and enzyme activity in cells are altered, and cellular dysfunction results. Other factors specific to the cause of the acidosis also may induce central nervous system depression, such as hyperosmolality with diabetic ketoacidosis. Severe metabolic acidosis predisposes to tachycardia, ventricular dysrhythmias (from myocardial intracellular acidity), and decreased cardiac contractility, which may be fatal. Death from brainstem dysfunction usually occurs when the pH falls below 6.9.

Arterial blood gases in metabolic acidosis show a bicarbonate concentration below normal. If metabolic acidosis is uncompensated, the pH also is below normal because the usual 20:1 ratio is decreased.

Uncompensated metabolic acidosis:

$$\frac{\text{Decreased}\,[HCO_3^-]}{\text{Unchanged}\,[H_2CO_3]} = \text{pH low}$$

Compensatory response. The respiratory compensation for metabolic acidosis is hyperventilation. The low blood pH stimulates the peripheral chemoreceptors, which then stimulate ventilatory neurons in the brainstem. The end result is increased rate and depth of respiration. As the rate and depth of respiration increase, more carbonic acid (carbon dioxide and water) is excreted. Although hyperventilation does not

remove metabolic acid from the body, it does change the ratio of bicarbonate ions to carbonic acid in a favorable direction. Because the bicarbonate ion concentration already is decreased by the metabolic acidosis, the compensatory decrease in carbonic acid brings the ratio (and thus the pH) back toward normal.

The arterial blood gases of a person who has compensated metabolic acidosis show decreased bicarbonate concentration (the primary imbalance), decreased $PaCO_2$ (compensation), and decreased or even normal pH, depending on the degree of compensation. A flowchart for interpreting laboratory measures specific to acid–base imbalances is presented in Fig. 25.2. Sample laboratory values for people with metabolic acidosis are presented in Table 25.4.

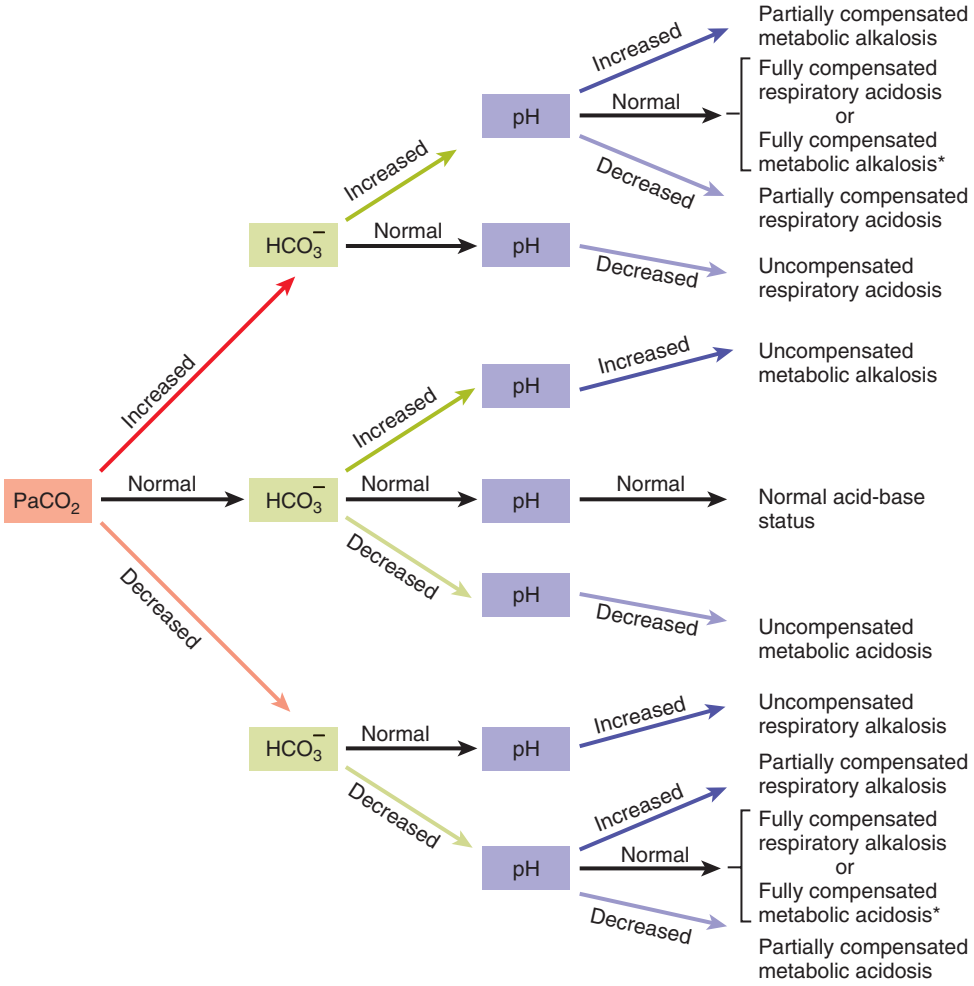
Compensated metabolic acidosis:

$$\frac{\text{Decreased } [HCO_3^-] \text{ (primary)}}{\text{Decreased } [H_2CO_3] \text{ (compensatory)}} = \text{pH somewhat low (partially compensated) } \textit{or} \text{ pH in the normal range (fully compensated)}$$

Respiratory Acidosis

Etiology. Respiratory acidosis is a condition that tends to cause an excess of carbonic acid. This condition is aptly named because carbonic acid is excreted by the lungs in the form of carbon dioxide and water during exhalation.

Respiratory acidosis is caused by factors that impair the respiratory excretion of carbonic acid. Such factors include impaired gas exchange, inadequate neuromuscular function, and impairment of respiratory control in the brainstem. Box 25.2 provides examples of factors that may cause respiratory acidosis. These factors all decrease the normal 20:1 ratio of bicarbonate ion to carbonic acid (and thus decrease the pH of the blood) by increasing the carbonic acid portion of the ratio. Chronic respiratory acidosis often develops in people who have type B chronic obstructive pulmonary disease (COPD). If an acute respiratory infection also develops, the acidosis may worsen. Such a condition is termed *acute-on-chronic respiratory acidosis.* The pathophysiology of COPD and other disorders that may cause respiratory acidosis is discussed in other chapters of this text.



*To differentiate between possible fully compensated imbalances with the pH in the normal range, look at the previous laboratory values for the patient. If no previous values are available, choose the acidosis if the pH is below 7.40 and the alkalosis if the pH is above 7.40.

FIG 25.2 Flowchart for interpretation of laboratory measurements specific for acid–base imbalances. Use this flowchart to determine the primary acid–base imbalance from a set of laboratory values. Begin on the left with $PaCO_2$ and follow the arrows. This flowchart does not include mixed acid–base imbalances.

TABLE 25.4 Sample Laboratory Values for People With Acid–Base Imbalances

Laboratory Value for Imbalance	Explanation
Partially Compensated Metabolic Acidosis (Diabetic Ketoacidosis)	
$PaCO_2$ 30 mm Hg	Decreased because of compensatory hyperventilation
HCO_3^- 12 mEq/L	Decreased because of buffering of ketoacids
pH 7.22	Decreased because of excess metabolic acids; would be even lower without respiratory compensation
Uncompensated Respiratory Acidosis (Acute Asthma Episode)	
$PaCO_2$ 55 mm Hg	Increased because of impaired gas exchange
HCO_3^- 24 mEq/L	Normal; renal compensation has not yet occurred in this acute condition
pH 7.26	Decreased because of excess carbonic acid
Fully Compensated Respiratory Acidosis (Type B COPD)	
$PaCO_2$ 60 mm Hg	Increased because of impaired gas exchange
HCO_3^- 36 mEq/L	Increased because of renal compensation in this chronic condition
pH 7.35	Normal because of renal compensation, but below 7.4
Uncompensated Metabolic Alkalosis (Repeated Emesis and ECV Depletion)	
$PaCO_2$ 42 mm Hg	Normal, but increasing because of compensatory hypoventilation
HCO_3^- 36 mEq/L	Increased because of loss of H^+ from emesis and renal retention of HCO_3^- from ECV depletion
pH 7.52	Increased because of metabolic acid deficit
Uncompensated Respiratory Alkalosis (Hypoxemia From Pulmonary Embolism)	
$PaCO_2$ 28 mm Hg	Decreased because of hyperventilation caused by chemoreceptor response to decreased PaO_2
HCO_3^- 24 mEq/L	Normal; renal compensation has not yet occurred in this acute condition
pH 7.52	Increased because of carbonic acid deficit

ECV, Extracellular fluid volume.

Clinical manifestations. Signs and symptoms of respiratory acidosis are headache, tachycardia, cardiac dysrhythmias, and neurologic abnormalities such as blurred vision, tremors, vertigo, disorientation, lethargy, or somnolence.

Headache occurs because of dilation of blood vessels in the brain. This cerebral vasodilation increases cerebrospinal fluid pressure; papilledema may result. Neurologic manifestations usually are more prominent in patients with respiratory acidosis than in those with metabolic acidosis because carbonic acid (in the form of carbon dioxide and water) crosses the blood–brain barrier relatively easily. The neurologic manifestations are due to the decreased pH of the cerebrospinal fluid and interstitial fluid in the brain. This decreased interstitial fluid pH causes decreased intracellular pH, with resulting cellular dysfunction. Cardiac dysrhythmias in patients with respiratory acidosis occur because of decreased pH inside myocardial cells. Severe respiratory acidosis causes peripheral vasodilation, and hypotension may result, especially if cardiac dysrhythmias also are present.

BOX 25.2 Common Causes of Respiratory Acidosis (Hypoventilation)

Impaired Gas Exchange
Type B chronic obstructive pulmonary disease (COPD)
End-stage type A COPD
Bacterial pneumonia
Severe asthma episode
Pulmonary edema
Acute (adult) respiratory distress syndrome

Impaired Neuromuscular Function
Guillain–Barré syndrome
Chest injury or surgery (pain limits ventilation)
Hypokalemic respiratory muscle weakness
Severe kyphoscoliosis
Respiratory muscle fatigue

Impaired Respiratory Control (Brainstem)
Respiratory depressant drugs (opioids, barbiturates)

Arterial blood gases in patients with respiratory acidosis show $PaCO_2$ above normal. If respiratory acidosis is uncompensated, the pH is below normal because the usual 20:1 ratio is decreased.

Uncompensated respiratory acidosis:

$$\frac{\text{Unchanged } [HCO_3^-]}{\text{Increased } [H_2CO_3]} = \text{pH low}$$

Compensatory response. The compensatory response to respiratory acidosis is increased renal excretion of metabolic acid. This mechanism requires several days to be effective. Although the kidneys cannot excrete carbonic acid, their ability to excrete more metabolic acid changes the ratio of bicarbonate ions to carbonic acid in a favorable direction so that the pH moves toward normal. As the kidneys excrete more metabolic acid, the bicarbonate concentration increases because fewer bicarbonate ions are used for buffering. Because carbonic acid concentration already is increased, the compensatory increase in bicarbonate concentration tends to normalize the ratio of HCO_3^- to H_2CO_3. The arterial blood gases of a person who has compensated respiratory acidosis show increased $PaCO_2$ (the primary imbalance), increased bicarbonate concentration (compensation), and decreased or even normal pH, depending on the degree of compensation. Table 25.4 presents sample laboratory values for people with respiratory acidosis.

Compensated respiratory acidosis:

$$\frac{\text{Increased } [HCO_3^-] \text{ (compensatory)}}{\text{Increased } [H_2CO_3] \text{ (primary)}} = \text{pH somewhat low (partially compensated) } \textit{or} \text{ pH in the normal range (fully compensated)}$$

Metabolic Alkalosis

Etiology. Metabolic alkalosis is a condition that tends to cause a relative deficit of any acid except carbonic acid. Metabolic alkalosis may be caused by an increase in base (bicarbonate), by a decrease in acid, or by a combination of the two. Bicarbonate may be ingested in antacids such as baking soda and over-the-counter bicarbonate products (e.g., effervescent antacids). With overuse of these agents, enough bicarbonate is absorbed from the gastrointestinal tract to increase the blood bicarbonate concentration, thus increasing the pH.

In addition to a gain in bicarbonate, metabolic alkalosis may be caused by a decrease in acid. The stomach is a major reservoir of acid. Emesis and gastric suction remove acid from the body and create a relative excess of base; this situation is, by definition, metabolic alkalosis. Increased renal excretion of acid with retention of bicarbonate occurs in extracellular fluid volume deficit. Metabolic alkalosis caused by mild or moderate extracellular fluid volume deficit often is called *contraction alkalosis* and is common during diuretic therapy. (Severe extracellular fluid volume deficit is associated with tissue anoxia, circulatory shock, and metabolic acidosis.) Hypokalemia causes metabolic alkalosis by shifting hydrogen ions into cells and increasing renal excretion of acid.

Causes of metabolic alkalosis are summarized in Box 25.3. The pathophysiology of the disorders that may cause metabolic alkalosis is discussed in other chapters of this text.

Clinical manifestations. Signs and symptoms in patients who have metabolic alkalosis may arise from the extracellular fluid volume depletion that caused the alkalosis. Thus postural hypotension may be present. Hypokalemia frequently coexists with metabolic alkalosis. As described previously, hypokalemia may cause metabolic alkalosis. In addition, metabolic alkalosis that arises from another cause frequently induces hypokalemia by shifting potassium ions into cells. Regardless of whether the hypokalemia preceded or followed the metabolic alkalosis, the bilateral muscle weakness of hypokalemia frequently is evident in individuals who have metabolic alkalosis.

In people who experience signs and symptoms from the metabolic alkalosis itself, the initial manifestations are those of increased neuromuscular excitability. Fingers and toes may tingle; signs of tetany may progress to seizures. Increased interstitial pH causes increased excitability of nerve cell membranes. In addition, alkalosis causes more ionized calcium to bind to albumin, thus causing an ionized hypocalcemia that contributes to increased neuromuscular excitability. People who develop metabolic alkalosis may become quite belligerent. With severe metabolic alkalosis, initial excitation may change to central nervous system depression. Confusion, lethargy, and coma may ensue from dysfunction of brain cells. Death usually occurs when the pH is around 7.8. The plasma bicarbonate concentration is elevated in people who have metabolic alkalosis.

Uncompensated metabolic alkalosis:

$$\frac{\text{Increased } [HCO_3^-]}{\text{Unchanged } [H_2CO_3]} = \text{pH high}$$

Compensatory response. The compensatory response to metabolic alkalosis is hypoventilation. This shallow breathing retains carbonic acid within the body, thus increasing the lower portion of the bicarbonate ion to carbonic acid ratio. Because the upper portion of the ratio has been increased by the elevated bicarbonate concentration of metabolic alkalosis, the respiratory compensation tends to move the pH toward normal. However, respiratory compensation for metabolic alkalosis usually is incomplete. The need for oxygen drives ventilation, even though the increased pH tends to depress it. Thus the arterial blood gases of a person who has compensated metabolic alkalosis usually show increased bicarbonate concentration (the primary imbalance), increased $PaCO_2$ (compensation), and increased pH. Table 25.4 presents sample laboratory values for people with metabolic alkalosis.

Compensated metabolic alkalosis:

$$\frac{\text{Increased } [HCO_3^-] \text{ (primary)}}{\text{Increased } [H_2CO_3] \text{ (compensatory)}} = \text{pH somewhat high (partially compensated)}$$

Respiratory Alkalosis

Etiology. Respiratory alkalosis is a condition that tends to cause a carbonic acid deficit. With a deficit of carbonic acid, the blood is relatively too alkaline.

Respiratory alkalosis is caused by hyperventilation. Carbonic acid is excreted during exhalation; when respirations are excessively rapid and deep (hyperventilation), too much carbonic acid is excreted. The resulting deficit of carbonic acid is respiratory alkalosis. For example, in gram-negative sepsis, respiratory neurons in the brainstem often are stimulated abnormally, causing hyperventilation. Hypoxemia, acute pain, and psychological distress are important clinical causes of hyperventilation that leads to respiratory alkalosis. Other causes of hyperventilation (and thus of respiratory alkalosis) are listed in Box 25.4.

Clinical manifestations. Clinical manifestations of respiratory alkalosis arise from increased neuromuscular excitability. Paresthesias (numbness and tingling) often occur in the fingers and around the mouth; carpal and pedal spasms may occur. Increased extracellular pH has a direct effect of increasing membrane excitability in both central and peripheral neurons. In addition, increased pH of the cerebrospinal fluid and cerebral interstitial fluid alters brain cell function, causing excitation and/or confusion in some people. Respiratory alkalosis causes cerebral vasoconstriction, which reduces blood flow in the brain. The alkalosis also decreases the availability of ionized calcium, which contributes to increased neuromuscular excitability.

The increased excretion of carbonic acid in people with respiratory alkalosis causes the $PaCO_2$ to be abnormally low. If the imbalance is uncompensated, the pH is abnormally high.

Uncompensated respiratory alkalosis:

$$\frac{\text{Unchanged } [HCO_3^-]}{\text{Decreased } [H_2CO_3]} = \text{pH high}$$

BOX 25.3 Common Causes of Metabolic Alkalosis

Increase in Base (Bicarbonate)

Excessive intake of bicarbonate or bicarbonate precursors (acetate, citrate, lactate)
Massive transfusion with citrated blood
Mild or moderate extracellular fluid volume deficit

Decrease in Metabolic Acid

Emesis
Gastric suction
Mild or moderate extracellular fluid volume deficit
Hyperaldosteronism
Hypokalemia

BOX 25.4 Common Causes of Respiratory Alkalosis (Hyperventilation)

Hypoxemia
Acute pain
Anxiety, psychological distress
Prolonged sobbing
Initial portion of acute asthma episode
Alcohol withdrawal
Stimulation of the brainstem (salicylate overdose, meningitis, head injury, gram-negative sepsis)

Compensatory response. The compensatory response to respiratory alkalosis is decreased renal excretion of metabolic acid. As metabolic acids accumulate in the blood, the bicarbonate ion concentration decreases because bicarbonate ions are used for buffering. Because the carbonic acid concentration already is decreased, renal compensation for respiratory alkalosis tends to return the ratio of bicarbonate ions to carbonic acid, and thus the pH, toward normal. Renal compensatory mechanisms take several days to be fully effective. Many of the causes of respiratory alkalosis, such as acute hypoxemia, pain, and psychological distress, are short lived; for that reason, they may not be compensated renally. Arterial blood gases of a person who has compensated respiratory alkalosis show decreased $PaCO_2$ (the primary imbalance), decreased bicarbonate concentration (compensation), and increased or (rarely) normal pH, depending on the degree of compensation. Table 25.4 presents sample laboratory values for people with respiratory alkalosis.

Compensated respiratory alkalosis:

$$\frac{\text{Decreased } [HCO_3^-] \text{ (compensatory)}}{\text{Decreased } [H_2CO_3] \text{ (primary)}} = \text{pH somewhat high (partially compensated) } \textit{or} \text{ pH in the normal range (fully compensated)}$$

Mixed Acid–Base Imbalances

In most people, only one of the four primary imbalances discussed in this chapter arises at a time. If the imbalance persists, a compensatory imbalance arises as well. This situation was discussed previously in this chapter. Occasionally, however, two primary imbalances arise in the same person. This latter situation is termed a *mixed acid–base imbalance.* For example, a patient who has bacterial pneumonia may develop respiratory acidosis. If severe *Clostridium difficile*–associated diarrhea develops during antibiotic therapy for the pneumonia, a concurrent metabolic acidosis may arise. In this mixed imbalance, the pH is likely to be very low because the two types of primary acidosis impair the effectiveness of the usual compensatory mechanisms. Specifically, the usual compensatory mechanism for metabolic acidosis is hyperventilation, which causes increased excretion of carbonic acid from the body. With bacterial pneumonia, however, the effectiveness of alveolar ventilation already is impaired and carbonic acid is being retained in the blood. Analogously, patients who have both types of primary alkalosis often have a very high pH because their usual compensatory mechanisms are impeded by the concurrent acid–base disorders.

Mixed acid–base disorders may also occur with a nearly normal pH if a primary acidosis and a primary alkalosis are involved. An example of this type of mixed disorder is a head-injured patient whose treatment includes hyperventilation by mechanical ventilation to reduce intracranial pressure (respiratory alkalosis) but who at the same time has a metabolic acidosis from oliguric acute kidney injury. In this situation, the $PaCO_2$ is decreased (respiratory alkalosis), the plasma bicarbonate concentration is decreased (metabolic acidosis), and the pH depends on the relative severity of the two imbalances.

KEY POINTS

- Acidosis is a condition that tends to cause a relative excess of acid. Alkalosis is a condition that tends to cause a relative excess of base (bicarbonate).
- Metabolic acidosis is characterized by pH below 7.40 and abnormally low HCO_3^- concentration. It arises from processes that lead to metabolic acid accumulation (e.g., lactic acidosis) or loss of HCO_3^- (e.g., diarrhea). Compensatory hyperventilation decreases the $PaCO_2$.
- Metabolic alkalosis is characterized by pH above 7.40 and abnormally high HCO_3^- concentration. It arises from processes that lead to metabolic acid loss (e.g., vomiting) or gain of HCO_3^- (e.g., bicarbonate antacids). Compensatory hypoventilation increases the $PaCO_2$.
- Respiratory acidosis is characterized by pH below 7.40 and abnormally high $PaCO_2$. It arises from processes that lead to hypoventilation by impairing gas exchange (e.g., lung diseases), neuromuscular function of the chest (e.g., hypokalemic muscle paralysis), or respiratory control mechanisms in the brainstem (e.g., opioid overdose). Compensatory excretion of H^+ and retention of HCO_3^- by the kidneys increase the HCO_3^- concentration.
- Respiratory alkalosis is characterized by pH above 7.40 and abnormally low $PaCO_2$. It arises from processes that lead to hyperventilation (e.g., hypoxemia, pain, anxiety). Compensatory retention of H^+ and excretion of HCO_3^- by the kidneys decrease the HCO_3^- concentration.
- Mixed acid–base disorders occur when two primary acid–base disorders are present independently. They may arise from simultaneous dysfunction of the respiratory system and kidneys. Depending on the combination of disorders, the pH may be nearly normal or grossly abnormal.

SUMMARY

Acid–base homeostasis involves the interplay of buffers, the respiratory system, and renal mechanisms. Metabolic acids are produced continually by cellular metabolism. These metabolic acids enter the blood, where they are buffered, and eventually are excreted by the kidneys. In healthy people, the kidneys adjust the rate of excretion of metabolic acids to meet the demands of the acid load being produced. The concentration of bicarbonate ions in the blood indicates the effectiveness of renal excretion of metabolic acids. Cellular metabolism also constantly produces carbonic acid (carbon dioxide and water) that is excreted by the lungs. In healthy people, changes in the respiratory rate and depth adjust the rate of excretion of carbonic acid appropriately. The $PaCO_2$ indicates the effectiveness of respiratory excretion of carbonic acid.

If one of the two acid excretion mechanisms becomes dysfunctional or overwhelmed, the other mechanism can produce a compensatory response that will help normalize the pH of the extracellular fluid, even though it will not correct the acid–base imbalance. Thus the kidneys adjust their excretion of metabolic acids when respiratory excretion of carbonic acid is altered abnormally. Similarly, the respiratory system adjusts the rate of excretion of carbonic acid if renal excretion of metabolic acids is impaired or overwhelmed. The pH of the blood at any time is the net result of the operation of these regulatory and compensatory mechanisms.

Primary acid–base imbalances arise when the normal regulatory mechanisms for acid–base homeostasis become impaired or are overwhelmed by a large acid or alkaline load. Pediatric and geriatric considerations are summarized in the following boxes. Primary metabolic acidosis arises when the kidneys are unable to excrete enough metabolic acid or bicarbonate is lost from the body. The compensatory response to metabolic acidosis is hyperventilation. Primary respiratory acidosis arises when the lungs are unable to excrete enough carbonic acid. The compensatory response to respiratory acidosis is increased renal excretion of metabolic acid.

Primary metabolic alkalosis arises when the kidneys excrete too much metabolic acid or there is a gain of bicarbonate. The compensatory response to metabolic alkalosis is hypoventilation. Primary respiratory alkalosis arises when the lungs excrete too much carbonic acid. The compensatory response to sustained respiratory alkalosis is decreased renal excretion of metabolic acid. The $PaCO_2$ reflects the respiratory component of an acid–base imbalance; the plasma bicarbonate concentration reflects the metabolic (renal) component of an acid–base imbalance.

A mixed acid–base imbalance occurs when two primary imbalances exist at the same time. The two primary imbalances may drive the pH to an extremely abnormal value or may nearly cancel each other's effect on the pH, although the $PaCO_2$ and plasma bicarbonate concentration may still be very abnormal.

PEDIATRIC CONSIDERATIONS

Acid–Base Imbalance

- Neonates often have mild metabolic acidosis. Infants younger than 1 month have a reduced ability to excrete a large acid load; their kidneys are less able to reabsorb bicarbonate, they produce less ammonia, and urinary buffers are limited in quantity. These factors increase the risk of metabolic acidosis from acid accumulation.
- Adolescents with eating disorders may develop metabolic alkalosis from repeated emesis and hypokalemia or metabolic acidosis from starvation and laxative-induced chronic diarrhea.

GERIATRIC CONSIDERATIONS

Acid–Base Imbalance

- Older adults are at increased risk of respiratory depression (and thus respiratory acidosis) from some opioids, due to decreased renal excretion of active metabolites.
- Older adults' kidneys are less able to excrete a large acid load, which increases the risk for metabolic acidosis from acid accumulation.
- Diarrhea from chronic laxative overuse may contribute to metabolic acidosis.

RESOURCES

Berend K, et al: Physiological approach to assessment of acid-base disturbances. *New Engl J Med* 371(15):1434–1445, 2014.

Carrera E, Kim DJ, Castellani G, et al: Effect of hyper- and hypocapnia on cerebral arterial compliance in normal subjects. *J Neuroimaging* 21:121–125, 2011.

Espay AJ: Neurologic complications of electrolyte disturbances and acid-base balance. *Handb Clin Neurol* 119:365–382, 2014.

Felver L: Acid-base balance. In Giddens J, editor: *Concepts for nursing practice*, ed 2, Philadelphia, 2017, Elsevier, pp 75–84.

Hale A, Hovey MJ: *Fluid, electrolyte, and acid-base imbalances*, Philadelphia, 2014, F.A. Davis.

Hall JE: *Guyton and Hall textbook of medical physiology*, ed 13, Philadelphia, 2016, Elsevier.

Ho CH, Lewis KE, Johnson JL, et al: A 20-year-old woman with fatigue and palpitations. *Cleveland Clin J Med* 81:283–288, 2014.

Kamel KS, Halperin MJ: Acid-base problems in diabetic ketoacidosis. *New Engl J Med* 372:546–554, 2015.

Kamel KS, Halperin ML: *Fluid, electrolyte and acid-base physiology: a problem-based approach*, ed 5, St Louis, MO, 2016, Elsevier.

Karpate SJ, Morsi H, Shehmar M, et al: Euglycemic ketoacidosis in pregnancy and its management: case report and review of literature. *Eur J Obstet Gynecol Reprod Biol* 171:386–387, 2013.

Kraut JA, Madias NE: Lactic acidosis. *New Engl J Med* 371:2309–2319, 2014.

Marston N, Devin Kehl D, Copp J, et al: Alkalotics anonymous: severe metabolic alkalosis. *Am J Med* 127(1):25–27, 2014.

Namakura M, Shirai A, Yamakazi O, et al: Role of renal proximal tubule transport in acid/base and blood pressure regulation. *BioMed Res International* 2014:article ID 504808, 2014.

Rice M, Ismail B, Tyson Pillow M: Approach to metabolic acidosis in the emergency department. *Emerg Med Clin N Am* 32:403–420, 2014.

Rogers KMA, McCutcheon K: Four steps to interpreting arterial blood gases. *J Perioperative Practice* 25(3):46–52, 2015.

Rose BD, Post T, Stokes J: *Clinical physiology of acid-base and electrolyte disorders*, ed 6, New York, 2017, McGraw-Hill.

Soifer JT, Kim HT: Approach to metabolic alkalosis. *Emerg Med Clin N Am* 32:453–463, 2014.

Weiner D, Verlander JW: Ammonia transport in the kidney by Rhesus glycoproteins. *Am J Physiol Renal Physiol* 306:F1107–F1120, 2014.

26

Renal Function

Jacquelyn L. Banasik

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- How does the structure of the glomerulus determine the composition of the tubular filtrate?
- What factors determine glomerular filtration rate?
- How do individual nephrons regulate their glomerular filtration rates?
- How are solutes and water transported across the renal tubular epithelium?
- What is the role of the kidney in fluid, electrolyte, and acid–base balance?
- How does renal function change across the life span?
- How are laboratory and diagnostic tests used to evaluate renal function and disease?

CHAPTER OUTLINE

Renal Anatomy, 552
Renal Parenchyma, 552
Renal Lymphatics and Innervation, 552
Renal Blood Supply, 553
Overview of Nephron Structure and Function, 554
Glomerulus, 555
Proximal Convoluted Tubule, 557
Loop of Henle, 558
Distal Convoluted Tubule, 559
Collecting Duct, 559
Regulation of Glomerular Filtration, 560
Physics of Filtration, 560
Factors Affecting Filtration Pressure, 562
Tubuloglomerular Feedback, 563
Effects of Glucose and Amino Acids, 563
Role of Mesangial Cells, 564
Transport Across Renal Tubules, 565
Reabsorption of Glucose, 565
Regulation of Acid–Base Balance, 565
Renal Compensation Process, 566
Secretion of Potassium, 566
Regulation of Blood Volume and Osmolality, 567
Antidiuretic Hormone, 567
Aldosterone, Angiotensin II, Natriuretic Peptides, Urodilatin, Uroguanylin, and Guanylin, 568
Diuretic Agents, 569
Endocrine Functions, 570
Erythropoietin, 570
Vitamin D, 570
Age-Related Changes in Renal Function, 570
Infant, 570
Adult and Elderly, 570
Tests of Renal Structure and Function, 570
Urine and Blood Studies, 570
Urinalysis, 571
Serum Creatinine and Blood Urea Nitrogen, 572
Measures of Glomerular Filtration Rate, 572
Diagnostic Tests, 572
Kidney, Ureter, and Bladder Roentgenography, 572
Intravenous Urography/Pyelography, 572
Radionuclide Studies, 573
Ultrasonography, 573
Computed Tomography, 573
Magnetic Resonance Imaging, 573
Renal Biopsy, 573

The kidneys are responsible for maintaining fluid and electrolyte homeostasis and ridding the body of water-soluble wastes. To accomplish these functions, the kidneys filter more than 7 L of fluid per hour, then reabsorb about 99%, producing a small amount of urine containing a high concentration of wastes. The kidneys can alter the amount and composition of urine to keep blood volume and electrolyte composition within normal limits. In addition, the kidneys perform two important endocrine functions: production of erythropoietin, which is a regulator of red blood cell (RBC) quantity, and activation of vitamin D, which is a cofactor for intestinal calcium absorption.

Most individuals have two kidneys, each containing approximately 1 million nephrons, which provide a large renal reserve. The nephron is the functional unit of the kidney, performing all filtration, reabsorption, and secretory functions. Removal of 50% of a person's nephrons, as occurs with kidney donation, results in no immediate significant impairment of renal function, although renal reserve is reduced and long-term risk for kidney impairment may be increased. Serious renal impairment generally does not occur until between 75% and 90% of the total nephrons have been damaged. Thus clinical findings may not be evident until late in the course of chronic kidney disease. A number of laboratory and diagnostic tests are used to assess renal structure and function and to identify disease processes. These are briefly described at the end of this chapter. Kidney diseases, renal failure, and abnormalities of the bladder are discussed in Chapters 27, 28, and 29, respectively. A discussion of fluid and electrolyte imbalances and acid–base disturbances can be found in Chapters 24 and 25, respectively. The essentials of kidney structure and nephron function are presented in this chapter.

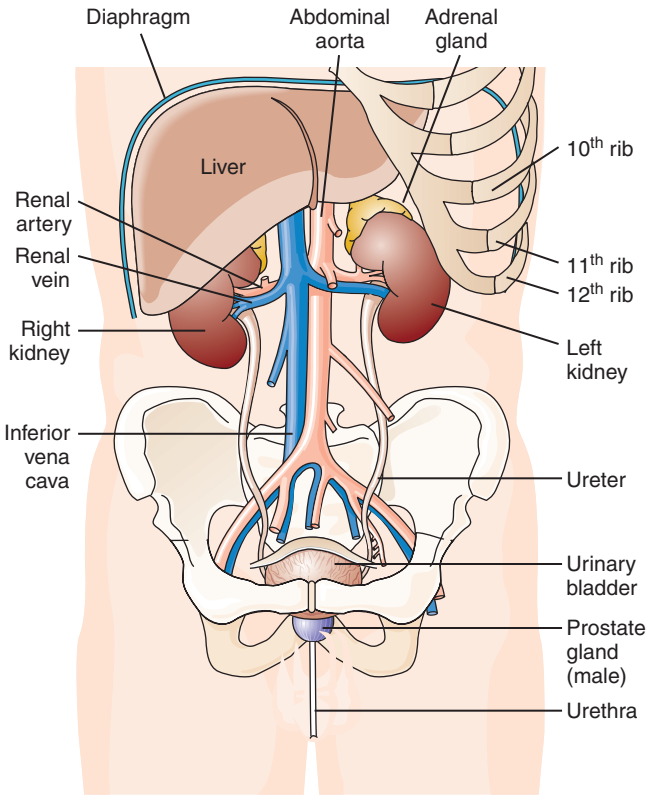


FIG 26.1 Structure of the urinary tract. The kidneys are located in the retroperitoneal space in the posterior abdominal cavity, in contact with the diaphragm and covered on the upper portions by ribs.

RENAL ANATOMY

The urinary system consists of the kidneys, ureters, urinary bladder, and urethra (Fig. 26.1). The kidneys are located in the retroperitoneal space in the posterior abdomen. One kidney is on each side of the vertebral column between the level of the twelfth thoracic and third lumbar vertebrae. The costovertebral angle (CVA), the point at which the bottom of the ribcage meets the spine, is commonly used as an external landmark for finding kidney position during physical examination. The right kidney is located beneath the liver and is placed slightly lower than the left kidney.

The kidneys are protected and surrounded by strong back and flank muscles, fascia, and fat. The kidneys are somewhat mobile and can be injured by high-impact activities, such as bouncing along on horseback or on a mountain bike, or by direct trauma, as might occur from falls or blunt trauma. Kidney hemorrhage results in bleeding into the retroperitoneal space, but not into the peritoneal cavity.

The kidneys drain urine into the ureters by gravity flow, and the ureters provide peristaltic action to move urine along to the bladder where it is stored. The two principal parts of the bladder are the body and the neck. The body stores urine and is made up of smooth muscle known as *detrusor muscle.* Detrusor muscle extends in all directions throughout the bladder and contracts as a unit in response to initiation of action potentials. The urinary bladder collects 300 to 500 mL before stretch receptors signal a need for bladder emptying. Urine is drained from the bladder by the urethra when the internal and external sphincters are relaxed. Innervation and control of bladder function are discussed in detail in Chapter 29.

An adult kidney weighs approximately 115 to 170 g; is 11 cm long, 6 cm wide, and 3 cm thick; and is shaped like a red kidney bean, with the concave portion, termed the hilum, facing the vertebral column. A thin, fibrous capsule covers each kidney and encloses blood vessels, lymphatic vessels, and nerve fibers, including pain receptors. Lymphatic vessels, blood vessels, and nerves enter and exit the kidney through the hilum.

Renal Parenchyma

On cross-section, the kidney is seen to contain three principal areas: the pelvis, the medulla, and the cortex (Fig. 26.2). The renal pelvis is a large collecting area for the urine that drains from the many collecting ducts of the nephrons. The minor (smaller) calices collect urine as it drains from the papilla of the renal pyramids. The normal kidney has 8 to 18 minor calices and 2 to 3 major calices. The major calices are large collecting spaces located between the minor calices and the upper part of the ureter.

The medulla contains 8 to 18 renal pyramids, the bases of which are adjacent to the outer cortex, whereas the apices open into the minor calices. The pyramids consist of collecting tubules, collecting ducts, long loops of Henle, and vasa recta. The papillae are the openings at the tips of the renal pyramids through which urine exits the collecting ducts.

The renal cortex, which is the outer rim of the kidney, is about 1 cm thick. The cortex contains all of the glomeruli as well as 85% of the nephron tubules. Fifteen percent of nephrons send their loops of Henle deep into the medulla and are called juxtamedullary nephrons. Columns of cortical tissue are found between the medullary pyramids and provide the passageway for the interlobar arteries.

Renal Lymphatics and Innervation

There are two lymphatic systems in the kidney. One system is composed of vessels that are located both in the renal capsule and immediately under the capsule in the outer cortex. The other lymphatic system is composed of vessels that accompany and wrap around the arterial blood

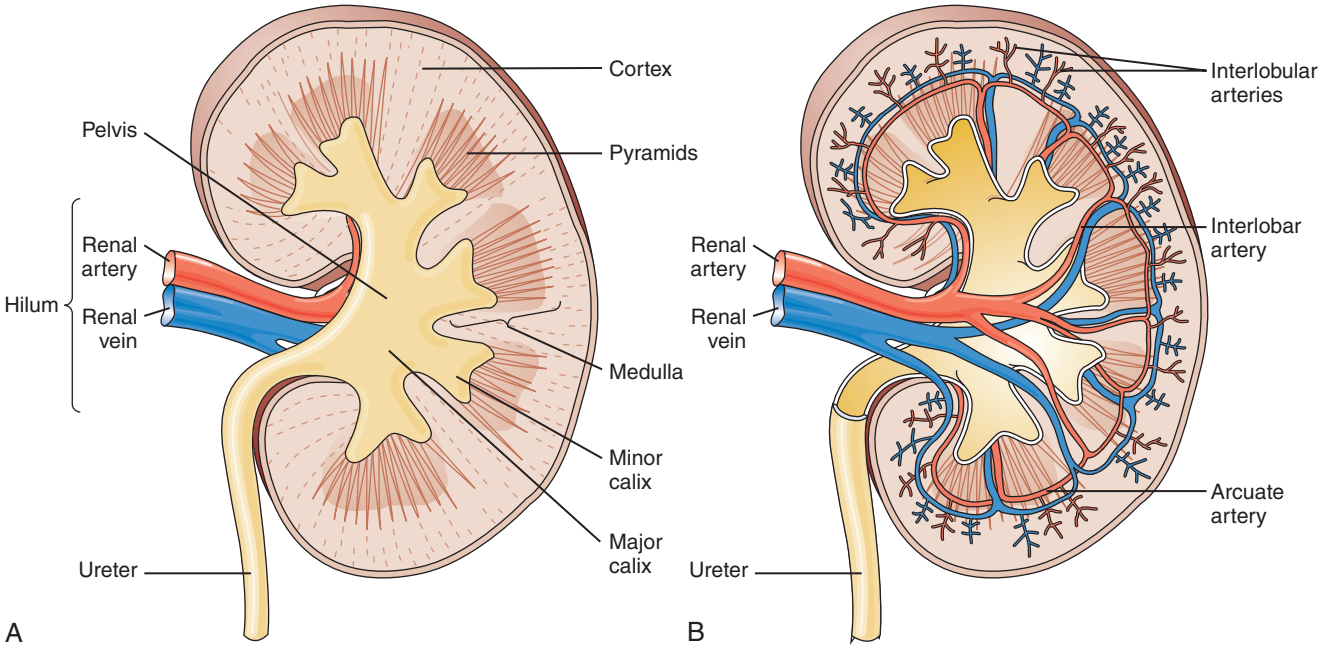


FIG 26.2 **A,** Cross-section of the kidney showing the renal pelvis, medullary pyramids, and cortex. Normal kidneys have 8 to 18 renal pyramids and a corresponding number of minor calices. The major calices drain urine into the ureter. Blood vessels, lymphatic vessels, and nerves enter and exit through the hilum. **B,** The arterial blood supply to the kidney is derived from the renal arteries, which branch from the abdominal aorta and enter the kidney through the hilus. The renal artery branches to form several interlobar arteries, which travel toward the cortex in the renal columns. The interlobar arteries branch to form the arcuate arteries, which divide further to form the interlobular arteries. Interlobular arteries branch multiple times to provide the afferent arterioles for each of the kidney's million nephrons.

vessels. All the lymphatic vessels, as well as blood vessels and nerves, exit the kidney through the hilum, and lymph drains into the paraaortic lymph nodes.

The kidneys are innervated by the sympathetic division of the autonomic nervous system. The lesser splanchnic nerves come from the renal plexuses, which are located next to the renal arteries. These nerve fibers travel with the renal arterial blood vessels and terminate in smooth muscle of the afferent and efferent arterioles, proximal and distal tubules, and the renin-secreting juxtaglomerular cells. Stimulation of the sympathetic nervous system results in renal vasoconstriction and renin release. The renal capsule and all structures between the renal pelvis and urinary meatus are innervated with pain receptors (see Chapter 27).

Renal Blood Supply

Approximately 25% of the cardiac output is delivered to the kidneys, the majority of which circulates through the cortex, whereas only 1% to 2% perfuses the medulla. Total renal blood flow in both kidneys is approximately 1250 mL/min. Blood flows to the kidneys from the abdominal aorta through the renal arteries, which then divide into several interlobar arteries. The interlobar arteries travel in the renal columns adjacent to the pyramids (see Fig. 26.2). When the interlobar arteries reach the border of the medulla and the cortex, they branch into the arcuate arteries. The arcuate arteries then travel along the cortical medullary border parallel to the renal capsule. The arcuate arteries branch further to form small interlobular arteries, which penetrate the cortex and branch extensively to form the afferent arterioles. The afferent arterioles divide to form glomerular capillaries, which coalesce to form the efferent arterioles (Fig. 26.3). The efferent arterioles branch again to form a second capillary bed. The peritubular capillaries wrap around the proximal and distal convoluted tubules (Fig. 26.4). Some capillaries, called *vasa recta*, dip down into the medulla to surround

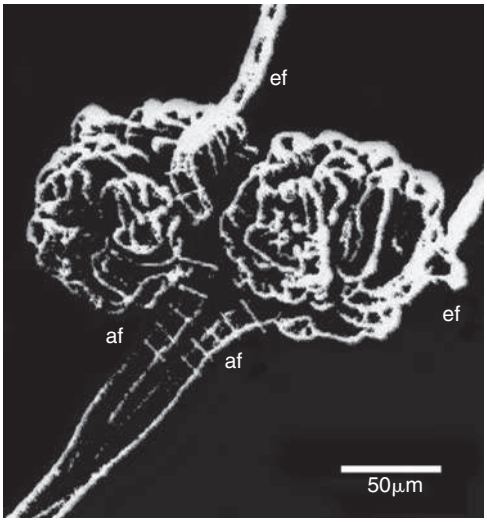


FIG 26.3 Scanning electron micrograph showing branching of an interlobular artery into two afferent arterioles *(af),* with associated glomerular tufts and efferent arterioles *(ef).* (From Kimura K et al: Effects of atrial natriuretic peptide on renal arterioles: morphometric analysis using microvascular casts, *Am J Physiol* 1990;259:F936. Used with permission.)

the loops of Henle and collecting ducts. The vasa recta have a specialized loop structure that enables them to pick up interstitial fluid without excessive removal of interstitial solutes. Solutes and water move into and out of the vasa recta passively such that the descending limb gains solute as it dips into the highly concentrated medulla, but then most of the solute is lost as the ascending loop makes its way back up to the cortex.

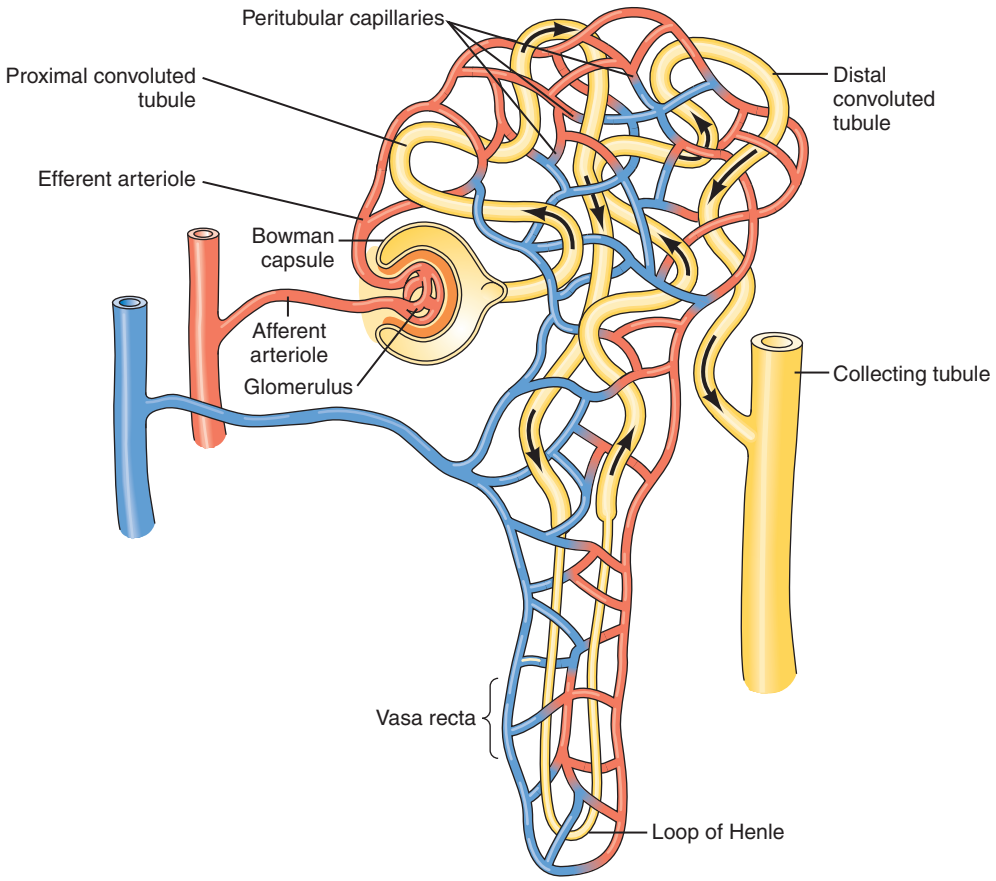


FIG 26.4 The nephron tubule is covered by peritubular capillaries and vasa recta, which pick up the fluid and solutes that have been reabsorbed by the tubular epithelium and return them to the general circulation.

The capillaries of the peritubular system and the vasa recta join together and drain into interlobular venules. The veins that drain blood from the kidney run parallel to the arteries and are similarly named (Fig. 26.5).

KEY POINTS

- The kidneys are located in the retroperitoneal space, just under the diaphragm. The right kidney is slightly lower than the left. The costovertebral angle (CVA) is an external landmark useful for locating the kidneys.
- The kidney can be divided into three principal anatomic sections: the pelvis, the medulla, and the cortex. The pelvis is composed of urinary collecting structures, called *calices*. The medulla is the middle portion and contains the renal pyramids. The cortex is the outer portion and contains glomeruli and nephron tubules.
- The kidneys are supplied with lymphatics to drain excess interstitial fluid and proteins and with sympathetic neurons to regulate blood supply and renin release.
- Blood is supplied to the kidneys by the renal artery, which divides several times to form the interlobar, arcuate, and interlobular arteries. The interlobular arteries branch multiple times to form afferent arterioles for each of the millions of kidney glomeruli.
- Each nephron has its own afferent arteriole, capillary network or tuft, and efferent arteriole. Efferent arterioles continue on to form peritubular capillaries, or vasa recta, which wrap around nephron structures and eventually drain into the renal veins. The loop structure of the vasa recta enables them to pick up interstitial fluid without removing excessive solute.

OVERVIEW OF NEPHRON STRUCTURE AND FUNCTION

Most of the physiologic functioning of the kidney can be understood by examining the function of an individual nephron. Thus the nephron is said to be the functional unit of the kidney. Nephrons are organized in parallel such that each must accomplish all the necessary processing before releasing urine into the collecting ducts. Complex autoregulatory mechanisms ensure that the workload is evenly distributed among the kidneys' many nephrons.

As the unit of kidney function, a nephron must accomplish three major functions: (1) filtration of water-soluble substances from the blood; (2) reabsorption of filtered nutrients, water, and electrolytes; and (3) secretion of wastes or excess substances into the filtrate. Different segments of the nephron are specialized to accomplish each of these processes. Each nephron is composed of a glomerulus, which includes the capillary tuft and Bowman capsule, and a tubule, which includes the proximal convoluted tubule, loop of Henle, distal convoluted tubule, and collecting tubule (Fig. 26.6). The nephron tubule is composed of a single layer of epithelial cells with an apical side facing the lumen and a basolateral side facing the interstitial space and capillaries (Fig. 26.7). The epithelial cells in each segment of the tubule are specialized for certain functions (Table 26.1). Nearly all cells in the nephron have a single cilium that protrudes from the apical surface into the lumen of the tubule. These cilia are mechanoreceptors and chemoreceptors that sense flow rate and composition of the tubular filtrate. Stimulation of the cilium triggers signaling cascades within the tubule cells that regulate cell proliferation, differentiation, and apoptosis. Abnormalities

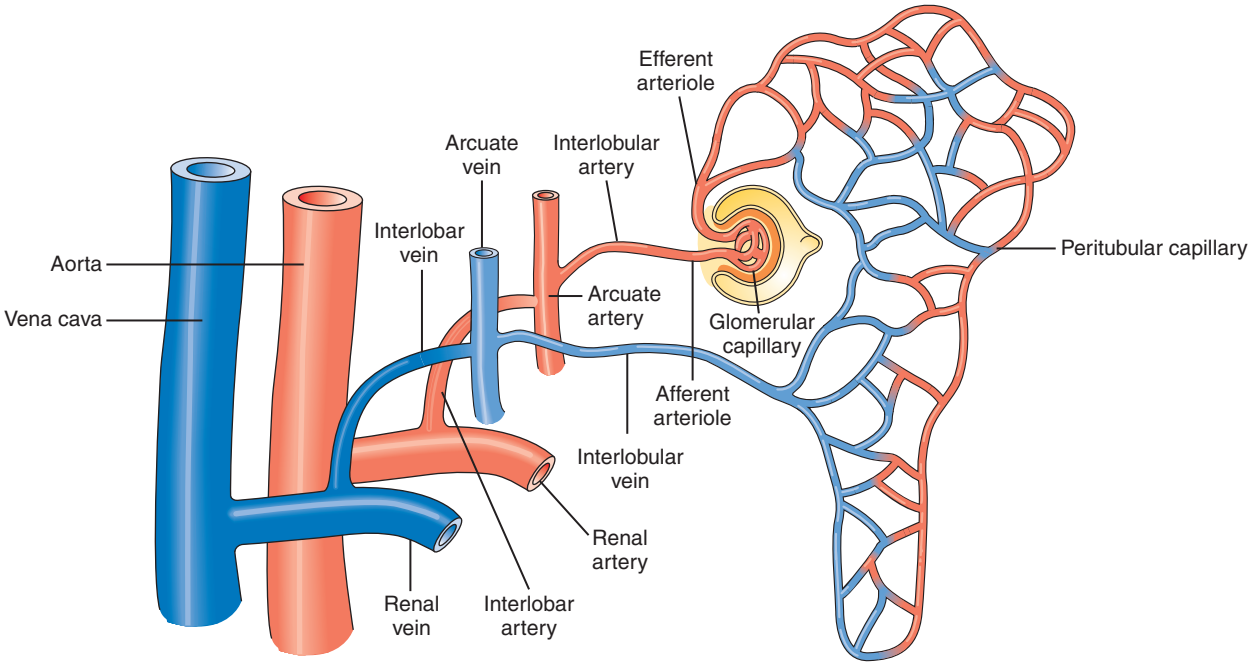


FIG 26.5 The venous vessels of the kidney parallel the arterial vessels and are similarly named.

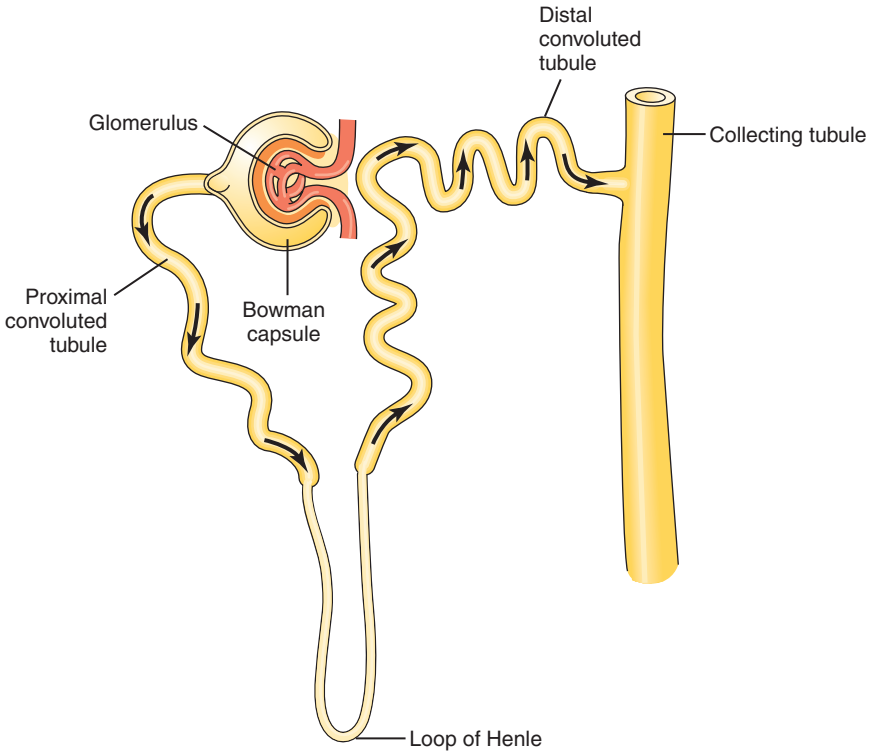


FIG 26.6 The nephron is composed of a glomerulus, proximal convoluted tubule, loop of Henle, distal convoluted tubule, and collecting tubule. Filtration occurs at the glomerulus, and the remaining tubule segments perform reabsorption and secretion functions.

in cilia signaling function have been found in patients with polycystic kidney disease, associated with mutations in genes *(polycystin, PKD1, PKD2)* that code for cilia membrane proteins (see Chapter 27).

Glomerulus

The glomerulus is the site of fluid filtration from the blood to the nephron tubule. It is formed by a capillary tuft, which lies between the afferent and efferent arterioles, and by the surrounding epithelial cells of Bowman capsule. The outer layer of the glomerular capsule is called the *parietal layer* and consists of a single thickness of epithelial cells resting on a layer of basement membrane (Fig. 26.8). The inner (visceral) layer of the capsule is composed of specialized epithelial cells called *podocytes*. Podocytes have foot processes (pedicels) that surround the glomerular capillary walls (Fig. 26.9). Between the podocyte and the

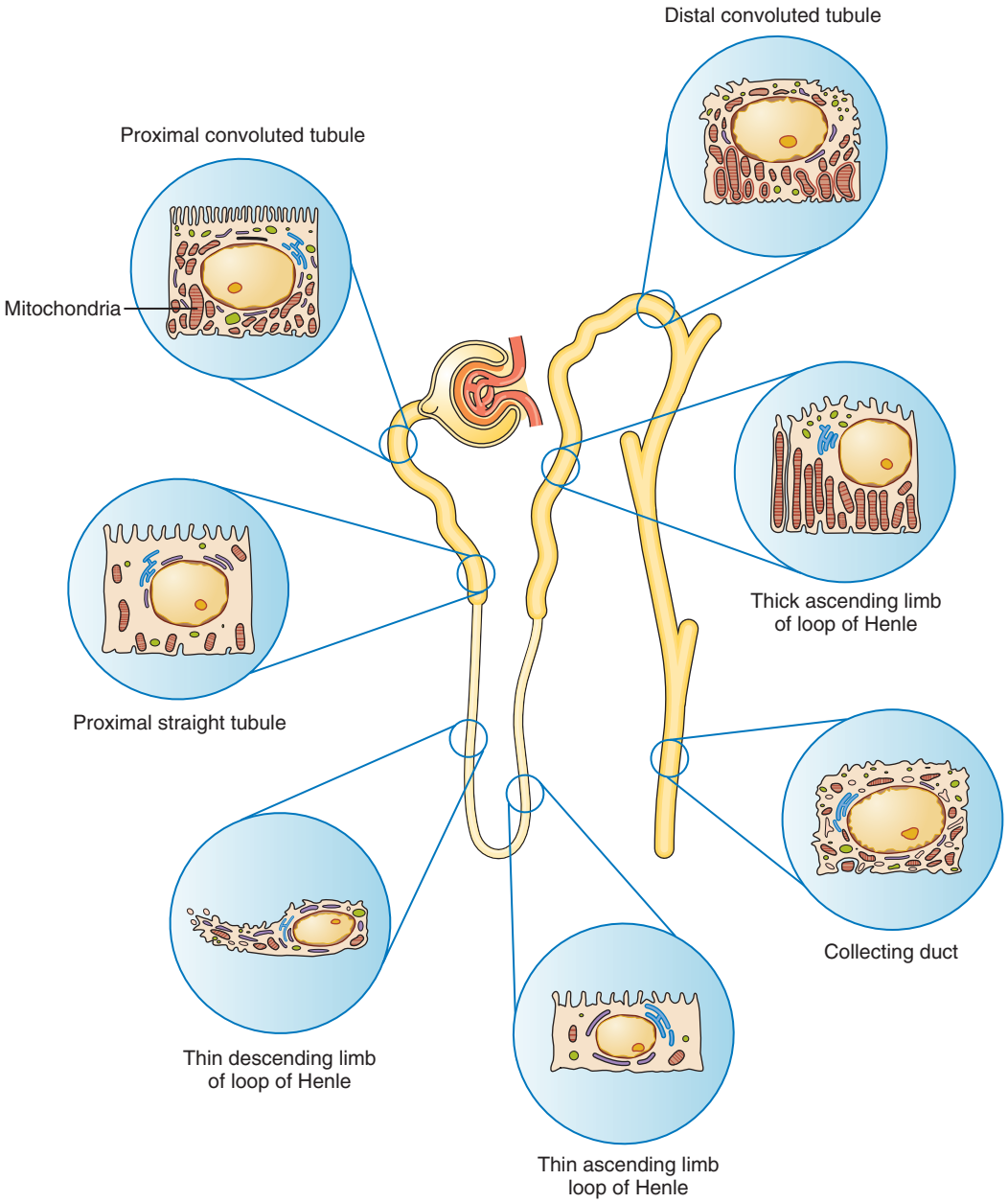


FIG 26.7 Each segment of the nephron is specialized for certain functions, which is reflected in the type of epithelial cells that compose the tubules.

TABLE 26.1 Functions of the Nephron Segments

Nephron Segment	Functions
Glomerulus	Filters fluid from blood into Bowman capsule; prevents passage of blood cells and proteins
Proximal convoluted tubule	Transports two-thirds of filtered water and electrolytes and all of the filtered bicarbonate, glucose, amino acids, and vitamins from filtrate to interstitium
Descending loop of Henle	Transports water; delivers a concentrated filtrate to ascending loop of Henle
Ascending loop of Henle	Actively transports Na^+, K^+, Cl^- to produce a hypoosmotic filtrate and a high interstitial osmolality
Distal convoluted tubule	Transports Na^+, Cl^-, water, and urea; responsive to aldosterone; site of macula densa regulation of GFR; secretes H^+ and K^+
Collecting tubule	Passively transports water under influence of ADH; secretes H^+ and K^+

ADH, Antidiuretic hormone; *GFR,* glomerular filtration rate.

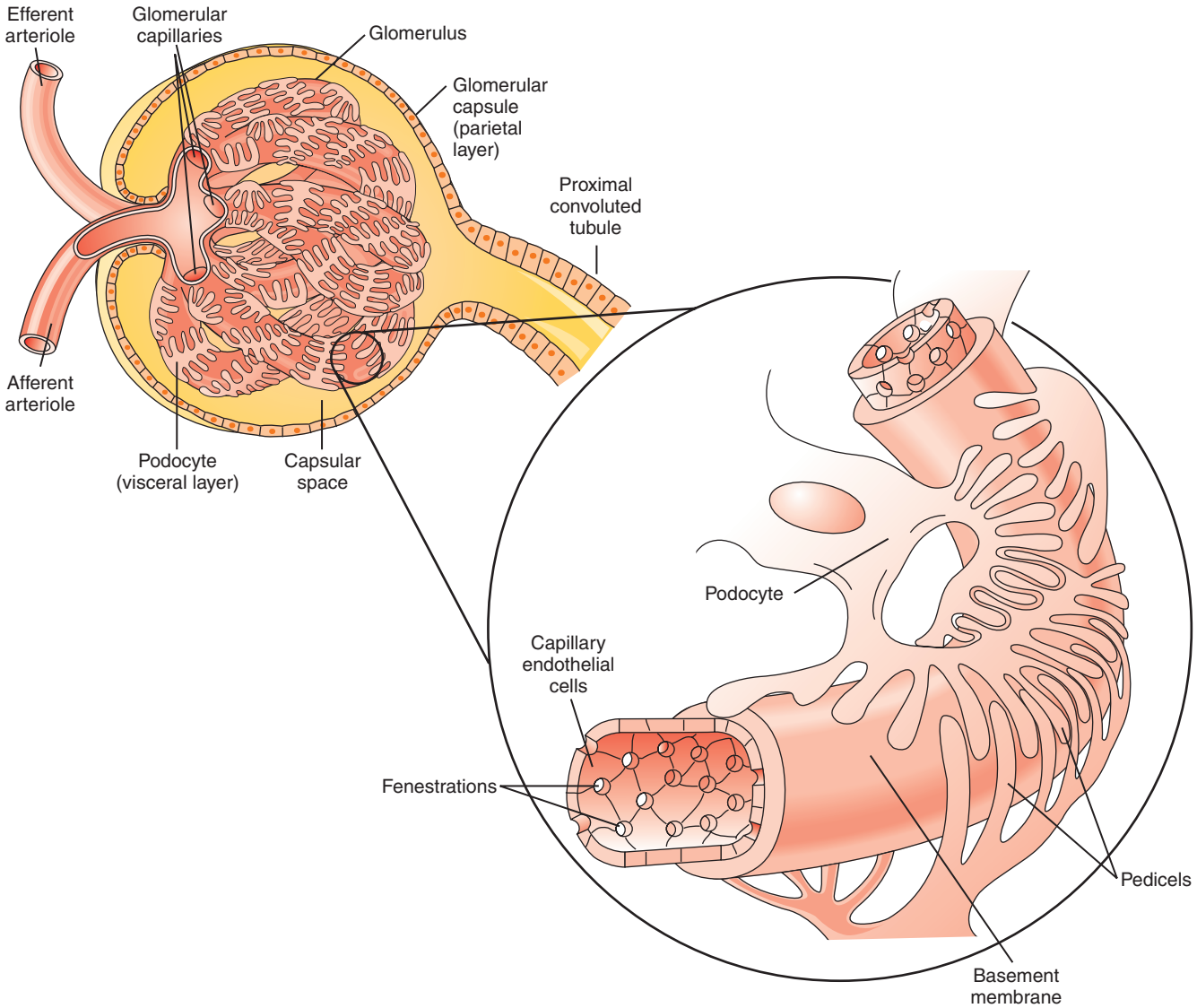


FIG 26.8 Structure of the glomerulus, including the afferent and efferent arterioles, capillary tuft, and surrounding epithelial membrane of Bowman capsule. The enlargement shows the glomerular membrane to be composed of the endothelial cells of the capillary, the podocytes of Bowman capsule, and the basement membrane between them.

capillary endothelium is a layer of extracellular matrix called the *basement membrane* (see Fig. 26.8).

Spaces between the endothelial cells are called *fenestra,* and spaces between the podocyte foot processes are called *slit pores.* These intercellular spaces provide the surface area for glomerular filtration and make the glomeruli considerably more permeable than other capillaries in the body (Fig. 26.10). The basement membrane is an important selectivity barrier of the glomerulus, preventing plasma proteins, erythrocytes, leukocytes, and platelets from passing through. Cells are too large to pass through pores, and plasma proteins are negatively charged and repelled to some extent by the basement membrane. Slit pores have a thin diaphragm of extracellular proteins that restricts the filtration of plasma proteins that make it through the basement membrane. *Nephrin,* podocin, NEPH1, CD2AP, and others are important proteins in the slit pores, as demonstrated by the proteinuria (protein in urine) that occurs when they are genetically mutated (Fig. 26.11). Proteins and blood cells are not usually present in the urine. If the glomerulus is injured, blood cells and proteins may filter through and be found in urine. Proteinuria is an important sign of basement membrane dysfunction. Except for the lack of proteins and cells, the glomerular filtrate is similar in composition to plasma.

Another important component of the glomerulus is the mesangium, which includes mesangial cells and mesangial matrix. Mesangial cells have a number of functions, including provision of structural support for glomerular capillaries, secretion of matrix proteins, phagocytosis, and regulation of the glomerular filtration rate (GFR). By contracting and relaxing, mesangial cells can alter the available surface area for filtration and affect GFR. The GFR averages about 125 mL/min.

Proximal Convoluted Tubule

The Bowman capsule drains the glomerular filtrate directly into the proximal tubule segment, where two-thirds of the water and electrolytes are rapidly transported from the filtrate to the interstitium for

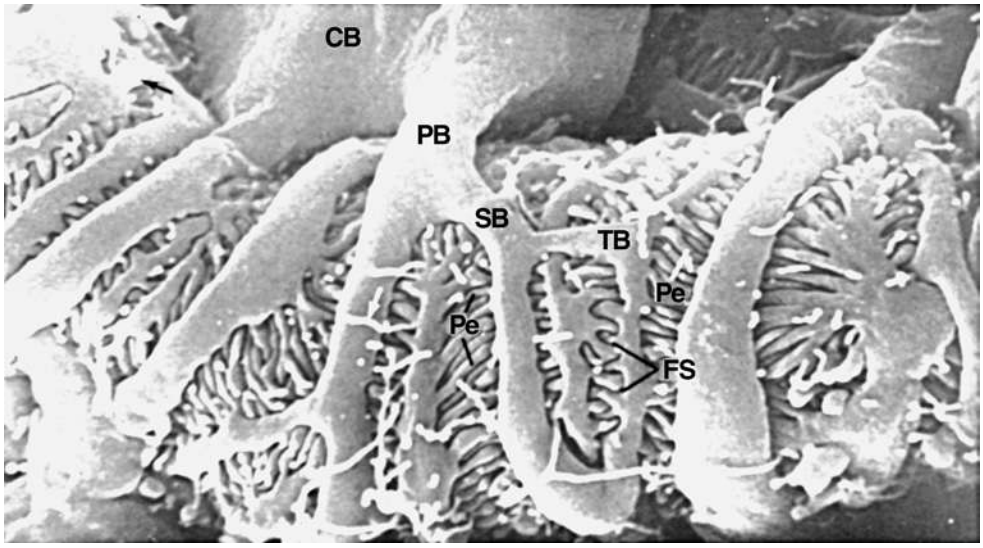


FIG 26.9 Electron micrograph showing a close-up view of podocyte foot processes of the glomerular capillary. Note the spaces between the podocyte foot processes that contribute to a highly permeable glomerular membrane. *CB,* Podocyte cell body; *FS,* filtration slits (slit pores); *PB,* primary branch; *Pe,* pedicle; *SB,* secondary branch; *TB,* tertiary branch. (From Kessel RG, Kardon RH: *Tissues and organs: a text-atlas of scanning electron microscopy*, San Francisco, 1979, WH Freeman.)

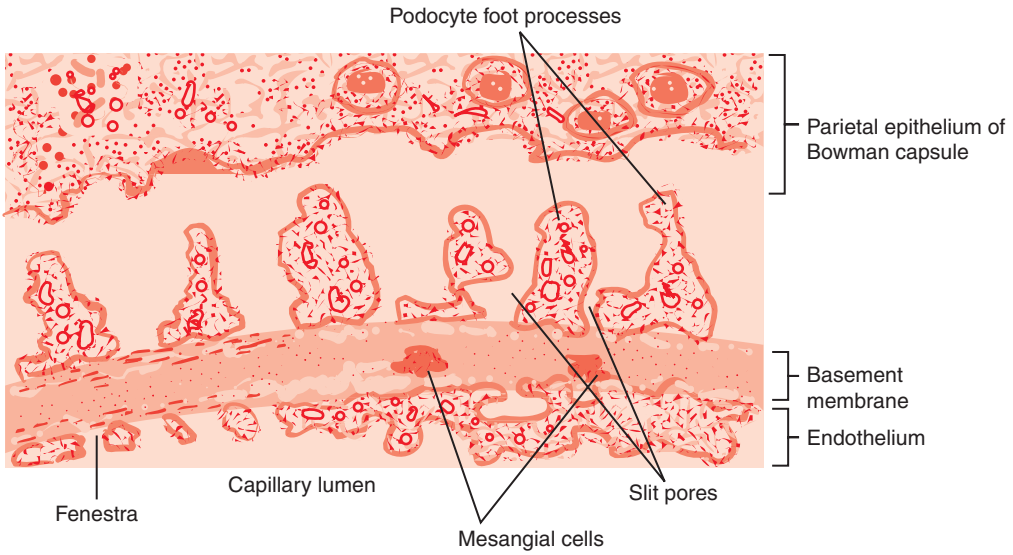


FIG 26.10 Section of the glomerular membrane showing the large spaces between the endothelial cells and podocyte foot processes. Filtration occurs through the fenestra and slit pores. The basement membrane provides the principal selectivity barrier of the glomerulus.

reabsorption by peritubular capillaries (Fig. 26.12). Nutrients, vitamins, and small proteins normally are reabsorbed completely in the early proximal tubule. The early proximal tubule is the site of most bicarbonate ion reabsorption, whereas chloride ion is reabsorbed in the late proximal tubule. The proximal tubule consists of cuboidal epithelium that is convoluted to provide a greater surface area for reabsorption. The epithelial cells in this segment have microvilli that form a brush border next to the filtrate and substantially increase the apical surface area. Proximal tubule cells have high adenosine triphosphate (ATP) requirements because most reabsorption utilizes active transport mechanisms that are dependent on Na^+–K^+ ion pumps in the basolateral membrane. Details of some of these transport mechanisms are discussed in subsequent sections. Water is reabsorbed passively through paracellular transport between the tubular cells and through water channels in the tubule cell membranes made of proteins called *aquaporin 1*. Reabsorption of solutes creates the osmotic force for passive water reabsorption.

Loop of Henle

The loop of Henle is divided into the descending and ascending limbs, which differ significantly in structure and function. The descending limb receives filtrate from the proximal convoluted tubule and delivers it to the ascending limb. The thin descending limb is permeable to water, but the thin and thick ascending part of the loop is not. The thick ascending segment contains powerful membrane pumps that cotransport ions (Na^+, K^+, $2Cl^-$) from the filtrate and deposit them in the interstitial fluid surrounding the loops of Henle and collecting ducts

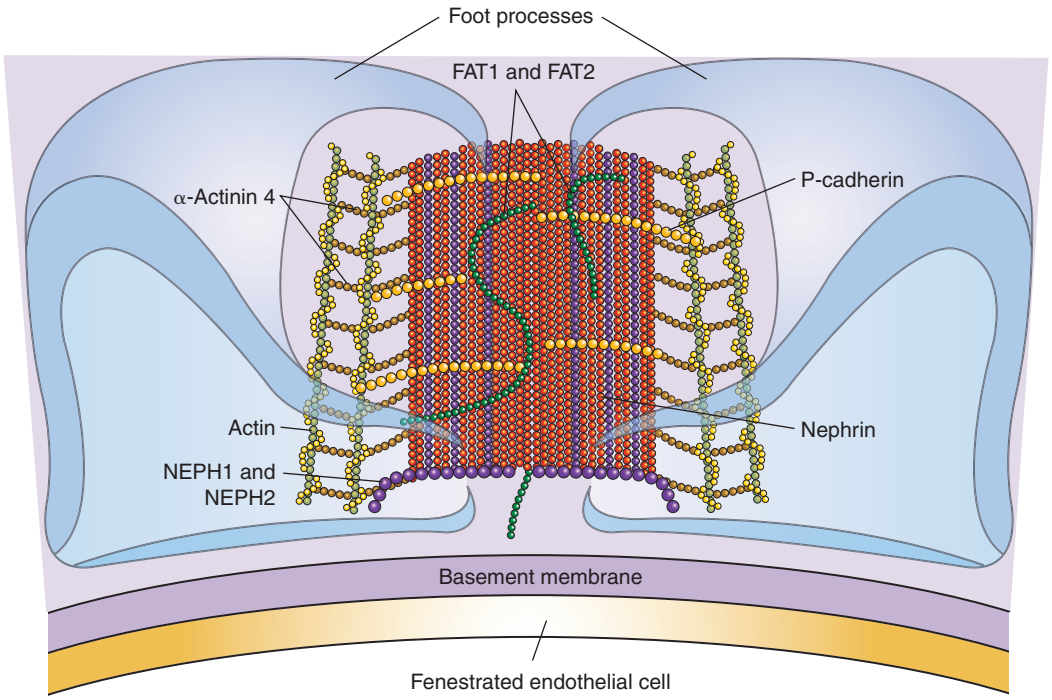


FIG 26.11 Diagram of the major proteins that constitute the matrix in the slit pore between renal podocyte foot processes. The proteins contribute to the ability of the matrix to selectively prevent serum proteins from filtering through the glomerulus.

(Fig. 26.13). About 15% of nephrons have extra-long loops of Henle that dip down into the medulla (juxtamedullary nephrons). These nephrons are vital for creating concentrated urine.

The loop formation of the loop of Henle creates a countercurrent mechanism, which allows the ascending loop of Henle to create a high interstitial gradient in the medulla of the kidney (Fig. 26.14). Because the ascending loop is impermeable to water, water cannot follow the Na^+, K^+, and Cl^- ions that are pumped into the interstitium. The descending loop is permeable to water, and water will be drawn out by the extra ions that were pumped into the interstitium by the ascending limb. Thus the filtrate that reaches the ascending limb will be more concentrated than the original filtrate. Delivery of a more concentrated filtrate to the ascending limb allows the Na^+–K^+–$2Cl^-$ cotransporter to pump out a greater number of ions and reach an even higher interstitial gradient (Fig. 26.15). This countercurrent mechanism creates a maximal osmolarity of about 600 mOsm/L at the tip of the loop of Henle compared with the usual extracellular osmolarity of 280 to 300 mOsm/L at the cortex. Another 600 mOsm/L is contributed by the accumulation of urea particles in the interstitium. Urea moves passively from the filtrate into the interstitium down its concentration gradient. It becomes concentrated in the tubule filtrate when electrolytes and water are removed in the proximal tubule and loop of Henle. An overall interstitial osmolarity is generated that begins in the cortex at about 300 mOsm/L and increases progressively to about 1200 mOsm/L at a point deep in the medulla. This high interstitial osmolarity provides a gradient for water reabsorption from the collecting ducts as they pass through the medulla on their way to the renal pelvis. The maximal interstitial gradient attained is dependent on the length of the loops of Henle. In some animals that survive in dry climates, such as the desert mouse, very long loops of Henle create a much higher interstitial osmolarity, which allows formation of extremely concentrated urine. Fluid that is reabsorbed from collecting tubules into the medullary interstitium is picked up by the specialized capillary network called the *vasa recta* and returned to the venous circulation. Like other capillaries, the vasa recta passively exchange ions according to concentration gradients and passively reabsorb fluid by filtration forces. The loop structure of the vasa recta allows the capillary to passively leak accumulated solute back into the interstitium as the capillary makes its way back to the cortex from the medulla. This process minimizes the washout of the interstitial osmolality and has been called the *countercurrent exchange mechanism.*

Distal Convoluted Tubule

The filtrate that reaches the distal tubule is normally hypoosmotic (100 mOsm/L) in comparison with plasma (280 mOsm/L) because electrolytes have been removed by the pumps in the ascending loop of Henle. At this point in the nephron, only 10% of the original glomerular filtrate volume remains, and further reabsorption in the distal tubule is largely under hormonal control. Aldosterone and angiotensin II (AII) stimulate the tubule cells to reabsorb sodium and water, whereas atrial natriuretic peptide (ANP) and urodilatin inhibit reabsorption.

Collecting Duct

The distal tubules of several nephrons empty into a single collecting tubule, which then merges into progressively larger and fewer collecting ducts that run parallel to the loops of Henle. Eventually the collecting ducts form the medullary pyramids, which empty into the minor calices through the papilla. The collecting ducts travel through the high interstitial gradient of the medulla on their way to the renal pelvis. The collecting ducts have two cell types called *principal cells (P cells)* and *intercalated cells (I cells).* The majority of cells are the P type that responds to antidiuretic hormone (ADH). In the presence of ADH, more than 99% of the original filtrate is reabsorbed by the time it reaches the renal pelvis, creating 30 to 60 mL of concentrated urine per hour. The I cells participate in acid–base balance by regulating the secretion of acid.

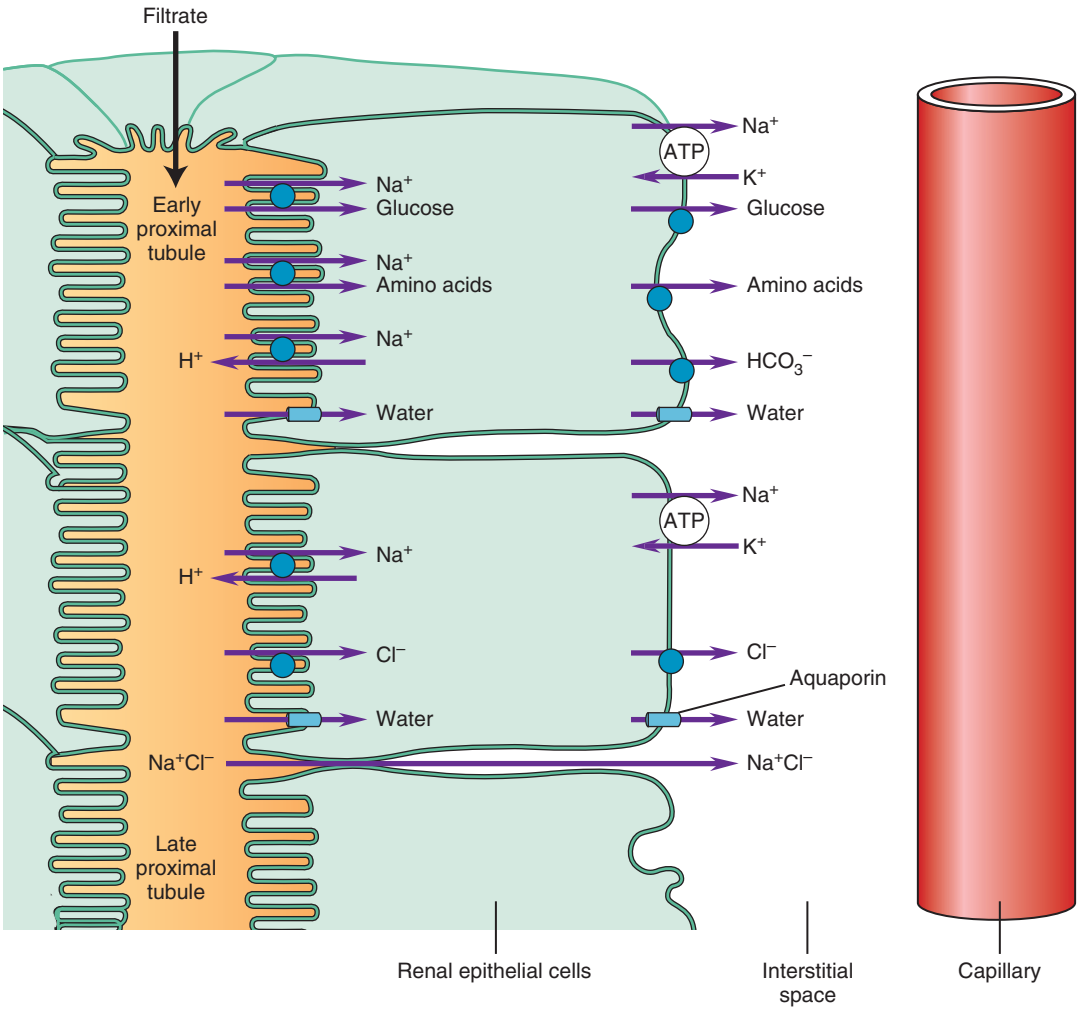


FIG 26.12 The proximal convoluted tubule has numerous membrane transporters that function to reabsorb filtered glucose, amino acids, water, and electrolytes. The early proximal tubule reabsorbs nearly all of the filtered bicarbonate ions (see Fig. 26.21), whereas the late proximal tubule reabsorbs chloride ions.

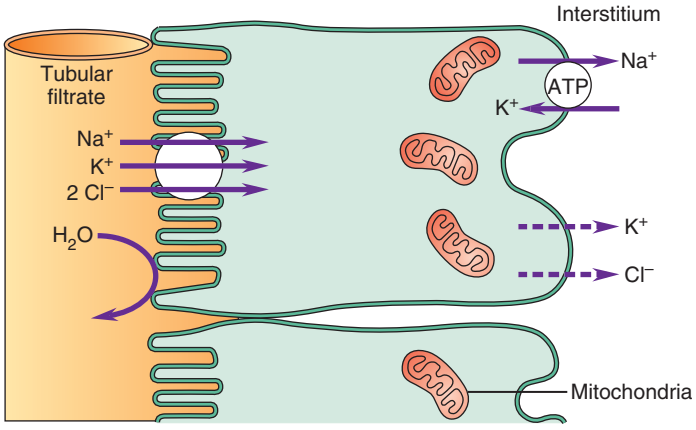


FIG 26.13 The epithelial cells of the thick ascending loop of Henle possess powerful ion pumps that cotransport Na^+, K^+, and $2Cl^-$ ions from the filtrate into the cell. The Na^+ is then pumped out of the basolateral membrane and into the interstitium. The loop of Henle ion cotransporter is responsible for creating a highly concentrated medullary interstitium.

REGULATION OF GLOMERULAR FILTRATION

The GFR is determined by the filtration pressure in the glomeruli and by the permeable surface area of the glomerular membrane (K_f). Filtration pressure varies considerably from the afferent end of the glomerulus to the efferent end and is difficult to measure directly. The average net filtration pressure for the capillary as a whole is about 10 mm Hg, and the permeability constant K_f is about 12.5 mL/min per mm Hg. GFR is the product of filtration pressure and K_f (10 mm Hg × 12.5 mL/min per mm Hg = 125 mL/min). It is determined by the physical principles of filtration across a capillary membrane (Fig. 26.16). These values are not easily measured in patients, so a global assessment of GFR is used clinically to assess renal function (see Tests of Renal Structure and Function); however, understanding the principles that underlie the global clinical measurement is essential to anticipate and prevent impaired GFR.

Physics of Filtration

Filtration rate is affected by factors that alter hydrostatic and oncotic pressure on either side of the glomerular membrane, as shown by the following filtration equation:

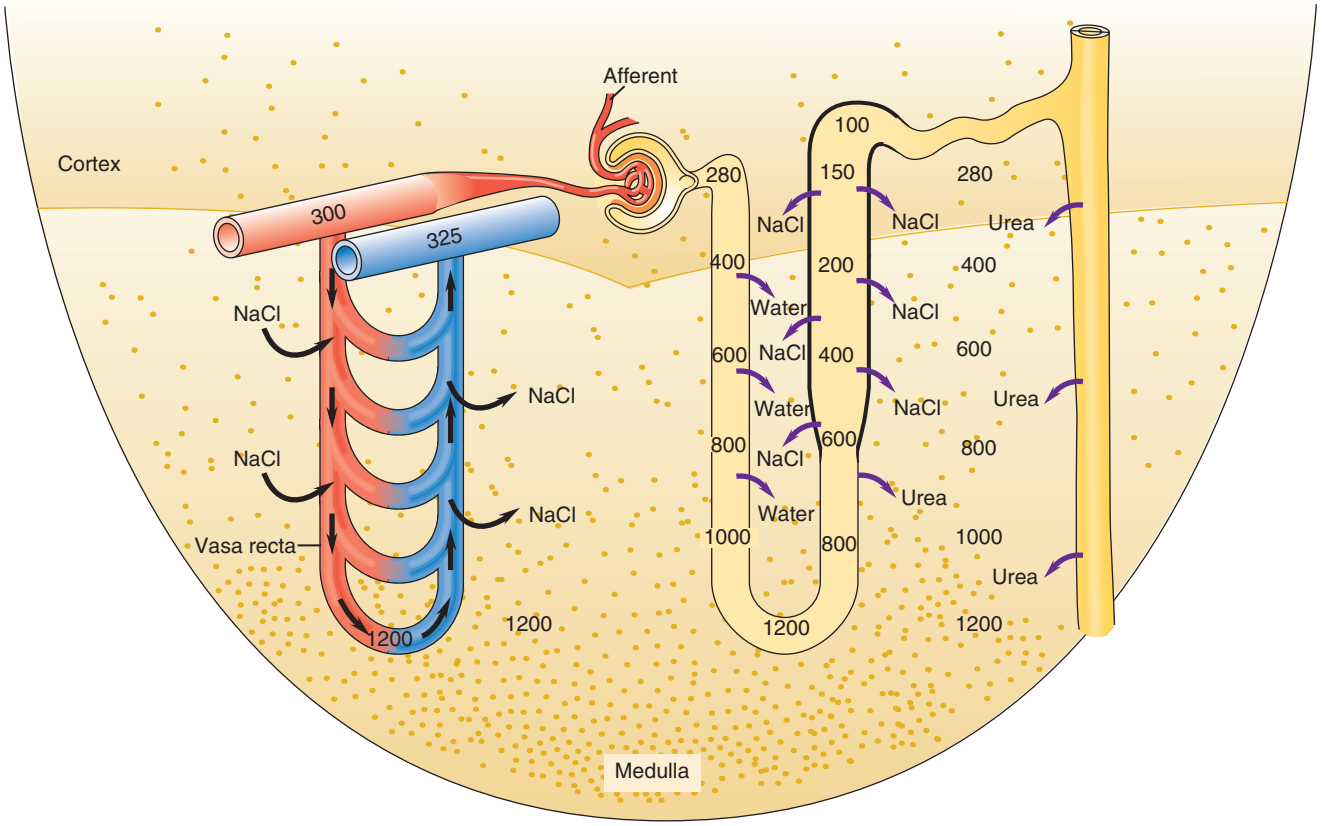


FIG 26.14 Mechanism of countercurrent multiplication. Ion pumps in the ascending loop of Henle create an interstitial gradient in the medulla of the kidney. The separation of solute from water in the ascending loop of Henle also produces a dilute tubular fluid, allowing the excretion of excess water by making a dilute urine in the absence of antidiuretic hormone (ADH). NaCl accumulation in the interstitium contributes about half of the total osmolality. Urea particles in the interstitium contribute the other half of the particles that produce the normal interstitial gradient in the medulla. As water is removed from the collecting duct (in the presence of ADH), urea becomes more concentrated and moves passively down its gradient into the interstitium. The specialized loop structure of the vasa recta allows it to pick up interstitial water from the medulla without significant solute removal. Although solutes are acquired in the descending segment of the vasa recta, they passively diffuse back out as the ascending segment reaches the cortex. This process is called *countercurrent exchange*.

$$GFR = K_f\,[(P_{GC} + \pi_{BC}) - (P_{BC} + \pi_{GC})]$$

where P_{GC} is glomerular capillary hydrostatic pressure (mm Hg); π_{BC} is oncotic pressure in the Bowman capsule (mm Hg); P_{BC} is Bowman capsule hydrostatic pressure (mm Hg); and π_{GC} is oncotic pressure in the glomerular capillary (mm Hg). The following is an illustrative example resulting in a normal GFR of 125 mL/min:

$$GFR = 12.5\,[(60 + 0) - (18 + 32)]$$

$$GFR = 125\ mL/min$$

The main driving force for filtration is hydrostatic pressure in the glomerular capillaries. The glomerular capillary hydrostatic pressure exerts a force against the glomerular capillary walls. As blood circulates through the capillaries, the hydrostatic pressure pushes blood against the walls, and fluid is filtered out. The hydrostatic pressure remains fairly constant along the length of the capillary and exerts an average force of approximately 60 mm Hg.

The glomerular capillary oncotic (colloid osmotic) pressure exists because proteins are present in the blood. Plasma proteins are negatively charged and attract positive ions, which subsequently attract water. Because ions and water are attracted to the proteins and are not pushed against the capillary wall, the glomerular capillary colloidal osmotic pressure opposes filtration by holding water and ions in the capillaries. The glomerular oncotic pressure is lower at the afferent end and becomes progressively higher along the length of the capillary (see Fig. 26.16).

The hydrostatic pressure in Bowman capsule is determined by the volume of filtrate present in the capsule. This pressure exerts a force against the walls of Bowman capsule and the glomerular capillaries and opposes filtration. The normal Bowman capsule hydrostatic pressure is about 18 mm Hg. Normally plasma proteins do not filter into Bowman capsule. If they did filter, then they would create Bowman capsule oncotic pressure. This pressure would enhance glomerular filtration because proteins attract cations and water. In a healthy kidney, this pressure is negligible.

In summary, the net filtration pressure across the glomerular membrane is approximately 10 mm Hg. The filtration pressure is higher at the afferent arteriole side of the capillary and diminishes as the blood reaches the efferent end. This occurs because the capillary oncotic pressure is lower at the afferent end. As blood passes through the capillaries, continued filtration leaves a greater concentration of proteins in the capillaries, which raises the oncotic pressure. As blood reaches the efferent arterioles, filtration may cease.

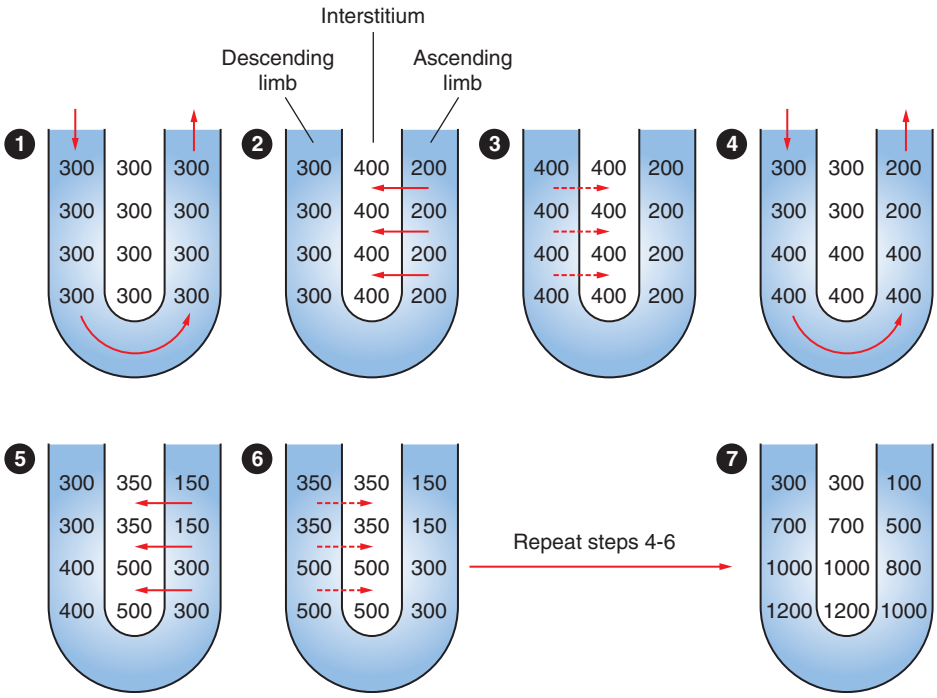


FIG 26.15 Sequence of events in development of the interstitial gradient by countercurrent multiplication. The Na^+–K^+–$2Cl^-$ cotransporters in the thick ascending loop of Henle can produce a gradient across the tubule wall of about 200 mOsm/L. The overall interstitial solute gradient is higher than these pumps could accomplish without the loop structure of the tubule. Countercurrent multiplication occurs because the descending loop is permeable to water and equilibrates with the rising interstitial solute concentration. Thus the filtrate reaching the ascending loop is increasingly concentrated with each step (1–7), allowing the ascending loop to further increase the osmolality of the interstitial fluid.

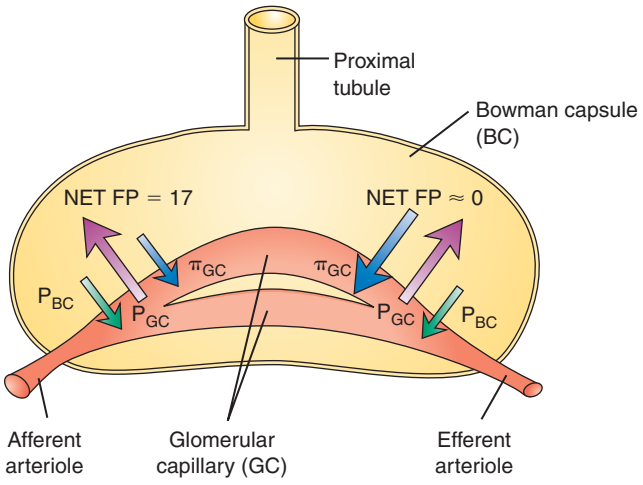


FIG 26.16 Net filtration is higher at the afferent end of the glomerular capillary because the hydrostatic blood pressure in the capillary exceeds the pressure in Bowman capsule and the oncotic pressure in the capillary. Toward the efferent end of the capillary, the filtration pressure is low because the oncotic pressure of the blood is high and offsets the hydrostatic blood pressure. Capillary oncotic pressure gets progressively higher along the capillary because fluid is filtering out of the blood into Bowman capsule and leaving the proteins behind so they become more concentrated and exert a greater oncotic pressure. *FP,* Filtration pressure; π, oncotic pressure; *P,* hydrostatic pressure.

Factors Affecting Filtration Pressure

One of the most important physiologic regulators of GFR is blood volume. When blood volume increases because of fluid intake, the blood pressure rises slightly and causes glomerular hydrostatic pressure to increase. GFR increases, and the extra fluid is pushed into the filtrate to be excreted from the body. The opposite also occurs: when blood volume is decreased, capillary hydrostatic pressure falls, resulting in a lower GFR, and fluid is conserved. The glomerular capillary is protected from large swings in blood pressure by autoregulation. Autoregulation adjusts the arteriolar resistance to maintain a relatively steady rate of blood flow despite changes in perfusion pressure. Autoregulation is effective when arterial blood pressure varies between 75 and 160 mm Hg. Autoregulation of renal blood flow is achieved in part by a stretch response in the vascular smooth muscle of the afferent arterioles. When blood pressure increases, the vascular smooth muscle cells reflexively constrict to keep blood flow at about the same rate. This mechanism is called *myogenic autoregulation.*

Other factors can affect GFR by altering the pressure within Bowman capsule or affecting plasma oncotic pressure. Obstruction in the tubules or collecting ducts can significantly elevate the pressure in Bowman capsule. According to the filtration equation, GFR would fall because filtration pressure would be reduced. Because plasma oncotic pressure is determined primarily by the concentration of plasma proteins, a low serum albumin concentration would increase GFR.

Although K_f is called a constant, it is subject to change for physiologic and pathologic reasons. Specialized mesangial cells located in the glomerulus are thought to be important regulators of K_f. These cells

contract and relax in response to various stimuli and alter the surface area for filtration. Contraction squeezes the capillary cells together and reduces GFR, whereas relaxation allows the permeable surface area to expand. Disease processes that damage the glomerular membrane also can affect permeability. Sclerotic processes reduce K_f, whereas some inflammatory injuries may increase it.

Tubuloglomerular Feedback

Each nephron is able to regulate its own individual GFR through a process termed *tubuloglomerular feedback.* A specialized group of cells forms the regulatory structure, called the juxtaglomerular apparatus**.** The juxtaglomerular apparatus is composed of the glomerulus, the macula densa, and specialized juxtaglomerular cells, which are located around the glomerular arterioles (Fig. 26.17). The macula densa cells are located near the end of the thick ascending loop of Henle, which loops up to come in contact with the glomerulus and juxtaglomerular cells.

Macula densa cells sense changes in the amount of NaCl delivered to the tubule. When glomerular filtration is increased, a higher load of NaCl is delivered to the distal tubule. The mechanism whereby macula densa cells sense GFR is not completely understood, but NaCl delivery to the macula densa cells is a critical part of the process. Macula densa cells possess the same Na–Cl–K transporters as other cells in the thick ascending limb, and when NaCl delivery is high, more is transported into the cells through this transporter (Fig. 26.18). The macula densa cells increase the activity of the Na–K pump in the basal membrane, which stimulates production of adenosine and ATP. Adenosine stimulates contraction of afferent arterioles and relaxation of efferent arterioles, thus decreasing filtration at the glomerulus. Tubuloglomerular feedback helps distribute GFR evenly among the kidneys' 2 million nephrons.

The juxtaglomerular cells that surround the afferent arteriole are also thought to be mediators of tubuloglomerular feedback. The juxtaglomerular cells produce and release renin**,** an enzyme that converts angiotensinogen to angiotensin I (AI). Angiotensin I is then converted to AII by endothelial cells in the glomerular capillary, which possesses angiotensin-converting enzyme (ACE) activity. AII is a potent vasoconstrictor that constricts both afferent and efferent arterioles. The signals that pass from the macula densa to the glomerulus to regulate tubuloglomerular feedback are not completely known; however, roles for a number of vasoactive factors, including adenosine and ATP, prostaglandins, and nitric oxide have been demonstrated. Some vasoactive factors have vasodilating effects, whereas others are vasoconstrictors, the relative balance of which determines the degree of arteriolar constriction.

The importance of prostaglandins and AII in regulating GFR is supported by the observation that drugs that inhibit their activity interfere with tubuloglomerular feedback in some persons. For example, ACE inhibitors block AII production and may interfere with constriction of the efferent arteriole. This can be particularly detrimental to renal function in patients who require high filtration pressures, such as those with polycystic kidney disease or collecting system obstructions. Drugs that inhibit cyclooxygenase, such as aspirin and nonsteroidal antiinflammatory drugs, interfere with prostaglandin production and may precipitate excessive renovascular constriction in some patients.

Effects of Glucose and Amino Acids

The amount of glucose and amino acids filtered into the tubular fluid may alter GFR through the tubuloglomerular feedback mechanism.

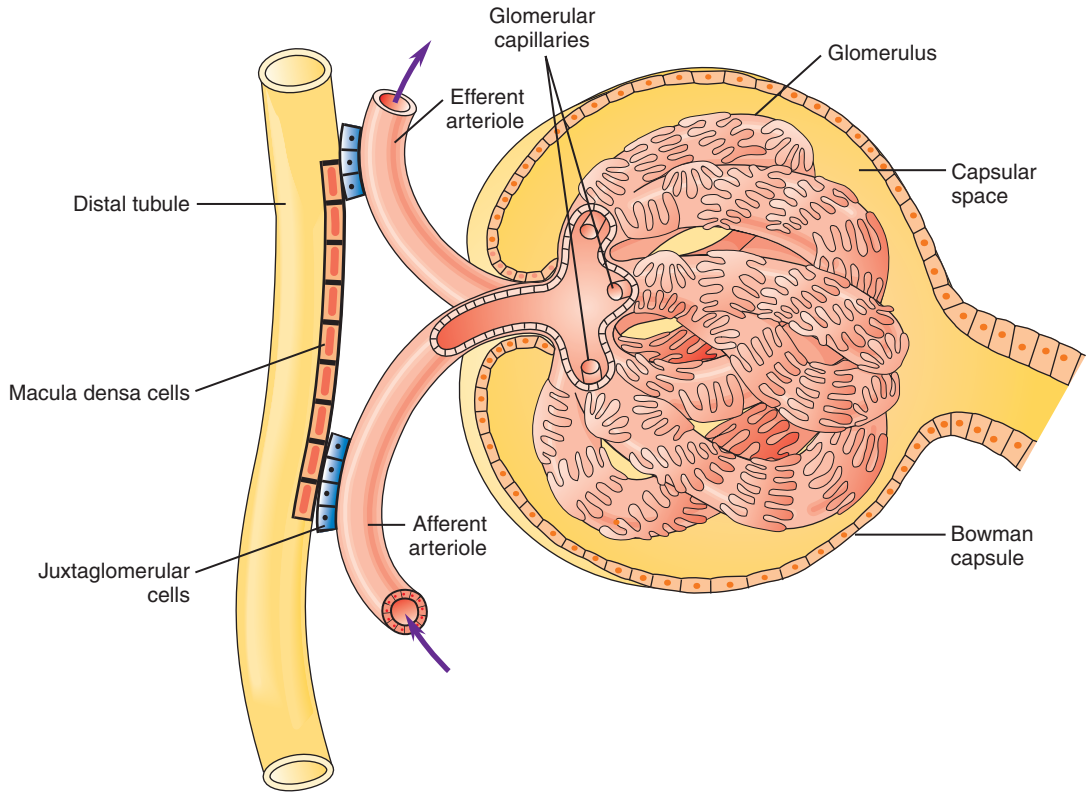


FIG 26.17 The juxtaglomerular apparatus is composed of the macula densa cells of the distal tubule, the afferent and efferent arterioles, and the renin-secreting juxtaglomerular cells. Macula densa cells sample the distal filtrate for NaCl content and send signals to the glomerulus to adjust the glomerular filtration rate.

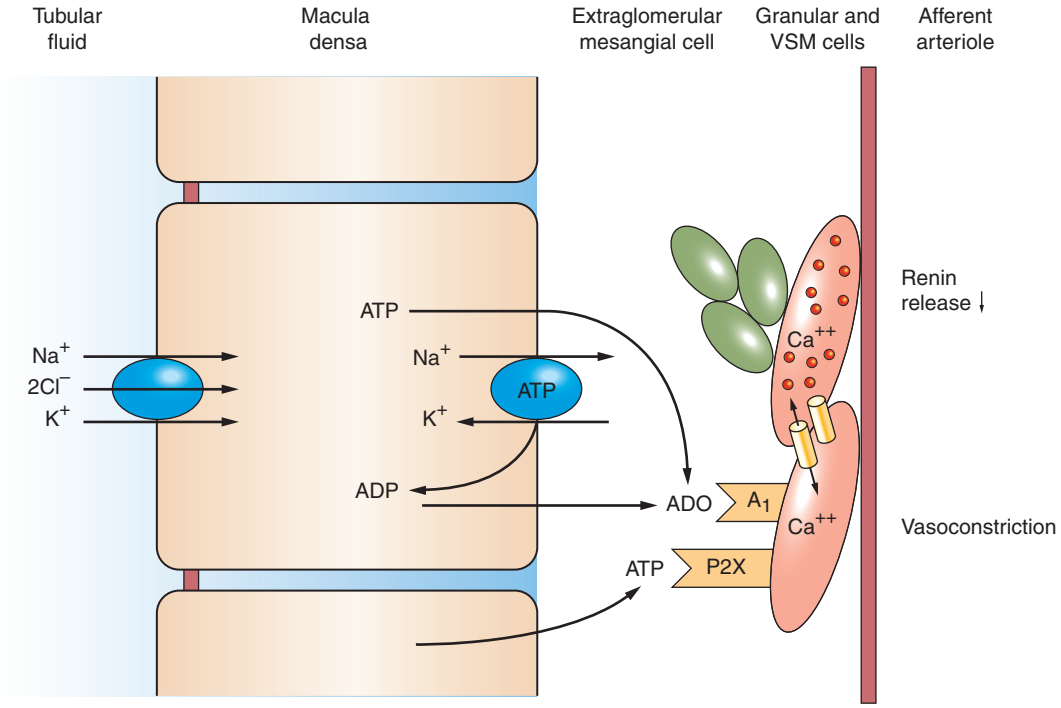


FIG 26.18 Mechanism of NaCl sensing by macula densa cells in tubuloglomerular feedback. Entry of ions through the apical Na–K–Cl transporter stimulates activity of the Na–K–ATPase pump on the basal side and stimulates production of ATP and adenosine. Receptors on vascular smooth muscle cells (VSM) of the afferent arterioles bind to ATP and adenosine and cause vasoconstriction. Renin release from the juxtaglomerular cells is also inhibited, thus reducing the amount of angiotensin II that reaches the efferent arteriole. The efferent arteriole becomes less constricted. Afferent constriction and efferent dilation reduce the filtration pressure in the glomerulus, reduce filtration, and reduce the delivery of NaCl to the macula densa. *ADO,* Adenosine; *A1,* adenosine-1 receptor; *P2X,* purineoreceptor (ATP receptor).

Both glucose and amino acids are filtered freely through the glomerular membrane and then are reabsorbed by active transport processes in the proximal tubule. Reabsorption occurs through transporters that use sodium ion entry into the cell to actively cotransport glucose and amino acids. The greater the load of tubular glucose and amino acids, the greater the amount of sodium reabsorbed by the proximal tubule. Fewer sodium ions are transported to the macula densa cells in the distal tubule, and the macula densa perceives this as a need to increase GFR. In addition, chronically high serum glucose concentrations, as occurs in poorly controlled diabetes mellitus, may induce excessive nitric oxide production, producing hyperfiltration, excessive GFR, and damage to the glomerulus.

Role of Mesangial Cells

Mesangial cells are located around the glomerular capillaries and are thought to regulate the surface area available for glomerular filtration. Contraction of the mesangial cells reduces surface area, and relaxation increases it. Mesangial cells are responsive to glomerular stretch and are stimulated to contract when more blood enters the glomerulus. This response provides a negative feedback that decreases surface area when filtration pressure is increased. In addition, mesangial cells respond to a number of chemical mediators, including AII and endothelin (peptides that favor mesangial contraction) and ANP and nitric oxide (substances that favor relaxation). Mesangial cells thus may regulate GFR by altering the filtration constant K_f.

KEY POINTS

- Glomerular filtration rate (GFR) is determined by the permeable surface area (K_f) and by capillary hydrostatic pressure and Bowman capsule oncotic pressure, which favor filtration, and by plasma oncotic pressure and Bowman capsule hydrostatic pressure, which oppose filtration. Normal GFR is about 125 mL/min.
- The filtration rate in an individual nephron is regulated by tubuloglomerular feedback in which macula densa cells sample the NaCl content (or other constituents) in the distal tubule and signal the juxtaglomerular apparatus to make appropriate adjustments in GFR.
- Afferent constriction and efferent dilation decrease GFR, whereas afferent dilation and efferent constriction increase GFR. Adenosine, AII, nitric oxide, and prostaglandins are important chemical mediators of arteriolar resistance.
- Excessive filtration of glucose and amino acids can result in increased GFR because their reabsorption in the proximal tubule requires cotransport of sodium ions. Fewer sodium ions are delivered to the macula densa, and GFR is increased.
- Mesangial cells respond to glomerular capillary stretch by contracting to reduce the surface area for filtration. Mesangial cells also respond to chemical signals that induce them to contract and relax, which alters GFR accordingly.

TRANSPORT ACROSS RENAL TUBULES

Reabsorption and secretion of substances across the nephron tubule are accomplished by two routes: the *transcellular* and the *paracellular* routes. Transcellular transport uses specific transporter proteins in the membranes of the tubular epithelial cells to move substances between the tubular filtrate and the interstitial fluid. Most of these transport processes are dependent on Na^+ reabsorption and are made possible by the Na^+–K^+ pump in the basolateral membrane. Paracellular transport refers to movement of substances through the tight junctions that hold the tubular epithelial cells together. Substances using the paracellular route therefore do not traverse the cell membrane and instead move passively through the spaces between cells. *Reabsorption* is the process of transporting a substance from the filtrate into the renal capillaries and requires several transport steps. First, the substance is moved from the filtrate into the tubular cell through transporters on the apical surface of the cell; then it passes through another transporter on the basolateral side of the tubular cell and into the interstitium. From the interstitium, it moves passively by diffusion or filtration into the capillary. In general, the reabsorption of cations, especially sodium, provides an electrical gradient to pull anions across the tubule and into the interstitium. Reabsorption of ions and solutes creates an osmotic force to pull water passively across the renal epithelium. A summary of transport in the various tubule segments is shown in Fig. 26.19. The details of glucose, bicarbonate, H^+, and K^+ transport are described next as important representative examples.

Reabsorption of Glucose

Glucose is filtered freely across the glomerular membrane such that the tubular load (in milligrams per minute) is determined by the product of serum glucose concentration (in milligrams per milliliter) and GFR (in milliliters per minute). Normally, all of the filtered glucose is reabsorbed in the proximal tubule by sodium-dependent cotransporters (e.g., SGLT 2) (Fig. 26.20). The transport proteins have a maximal rate of transport that can be exceeded if the tubular load of glucose is too great. The transport maximum for normal kidneys is about 375 mg/min. A tubular load of glucose in excess of this amount results in glycosuria. In fact, some spillage of glucose begins at a much lower tubular load because of the uneven distribution of GFR to individual nephrons or differences in the number of transporters in different nephrons. Some nephrons with higher GFR or fewer transporters may exceed their transport maxima, whereas other nephrons are working below capacity. The point at which glucose begins to spill into the urine is called the *renal threshold*. In normal kidneys with a GFR of 125 mL/min, the renal threshold will be reached when serum glucose concentration approaches 180 mg/dL, but significant glycosuria will not occur until the transport maximum is reached at a serum glucose level of about 300 mg/dL. Persons with low GFR associated with renal disease may not experience spillage of glucose until the serum glucose level is much higher, and glycosuria is not a reliable indicator of serum glucose level in these individuals. For example, a patient with a GFR of 50 mL/min and a serum glucose concentration of 300 mg/dL will have a tubular glucose load of only 150 mg/min, which is well below the normal renal threshold. No glycosuria would occur despite the high serum glucose level.

Regulation of Acid–Base Balance

The kidney tubules have an important role in maintaining the pH of the blood. In addition to excreting excess H^+, the kidneys regulate the concentration of bicarbonate (HCO_3^-) in the blood. The pH of the blood normally ranges between 7.35 and 7.45 and is determined by the ratio of acid (H_2CO_3) to base (HCO_3^-). The lungs and kidneys work together to maintain this balance. Metabolic processes create an excess of acid, which is excreted by the lungs in the form of CO_2 and by the kidneys in the form of H^+. In addition, HCO_3^- is filtered freely through the glomerulus and must be efficiently reabsorbed to maintain acid–base balance. Most HCO_3^- is reabsorbed in the proximal tubule; however, the distal segment also participates in regulating HCO_3^- and H^+ transport.

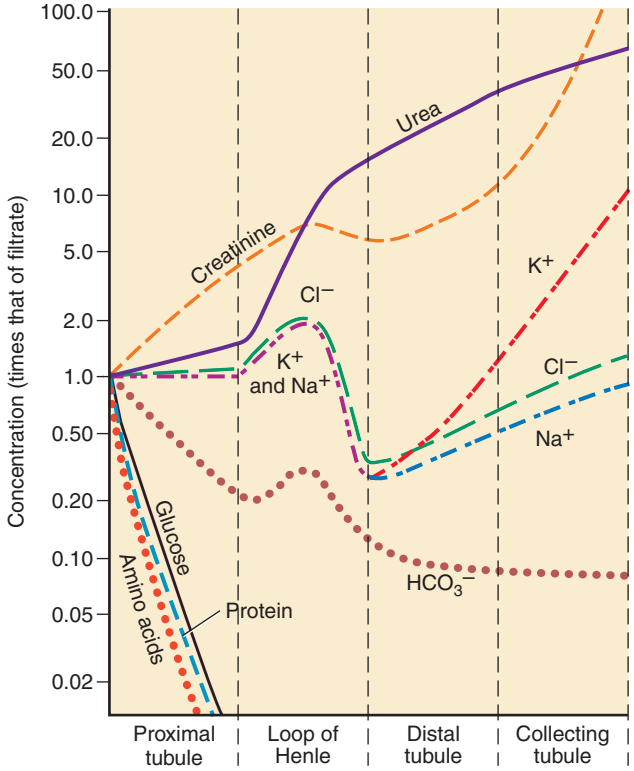


FIG 26.19 Summary of nutrient and electrolyte composition of the filtrate in each segment of the nephron. Two-thirds of the filtrate is reabsorbed in the proximal tubule. (Adapted from Hall JE, editor: *Guyton and Hall textbook of medical physiology,* ed 13, Philadelphia, 2015, Saunders, p 359.)

Reabsorption of HCO_3^- is complex because it is not directly transported across the apical membrane; rather, it is combined with H^+ in the tubule to form H_2CO_3, which dissociates into CO_2 and water (Fig. 26.21). The H^+ for this reaction is secreted into the filtrate in exchange for Na^+. Carbonic anhydrase present in the brush border of the proximal tubule cell catalyzes the reaction. Carbon dioxide is lipid soluble and diffuses passively into the tubular cell. Once inside, intracellular carbonic anhydrase catalyzes the reverse reaction to once again form HCO_3^- and H^+. The HCO_3^- is transported out through the basolateral membrane, whereas the H^+ is recycled to the tubular fluid to bind with another HCO_3^-. The energy to power this reabsorptive process is provided by the Na^+–K^+ pump, which keeps intracellular Na^+ concentration low so that the sodium gradient can continue to move H^+ into the tubule lumen through the Na^+–H^+ exchanger.

Normally, all of the filtered HCO_3^- is reabsorbed by this mechanism to help maintain acid–base balance. Excess H^+ ions that find no HCO_3^- in the filtrate with which to bind are excreted in the urine, and urine is normally acidic. The number of H^+ ions that can be excreted in urine is limited to a pH of about 4.0. However, urine buffers, including HPO_4^{2-} and NH_3, are secreted into the filtrate and bind with excess H^+, greatly increasing the ability of the kidney to excrete an acid load (see Fig. 26.21). Ammonia (NH_3) is produced by the renal epithelium via metabolism of amino acids. Ammonia binds to H^+ to form ammonium

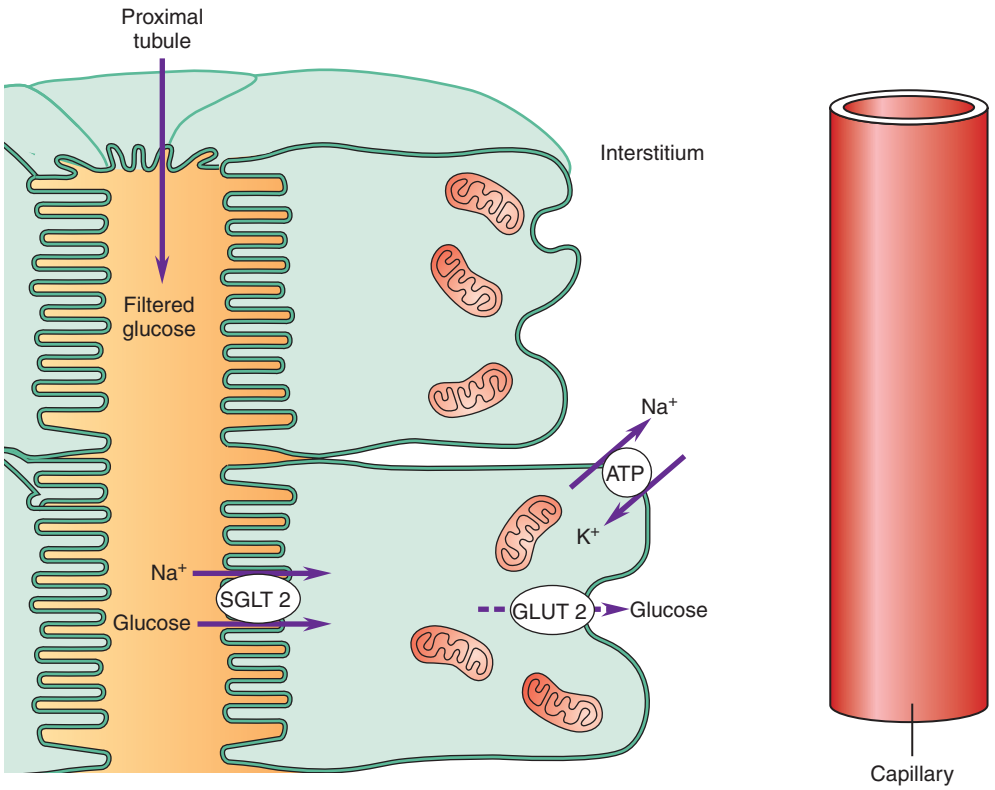


FIG 26.20 The glucose transporter in the proximal tubule (SGLT 2) is dependent on sodium reabsorption from the filtrate. The Na^+–K^+ pump in the basolateral membrane keeps the intracellular sodium level low and maintains a gradient for sodium and glucose reabsorption. Glucose diffuses out of the tubule cell and back into the interstitial fluid through passive carrier proteins (GLUT 2). *ATP,* Adenosine triphosphate.

ion (NH_4^+), whereas HPO_4^{2-} binds to H^+ to form $H_2PO_4^-$. The amino acid glutamine can also be metabolized to generate new HCO_3^- with the concurrent production of NH_4^+ that must be excreted in the urine.

Renal Compensation Process

In some cases, the kidneys are called on to compensate for an abnormality in lung function. The lungs normally regulate the amount of carbon dioxide in the blood ($PaCO_2$). When $PaCO_2$ is high, more carbonic acid is formed, and the blood pH becomes acidic. The kidneys compensate by excreting more H^+ and by creating new HCO_3^- to enhance the buffering capacity of the blood. These HCO_3^- ions are additional to those already being reabsorbed from the filtrate and are thus new. First, excess circulating CO_2 from respiratory acidosis diffuses into the renal cell and is converted to HCO_3^- and H^+ by the enzyme carbonic anhydrase. The new HCO_3^- is sent back to the bloodstream, and the new H^+ is secreted into the urine filtrate, where it binds with a renal buffer and is excreted. As mentioned previously, new HCO_3^- also can be formed by the metabolism of glutamine, and the new HCO_3^- is sent back to the bloodstream (see Fig. 26.21). Gradually, the creation of new HCO_3^- in this way increases the serum HCO_3^- concentration and restores the pH toward normal. This process may take hours to days.

The kidneys are also able to compensate for respiratory alkalosis by excreting some of the filtered HCO_3^-. Alkalosis reduces the number of H^+ ions available for transport into the filtrate. Some of the filtered HCO_3^- escapes the proximal tubule without being converted to CO_2 and is excreted in the urine.

Secretion of Potassium

There is normally a net excess of potassium from dietary sources that must be excreted by the kidneys. The primary transporter responsible for this process is the Na^+–K^+ pump in the basolateral cell membrane. The Na^+–K^+ pump moves K^+ into the tubular cell and increases the gradient for diffusion of K^+ through the apical membrane and into the filtrate (Fig. 26.22). Principal cells in the distal tubule and collecting duct are the site of potassium excretion. The activity of Na^+–K^+ pumps in these segments is sensitive to aldosterone, a steroid hormone secreted by the adrenal cortex. Aldosterone increases reabsorption of Na^+ and water and excretion of K^+. Potassium excretion also is affected by the activity of the K^+–H^+ exchanger and by the plasma K^+ concentration.

KEY POINTS

- Reabsorption across the tubular epithelium occurs by transcellular and paracellular routes. The transcellular route utilizes transporters in the apical and basolateral membranes of tubular cells to move substances from the filtrate, through the cell, to the interstitium. The paracellular route allows passive transport of substances between the tubular cells. Renal capillaries passing through the interstitium passively take up substances through filtration and diffusion and return them to the venous circulation.
- Reabsorption of glucose is accomplished by proximal tubule cell sodium-dependent transporters. These transporters have transport maxima that can be overwhelmed by excessive tubular loads of glucose, in which case glycosuria results.
- The kidneys participate in acid–base regulation through secretion of excess H^+ and reabsorption and creation of HCO_3^-. Urine buffers HPO_4^{2-} and NH_3 bind excess H^+ and increase the ability of the kidney to excrete an acid load.

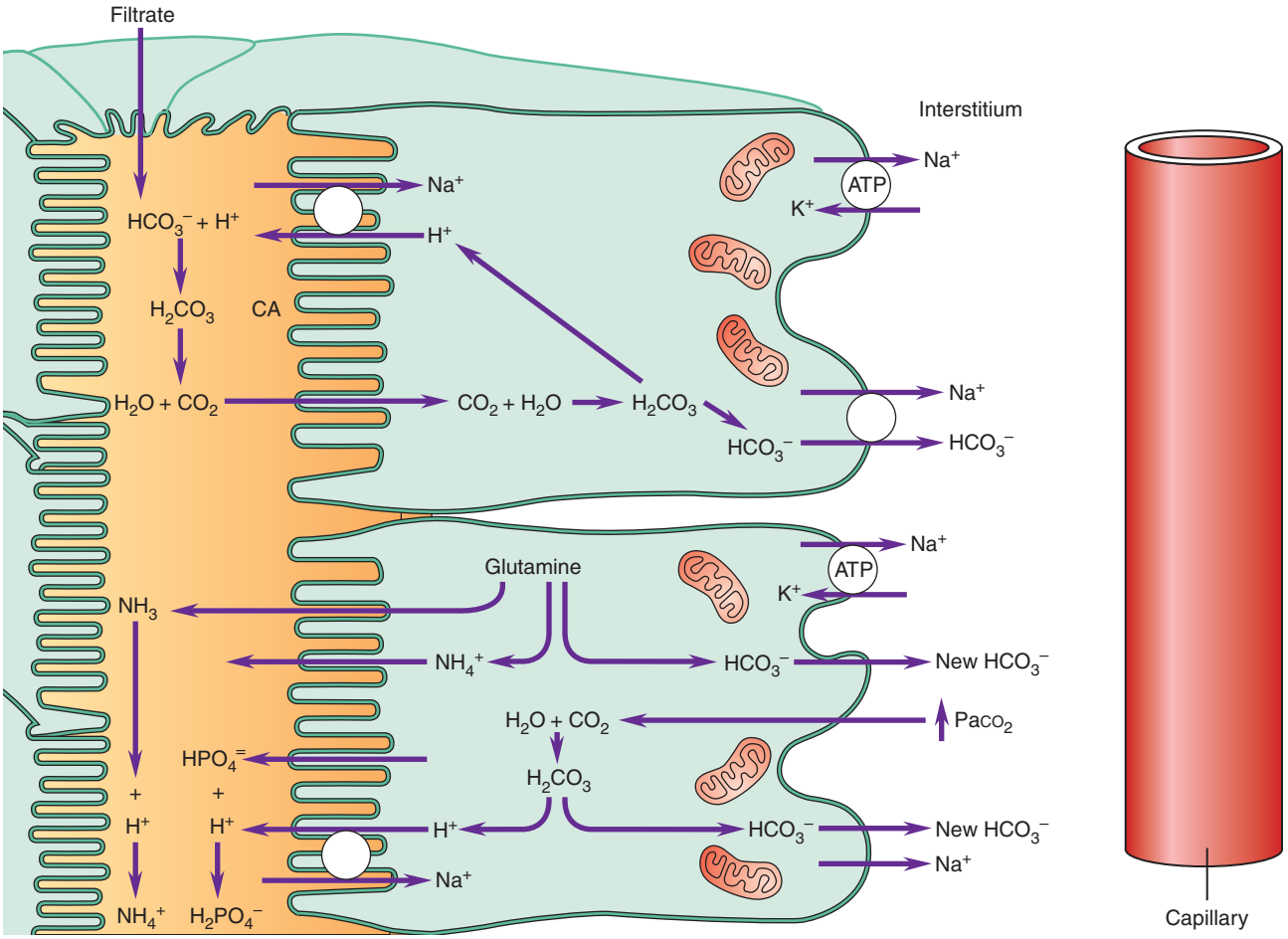


FIG 26.21 Bicarbonate ion reabsorption across the renal tubule. Filtered HCO_3^- is combined with secreted H^+ to form carbonic acid, which dissociates into water and carbon dioxide. Carbon dioxide is lipid soluble and diffuses into the cells, where the reverse reaction converts it back to HCO_3^- and H^+. The bicarbonate ion moves out of the basolateral membrane and returns to the bloodstream, whereas the H^+ is returned to the lumen to bind with another HCO_3^- ion. Excess H^+ ions are excreted in the urine in combination with phosphate and ammonia buffers. The kidney is able to create new bicarbonate as needed to maintain pH balance. *ATP,* Adenosine triphosphate; *CA,* carbonic anhydrase.

- HCO_3^- is not directly reabsorbed across the renal epithelium; it is first converted to CO_2 by the enzyme carbonic anhydrase. The H^+ ions needed for this reaction are provided by Na^+–H^+ pumps on the apical cell membrane.
- Secretion of potassium ions is promoted by activity of the Na^+–K^+ pump on the basolateral cell membrane. In the distal tubule, these pumps are regulated by aldosterone, which increases potassium excretion.

REGULATION OF BLOOD VOLUME AND OSMOLALITY

The kidneys play a vital role in maintaining normal blood volume and osmolality. As previously discussed, changes in blood volume alter the pressure in the glomerulus and affect GFR. An increase in blood volume results in a pressure diuresis, whereas a fall in blood volume reduces urine output. The kidney tubules are responsive to a number of hormonal signals that fine-tune tubular reabsorption (Table 26.2). These hormones include ADH, aldosterone, AII, ANP, urodilatin, uroguanylin, and guanylin. ADH is the principal regulator of osmolality, and the others regulate extracellular volume by increasing or decreasing NaCl and water reabsorption.

Antidiuretic Hormone

ADH (also called *vasopressin*) is secreted from the posterior pituitary when osmoreceptors located in the hypothalamus detect a high osmolality of the extracellular fluid. Principal cells in the collecting tubules respond to ADH by translocating water pores called *aquaporin 2* to the apical membrane (Fig. 26.23). These pores make the tubule permeable to water and allow water to be reabsorbed from the urinary filtrate. The high interstitial gradient of the medulla provides the osmotic force for water reabsorption. Recall that this gradient was formed by the action of powerful ion pumps in the thick ascending limb of the loop of Henle.

As water is reabsorbed into the medullary interstitium, it creates a high tissue pressure that pushes fluid into the vasa recta. The vasa recta return the reabsorbed water to the general circulation. The reabsorbed water dilutes the blood and reduces osmolality. Osmoreceptors in the brain detect the reduced osmolality and inhibit further production of ADH. When blood osmolality is too low, ADH secretion is completely inhibited, and the collecting tubules become impermeable to water. Water is not reabsorbed from the filtrate, and a large quantity of dilute urine is produced. Loss of water in excess of solute returns the blood osmolality toward normal.

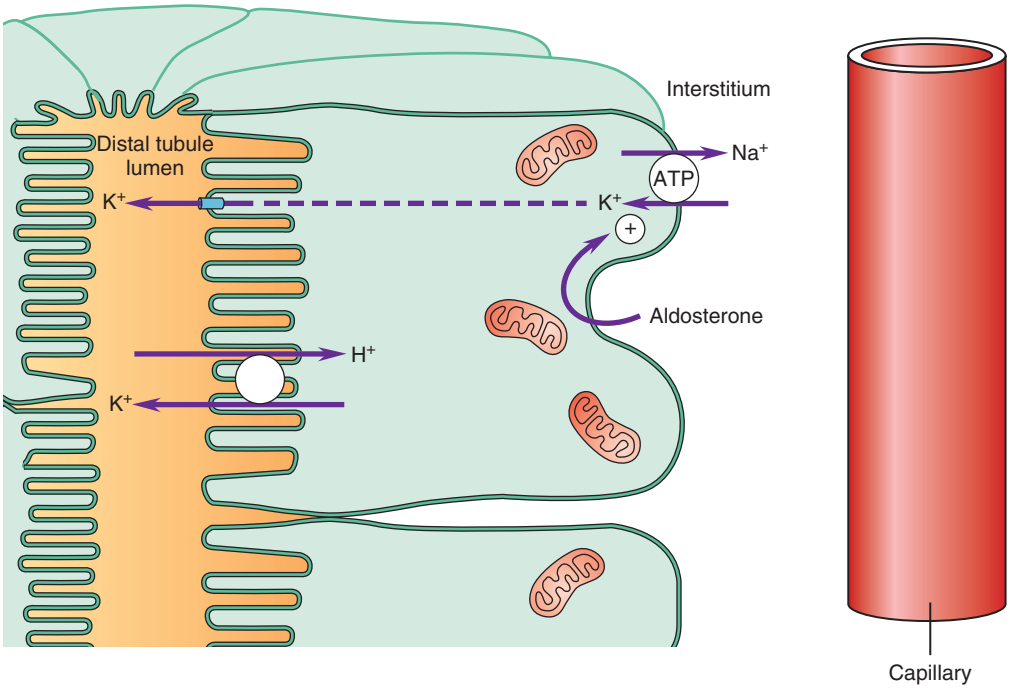


FIG 26.22 Tubular secretion of potassium ion. Increased serum potassium concentration and aldosterone increase the activity of the Na^+–K^+ pump and enhance K^+ secretion into the filtrate. The H^+–K^+ exchanger also regulates the secretion of K^+ ions. *ATP,* Adenosine triphosphate.

TABLE 26.2 Hormones that Regulate NACL and Water Reabsorption

Hormone	Major Stimulus	Nephron Site of Action	Effects on Transport
Angiotensin II	↑Renin, ↓BP	PT, TAL, DT/CD	↑NaCl and H_2O reabsorption
Aldosterone	↑Angiotensin II, ↑ $[K^+]p$	TAL, DT/CD	↑NaCl and H_2O reabsorption*
ANP, BNP, urodilatin	↑ECFV	CD	↓H_2O and NaCl reabsorption
Uroguanylin, guanylin	Oral ingestion of NaCl	PT, CD	↓H_2O and NaCl reabsorption
Sympathetic nerves	↓ECFV, ↓BP	PT, TAL, DT/CD	↑NaCl and H_2O reabsorption*
Dopamine	↑ECFV	PT	↓H_2O and NaCl reabsorption
ADH	↑P_{osm}, ↓ECFV	DT/CD	↑H_2O reabsorption*

Data from Koeppen B, Stanton B: *Berne & Levy physiology,* ed 6, Philadelphia, 2010, Mosby, p 610.
ADH, Antidiuretic hormone; *ANP,* atrial natriuretic peptide; *BNP,* B-type natriuretic peptide; *BP,* blood pressure; *CD,* collecting duct; *DT,* distal tubule; *ECFV,* extracellular fluid volume; *$[K^+]p$,* plasma K^+ concentration; *P_{osm},* plasma osmolality; *PT,* proximal tubule; *TAL,* thick ascending limb.
*The effect on H_2O reabsorption does not include the thick ascending limb or the early portion of the distal tube.

An insufficiency of ADH secondary to pituitary damage results in the condition of *diabetes insipidus* in which large volumes of dilute urine are excreted, leading to severe fluid imbalances. A similar problem occurs when the collecting tubules are unresponsive to ADH. This condition is called *nephrogenic diabetes insipidus* and usually results from genetic defects in either the ADH receptor (V_2) or the aquaporin 2 genes.

Aldosterone, Angiotensin II, Natriuretic Peptides, Urodilatin, Uroguanylin, and Guanylin

Aldosterone, AII, ANP, urodilatin, uroguanylin, and guanylin alter blood volume without affecting its concentration. Aldosterone and AII increase reabsorption of Na^+, which provides a gradient for water reabsorption. Because salt and water are reabsorbed together, the osmolality of the reabsorbed fluid is isosmotic with plasma.

AII and aldosterone are produced when the juxtaglomerular cells in the kidney are stimulated to release renin. Renin is released in response to (1) decreased blood flow to the kidney, (2) reduced serum sodium levels, and (3) activation of sympathetic nerves to the juxtaglomerular cells. Renin begins a cascade of reactions that result in the production of AII and aldosterone. When AII and aldosterone restore blood volume and blood pressure to normal, the stimuli for renin release are removed and the concentrations of AII and aldosterone fall.

ANP is released from atrial cells in the heart when the chamber is overstretched by excessive blood volume. ANP inhibits all of the actions of AII and results in loss of sodium and water in the urine. Thus ANP reduces extracellular volume, but the fluid losses are isosmotic with plasma, and blood osmolality remains unchanged. Urodilatin is a peptide that is secreted by distal and collecting tubule cells in response to increased circulating volume. It is similar in structure and function to ANP and inhibits Na^+ and water reabsorption by the collecting duct.

Uroguanylin and guanylin are peptide hormones produced by neuroendocrine cells in the intestine in response to NaCl ingestion. The targets for these hormones are guanylyl cyclase receptors located

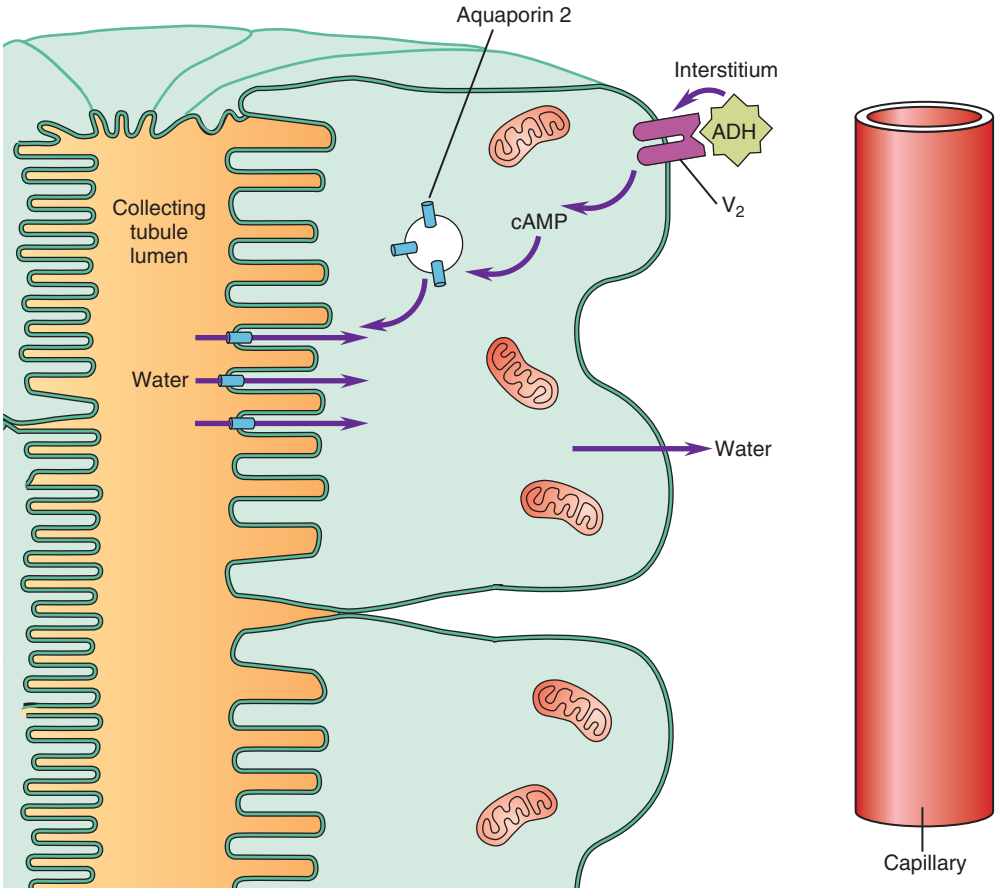


FIG 26.23 Antidiuretic hormone (ADH) action on the collecting tubule epithelium. ADH binds to receptors on the basolateral cell membrane, resulting in translocation of water pores (aquaporin 2) to the apical surface. Increased water permeability results in reabsorption of water from the filtrate and into the interstitium. *cAMP,* Cyclic adenosine monophosphate; *V2,* vasopressin-2 receptor.

TABLE 26.3 Commonly Used Diuretics

Diuretic	Action
Osmotic diuretics	Increase solute load in tubule
ACE inhibitors	Block production of AII and aldosterone
Loop diuretics	Block Na^+–K^+–$2Cl^-$ transporter in ascending loop of Henle
Thiazide-like diuretics	Block Na^+ reabsorption in distal tubule
Aldosterone inhibitors	Block action of aldosterone on distal tubule Na^+–K^+ transporters

ACE, Angiotensin-converting enzyme.

on cells in the proximal tubule and collecting duct. Binding to these receptors generates cGMP, which inhibits Na^+, Cl^-, and water reabsorption and produces an effect similar to that of natriuretic peptides and urodilatin.

Diuretic Agents

The ability of the kidneys to reabsorb fluid can be inhibited by drugs that block sodium and water reabsorption. These agents are called *diuretics* and include osmotic diuretics, ACE inhibitors, loop diuretics, thiazide-like diuretics, and inhibitors of aldosterone activity (Table 26.3). Diuretics work by altering osmotic gradients in the kidney tubules so that reabsorption of water is inhibited. Recall that water always moves passively according to an osmotic gradient. When the solute content of the filtrate is elevated, reabsorption of water is inhibited, resulting in a larger output of urine.

Osmotic diuretics (e.g., mannitol) are filtered through the glomerulus and are not reabsorbed by the tubules. The osmolality of the filtrate is increased by the presence of the solute, and more water remains in the tubule and is excreted in the urine. ACE inhibitors (e.g., captopril) inhibit the formation of AII and aldosterone, which normally stimulate the kidney tubules to reabsorb Na^+. In the absence of these hormones, more Na^+ stays in the urinary filtrate, resulting in less reabsorption of water.

Loop diuretics (e.g., furosemide) block the Na^+–K^+–$2Cl^-$ pumps in the ascending loop of Henle. The ions that would normally have been pumped into the interstitium stay in the filtrate and hold water with them. In addition, the maintenance of the high interstitial gradient in the medulla may be impaired. Washout of the gradient reduces the force for water reabsorption from the collecting ducts.

Thiazide-like diuretics (e.g., hydrochlorothiazide) block Na^+ reabsorption in the distal tubule. Sodium ions remain in the filtrate and oppose the action of the interstitial osmotic gradient.

All of these agents also increase the excretion of K^+ and are called *potassium-wasting diuretics.* Patients receiving chronic diuretic therapy with these agents usually require potassium replacement therapy.

In contrast, the aldosterone-blocking agents (e.g., spironolactone) are potassium sparing. Recall that aldosterone increases activity of the Na^+–K^+ pumps on the basolateral membrane of the distal tubule cells.

These pumps promote Na^+ and water reabsorption and potassium secretion. Blockage of aldosterone reduces the activity of these pumps and results in less sodium and water reabsorption as well as less potassium excretion. Significant elevations in serum K^+ concentration can occur with these agents.

Diuretics are used primarily in the management of high blood pressure (see Chapter 16) and congestive heart failure (see Chapter 19), but they also may be used in the diagnostic phase of acute renal failure or to manage potassium overload.

KEY POINTS

- The kidneys regulate blood volume and osmolality by altering GFR and reabsorption from the urinary filtrate.
- Changes in blood volume alter the filtration pressure in the glomerulus, resulting in a pressure diuresis when blood volume is high and in reduced filtration and fluid conservation when blood volume is low.
- The kidney tubules are responsive to hormones that alter their reabsorptive properties. Antidiuretic hormone (ADH) increases the permeability of the collecting tubule to water, resulting in increased reabsorption and reduced blood osmolality.
- Aldosterone, AII, ANP, and urodilatin alter blood volume without affecting blood osmolality. Aldosterone and AII increase sodium and water reabsorption, whereas ANP and urodilatin inhibit their reabsorption.
- Diuretics alter the osmolality of the urinary filtrate and oppose the reabsorption of water, resulting in an increase in urine volume.

ENDOCRINE FUNCTIONS

The kidney is the source of two important endocrine hormones: erythropoietin and active vitamin D. Secretion of these hormones is impaired in chronic kidney disease and contributes to the anemia and osteodystrophy found in this disorder (see Chapter 28).

Erythropoietin

Erythropoietin is a peptide growth factor that stimulates erythrocyte development in the bone marrow. The regulation of erythropoietin secretion is not completely understood; however, hypoxemia and decreased circulating red cell mass are known to increase its release (see Chapter 13). Erythropoietin is commercially available in a recombinant form that can be given parenterally. The anemia of chronic kidney disease usually responds well to erythropoietin replacement therapy.

Vitamin D

Synthesis of active vitamin D is an interdependent function of the skin, liver, and kidney. The precursors to active vitamin D can be formed in the skin in response to the ultraviolet rays in sunlight, or they can be ingested in fortified food products. These precursors (cholecalciferol) then must undergo a series of two hydroxylations to become active. The first occurs in the liver, resulting in the formation of 25-hydroxycholecalciferol. The kidney performs the second hydroxylation to form 1,25-dihydroxycholecalciferol, which is the active form of vitamin D. Vitamin D is a necessary cofactor for calcium absorption from the intestine. It may also facilitate calcium reabsorption in the kidney tubules.

In chronic kidney disease, the production of active vitamin D is impaired, resulting in poor calcium absorption from the intestine and low serum calcium levels. A low serum calcium level is the stimulus for parathyroid hormone release, resulting in removal of calcium and phosphate from the bones. Eventually excessive parathyroid hormone activity leads to the condition of osteodystrophy and predisposes to skeletal fractures (see Chapter 28).

KEY POINTS

- The kidney secretes two important endocrine hormones: erythropoietin, a growth factor for red blood cells (RBCs), and active vitamin D, a necessary cofactor for calcium absorption from the intestine.
- In chronic kidney disease, impaired production of these hormones results in anemia and osteodystrophy.

AGE-RELATED CHANGES IN RENAL FUNCTION

Infant

In the early postnatal period, the GFR is less than half the adult rate. Due to an immature ability of the infant kidneys to regulate urine osmolality, infants are predisposed to volume depletion during fluid losses, such as those that occur with diarrhea, fever, fluid restrictions, or decreased intake. Volume regulation improves as the kidney matures. In addition, the glomerular and tubular basement membranes thicken, the glomeruli become increasingly permeable, and the loops of Henle lengthen. Systemic changes, such as increased cardiac output and increased levels of plasma proteins, also influence the improvement in renal function. Between the first and second years of life, renal function essentially reaches maturity. Thereafter the kidney grows in proportion to overall body growth, reaching maximal size between 35 and 40 years of age. Renal function in infants is described in the Pediatrics Considerations box.

Adult and Elderly

As part of the normal aging process, the kidney begins to diminish in size and function after the fourth decade and more significantly by the middle of the sixth decade. The effects of aging on renal function are described in Geriatric Considerations: Changes in the Renal System. After age 40 years, the number of glomeruli begins to decrease, and by age 70 years it is estimated that as many as 30% to 50% of the glomeruli have been lost. Renal blood flow decreases after the fourth decade at an approximate rate of 10% per decade because of vascular changes, especially in the cortical blood vessels.

As the number of functional nephrons decreases with aging, there is less renal reserve. Under normal conditions, the kidney functions adequately; however, elderly people are much more susceptible to fluid and electrolyte imbalances and renal damage. The elderly are also very susceptible to kidney damage from drugs and medications, including contrast media. Consequently, it is essential that renal function be evaluated before, during, and after they receive these agents or substances.

KEY POINTS

- Renal function is impaired at both ends of the life span. Infants have reduced ability to make concentrated urine because of kidney immaturity. Aged individuals have reduced numbers of functioning nephrons, reduced renal blood flow and glomerular filtration rate (GFR), and decreased ability to conserve salt and water.
- The very young and very old are at increased risk for fluid and electrolyte imbalances and drug toxicity.

TESTS OF RENAL STRUCTURE AND FUNCTION

Urinalysis, serum creatinine level, blood urea nitrogen (BUN) levels, and tests of GFR are most helpful in evaluating kidney function, whereas other diagnostic tests are best for evaluating kidney structure.

Urine and Blood Studies

Routine assessment of urine is commonly performed to screen for a variety of kidney and metabolic disorders. Serum creatinine and BUN

TABLE 26.4 Normal Composition of Urine

Characteristics	Normal Value
Color	Light yellow to amber
pH	4.5–8.0
Specific gravity	1.003–1.030
Red blood cells	<5/HPF
White blood cells	<5/HPF
Protein	Negative
Glucose	Negative
Ketones	Negative
Nitrites	Negative
Casts	None
Crystals	None

HPF, High-powered field.

levels are used to monitor the progression of renal disease or to screen for occult renal insufficiency.

Urinalysis

Most often urinalysis is done on a single voided sample; however, longer collections may be done for quantitative analysis. Urinalysis assesses urine color; clarity; odor; specific gravity; pH; and concentrations of glucose, ketones, protein, and sediment (including cells, crystals, casts, and bacteria or other organisms). The first urine voided in the morning is the most concentrated, due to overnight fasting, and therefore is the best specimen to use for a routine or baseline urinalysis, especially to assess pH, osmolality, and sediment (Table 26.4).

A 24-hour urine collection measures the total quantity of a substance or substances excreted in a day. This is helpful for evaluating substances that are excreted in varying concentrations throughout the day, such as hormones, creatinine, protein, urea, and glucose.

Urine microscopy and culture and sensitivity tests assess the urine for the presence of microorganisms and accompanying cells and to determine the medications or drugs to which the organisms are most sensitive. For these tests, a few milliliters of urine is collected by the clean-catch method and placed into a sterile container.

Urine is approximately 95% water and contains varying amounts of water-soluble waste products. Freshly voided urine has a slight odor attributable to the breakdown of urea to ammonia. If urine stands for a period of time or has a large bacteria population, it will have a strong ammonia smell. The ingestion and excretion of certain foods, such as asparagus, or of certain medications, such as vitamins, may cause urine to have a different odor.

The pale yellow to amber color of urine is due to the presence of urochrome pigments. Urine color can change because of the presence of cells or because of an increased urine concentration. The presence of RBCs, or hematuria, can cause urine color to range from brown to bright red. White blood cells (WBCs) can make urine look cloudy. Concentrated urine is usually dark yellow to orange. Certain foods and drugs can change urine color. For instance, if beets have been eaten, the urine may be burgundy, and if the individual has taken phenazopyridine (Pyridium), the urine may be orange.

Normally, urine is clear and slightly acidic, although the pH range is 4.5 to 8.0. Urine allowed to stand undisturbed will become cloudy and alkaline because of the breakdown of urea to ammonia, which increases the pH. Cloudiness can result from the presence of cells, bacteria, crystals, casts, or fat substances.

Urine specific gravity and urine osmolality are measures of the concentration of solute in the urine. Urine specific gravity varies with the amount of solids in the urine, such as cells, casts, and microorganisms,

PEDIATRIC CONSIDERATIONS

Changes in the Kidneys in Infants

At birth, the loss of placental blood flow and the rapid increase in renal blood flow lead to high vascular resistance in the kidneys. The immature kidneys respond by temporarily reducing renal blood flow and filtration to compensate. The filtration capacity of the glomeruli is reduced. The kidneys cannot adequately concentrate urine to conserve body water. This makes the child susceptible to water loss. In the first 24 hours of life, the newborn will have decreased urine output. Ninety-five percent of infants will pass urine in the first 24 hours of life, although the amount is small (about 20 mL) (Hockenberry & Wilson, 2011). As the infant increases the number and volume of feedings in the first few weeks of life, the capillary resistance is reduced and filtration is increased by the glomeruli.

The kidneys have small immature nephrons. The nephrons are lined with cuboid epithelium, which limits the function of the nephrons. The cuboid epithelium is not replaced by pavement epithelium and fully functioning until after the first year of life (MacGregor, 2008). The immaturity of the nephrons also makes the distal convoluted tubules resistant to aldosterone. The immature nephrons also have short loops of Henle, where water and sodium levels are normally adjusted. The short loops of Henle and the resistance to aldosterone make it difficult for an infant to excrete excess sodium. The presence of excess sodium increases the interstitial osmolality, which will decrease the glomerular filtration rate in the infant. The decreased glomerular filtration rate contributes to the reduced production of urine that is seen with the infant. As the child grows and the kidney matures, the glomerular filtration rate will increase. The glomerular filtration rate triples by 9 months of age and reaches 30% of adult values by 2 years of age (MacGregor, 2008).

Hockenberry MJ, Wilson D: *Wong's nursing care of infants and children,* ed 9, St Louis, MO, 2011, Mosby. MacGregor J: *Introduction to the anatomy and physiology of children: a guide for students of nursing, child care and health,* ed 2, New York, 2008, Routledge.

GERIATRIC CONSIDERATIONS

Changes in the Renal System

In the aging individual, there is a 30% to 50% decrease in the number, size, weight, and function of the nephrons, with an accompanying reduction in the size and weight of the kidney. There is increasing interstitial fibrosis of the renal afferent arterioles. Loss of nephrons and diminished renal blood flow contribute to a decrease in the GFR.

There is also a decrease in the length and the excretory and reabsorptive capabilities of the tubules. The tubule changes affect the countercurrent mechanism, leading to significant changes in urine concentration, excretion, and absorption. Specifically, the changes include reduced urine concentration, decreased sodium retention, diminished drug and metabolite excretion, decreased hydrogen ion (H^+) secretion, and increased renal threshold for glucose. With aging, the kidney does not respond quickly to correct pH or sodium imbalances.

With aging, urinary muscles weaken, and sphincter tone and bladder capacity decrease. This increased muscular weakness can lead to a rise in the residual volume in the bladder and difficulty in starting the urinary stream. The length of the urethra decreases. There is also less bladder innervation and a reduced sensation of filling. A loss of the diurnal excretory pattern induces nocturia.

but urine osmolality is not affected by these substances. Thus urine osmolality is a more accurate measure of the kidneys' ability to concentrate and dilute the urine. The range for specific gravity is 1.003 to 1.030, with the higher number indicating a more concentrated urine. Usually urine osmolality and specific gravity vary throughout the day and from day to day. Results that remain fixed over consecutive voidings and days could be an indicator of renal disease.

Normal urine contains little or no protein. A small amount of protein in the urine is insignificant, but excretion of more than 150 mg per 24 hours should be investigated because it could indicate glomerular capillary disease. Proteinuria can cause urine to be foamy.

Glycosuria, or glucose in the urine, is abnormal and usually indicates hyperglycemia (elevated blood glucose level), which can occur with diabetes mellitus or after an excessive ingestion of sugar. Rarely does glycosuria indicate renal disease.

A few epithelial cells, erythrocytes, leukocytes, and bacteria are normally found in urine. Fewer than five RBCs or WBCs per high-powered field is considered to be within normal limits. An excess of any of these cells may indicate a pathologic process; however, collection technique and presence of menstrual blood may be confounding factors.

Crystals and stones are not usually found in the urine. Either can originate anywhere along the urinary tract. If found in the urine, their composition should be identified and the urinary tract assessed for more crystals and stones (see Chapter 27).

When urinary casts are present, they provide important clues for differentiating renal diseases. Casts are formed in the nephron tubule and are composed of a protein meshwork with entrapped cells or cell fragments. Cells in the thick ascending loop of Henle produce and secrete small amounts of a glycoprotein called *Tamm-Horsfall* protein (also called *uromodulin*), into the tubular lumen. This protein forms the meshwork that entraps cells in the tubule to form casts and is found in normal urine. Normal protein casts that do not have cells in them are called *agranular casts.* There are many types of abnormal casts, each associated with certain renal pathologic conditions. For example, WBC casts are associated with renal infections (pyelonephritis), RBC casts indicate inflammation of the glomerulus (glomerulonephritis), and epithelial cell casts indicate sloughing of tubular cells (acute tubular necrosis).

Serum Creatinine and Blood Urea Nitrogen

Creatinine is an end product of muscle metabolism that is excreted exclusively by the kidney. The serum creatinine level averages approximately 0.7 to 1.5 mg/dL and is relatively constant throughout the day and from day to day. Creatinine levels are slightly higher in men than in women because of men's larger muscle mass.

Serum creatinine level is a fairly reliable indicator of renal function because it is affected by only two factors: (1) the rate of creatinine produced from muscle, which is relatively constant in the absence of muscle breakdown; and (2) the rate of creatinine excreted by the kidney, which is determined primarily by the GFR. Therefore the GFR is reflected in the serum creatinine level. For instance, when the GFR decreases by half, the concentration of creatinine in the serum doubles. A rise in serum creatinine level indicates a decrease in renal function.

Urea is an end product of protein metabolism. It is excreted primarily by the kidney and measured in the blood as blood urea nitrogen (BUN). The BUN averages approximately 10 to 20 mg/dL and rises with a decrease in renal function, a decrease in fluid volume, and an increase in catabolism and dietary protein intake. When a change in renal function occurs, the BUN measurement tends to change more rapidly than the creatinine level; however, the BUN value is less specific. Often BUN and creatinine are measured together, and the ratio is determined. Acute changes in GFR are reflected in a higher BUN-to-creatinine ratio, usually greater than 20:1.

Measures of Glomerular Filtration Rate

GFR is an important parameter in the assessment of renal function. It is commonly measured by evaluating the clearance of a filterable substance from the plasma. Creatinine clearance is frequently used to assess GFR, but it is not completely accurate because some secretion and reabsorption occur in the nephron tubules. At low GFR, creatinine clearance is quite unreliable.

A more accurate measurement of GFR is obtained by using inulin, an inert substance that is filtered freely at the glomerulus and is completely unaffected by tubular secretion and reabsorption. The use of inulin is more expensive and cumbersome than creatinine clearance because it must be injected. The formula for measuring clearance is the same regardless of the marker substance used. Creatinine clearance is used in the following example, but the corresponding values for inulin can be substituted in the equation.

Creatinine clearance estimates the GFR by measuring the amount of blood that is cleared of creatinine each minute. Usually a 24-hour urine specimen and a blood specimen at the midpoint of the urine collection are used to determine creatinine clearance; however, shorter intervals can be used. The measured values are calculated in the following formula:

$$\text{Clearance} = \frac{\text{Urine volume (ml/min)} \times \text{urinary creatinine (mg/dL)}}{\text{Plasma creatinine (mg/dL)}}$$

An estimate of GFR called the *Modification of Diet in Renal Disease study equation* (MDRD) can be made using only patient demographics and serum creatinine value (S_{cr}). This estimate is based on an average body surface area for an adult of 1.73 m^2. It does not require urine collection, making it simple compared with other measures of clearance:

$$\text{GFR (mL/min/1.73 m}^2) = 186 \times (S_{cr})^{-1.154} \times (\text{Age})^{-0.203} \times (0.742 \text{ if female}) \times (1.210 \text{ if African American})$$

Diagnostic Tests

Although studies of urine and blood are good indicators of renal function, they often are not adequate to determine the underlying pathologic process. Diagnostic tests are helpful in assessing structural abnormalities, such as tumors or obstructions, congenital anomalies, perfusion defects, and histologic abnormalities. Sometimes a combination of diagnostic tests is necessary.

Kidney, Ureter, and Bladder Roentgenography

A kidney, ureter, and bladder (KUB) roentgenography is a plain radiograph (x-ray) taken of the abdomen to visualize the kidneys, ureters, and bladder. A KUB study shows the position, shape, size, and number of macroscopic or gross renal, ureteral, and bladder structures and surrounding bones. In addition, foreign bodies, radiopaque objects, stones, and neoplasms can be seen on KUB. The KUB may serve as a screening examination to inform further diagnostic testing.

Intravenous Urography/Pyelography

During intravenous urography, also called *intravenous pyelography (IVP),* an iodine-containing radiopaque dye is injected into a vein; it circulates through the kidney and is excreted in the urine. A rapid series of radiographs is made as the dye is being excreted. This test shows the size, shape, and location of urinary tract structures and can be used to

evaluate renal excretory function. The dye is nephrotoxic, meaning poisonous to the kidney, and allergenic to some people. A state of hydration helps the dye pass through the kidney and prevents renal damage. Because fecal matter and gas in the intestinal tract will interfere with visualization of the kidneys and ureters on the radiographs, a laxative or enema may be indicated before IVP.

Radionuclide Studies

Renograms and renal scans are diagnostic studies that use radioactive isotopes to assess kidney structure and function. In general, the renogram is more useful for assessing function, whereas the renal scan is better at detecting structural anomalies. During a renogram procedure, a small amount of filterable radioactive material is administered intravenously. It circulates through the kidney and is excreted in the urine. While the radionuclide circulates through the renal vessels and nephrons, a radiation detection probe counts the activity of the radioactive substance and simultaneously creates a graphic record of the activity. This test assesses renal function by measuring renal blood flow, glomerular filtration, and tubular secretion.

The renal scan uses a radionuclide that tends to accumulate in areas that are well perfused by blood. The renal scan images depict the concentration of the radionuclide in the kidney and provide anatomic and some physiologic information. In the presence of tumors or nonfunctioning areas, the radioactive material will not be detected by the scan.

A more dynamic assessment of renal physiology can be obtained using positron emission tomography (PET) or single-photon emission computed tomography. These modalities use scintigraphic imaging to view the kidney and can pick up subtle, dynamic changes. Regional differences in GFR, for example, can be detected by PET scan.

Ultrasonography

Ultrasonography is a noninvasive, painless procedure that uses high-frequency sound waves to image renal structures. The sound waves are at a frequency above the limit of human hearing. Ultrasound is used because its short wavelength produces a more detailed picture or image than other types of sound waves. A probe with a transducer inside is held against the back and emits ultrasound waves that travel through tissue to the kidney and reflect off the kidney, back to the probe. Ultrasonography demonstrates gross renal anatomy, true kidney depth, structural abnormalities, and perirenal masses, and it can be used to distinguish between a fluid-filled cyst and a solid tumor.

Computed Tomography

Computed tomography (CT) combines roentgenography with computer technology and is a noninvasive, painless procedure. Instead of using broad x-ray beams, CT uses thin x-ray beams, each about 10 degrees apart. The information obtained during scanning is transmitted to a computer, which constructs a tomograph and calculates its density. Because the kidneys are located deep within the abdominal cavity, they opacify better after an IV injection of a contrast agent. CT shows more detail than ultrasonography. CT can demonstrate perirenal and renal masses, renal vascular disorders, and filling defects of the collecting system.

Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) is a painless, noninvasive procedure that does not use x-rays or radioactive markers. The imager applies a strong magnetic field that causes protons to align themselves with the magnetic field. Pulses of radio waves are emitted that cause the magnetic fields to rotate or resonate. The rotating fields induce electrical signals that the computer analyzes and uses to create images or pictures on a screen. The renal images are available in all planes and show more detail than the images achievable with CT. Newer methods of MRI have been developed to obtain dynamic images using the movement of contrast dye through the kidney. Sequential fast-pulse imaging (functional MRI) allows assessment of obstructions, vascular disorders, and renal insufficiency.

Renal Biopsy

The purpose of a kidney biopsy is to obtain renal tissue that may be studied to determine the nature and extent of renal disease for diagnosis, management, and prognosis. The renal tissue is studied histologically by light and electron microscopy and immunofluorescence. Some indications for a kidney biopsy are persistent proteinuria, hematuria originating from the kidney, unexplained acute renal failure, glomerular disease, renal mass, rejection of a transplanted kidney, and renal involvement in systemic disease.

KEY POINTS

- Urinalysis provides important information about kidney function. Normal urine is clear, pale yellow to amber, and slightly acidic, and it may contain a few cells. Urine osmolality and specific gravity normally vary over the course of the day, depending on fluid intake. Urine is abnormal if it is cloudy or malodorous or contains protein, red blood cells (RBCs), crystals, stones, or casts. A fixed osmolality or specific gravity may indicate renal impairment.
- Serum creatinine and blood urea nitrogen (BUN) measurements are useful indicators of renal function. Serum creatinine level is a more reliable indicator of renal function than BUN measurement. In conditions of reduced glomerular filtration rate (GFR), serum creatinine and BUN levels increase.
- GFR can be estimated by measuring the clearance of a filterable substance from the urine. Creatinine clearance is frequently used for this purpose, but it is not completely accurate because of some tubular processing. Inulin clearance provides a more accurate measurement of GFR. The MDRD is a simple calculated method of estimating GFR using serum creatinine values.
- Diagnostic studies used to evaluate kidney structure and function include plain radiography, pyelography, radionuclide studies, ultrasound, CT, and MRI. Renal biopsy may be performed to obtain tissue for histologic examination.

SUMMARY

The kidneys have a vital role in excreting water-soluble waste products and maintaining fluid, electrolyte, and acid–base homeostasis. To perform these functions, the kidneys must have a sufficient GFR. Most waste products are removed by filtration rather than by secretion; thus a reduced GFR results in accumulation of wastes in the blood. The kidney has a large renal reserve and accomplishes its functions well until more than 75% of the nephron mass is dysfunctional.

The nephron is the structural and functional unit of the kidney. It performs three essential functions: filtration, secretion, and reabsorption. Filtration occurs at the glomerulus at a rate of about 125 mL/min. The composition of filtrate is similar to that of blood except that proteins and blood cells are absent. Normally 99% of the filtrate is reabsorbed along the nephron tubules, resulting in the elimination of 30 to 60 mL/hr of concentrated urine. Each nephron regulates its own GFR through

tubuloglomerular feedback to prevent overloading its reabsorptive capacities.

The kidneys are responsive to a number of endocrine hormones that regulate blood osmolality and volume, including ADH, aldosterone, AII, ANP, urodilatin, uroguanylin, and guanylin. In addition, the kidneys produce two important endocrine hormones: erythropoietin and vitamin D. Urinalysis, serum creatinine and BUN levels, and tests of GFR are important indicators of renal function. Structural abnormalities can be assessed by a variety of imaging techniques.

RESOURCES

Fehrman-Ekholm I, Kvarnström N, Söfteland JM, et al: Post-nephrectomy development of renal function in living kidney donors: a cross-sectional retrospective study. *Nephrol Dial Transplant* 26(7):2377–2381, 2011.

Hall JE, editor: *Guyton and Hall textbook of medical physiology*, ed 13, Philadelphia, 2015, Saunders.

Harris RC: COX-2 and the kidney. *J Cardiovasc Pharmacol* 47(Suppl 1):S37–S42, 2006.

Jarvis C: *Physical examination and health assessment*, ed 7, Philadelphia, 2015, Saunders.

Levey AS, Bosch JP, Lewis JB, et al: A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group. *Ann Intern Med* 130(6):461–470, 1999.

Mjøen G, et al: Long-term risks for kidney donors. *Kidney Int* 86:162–167, 2014.

Ramcharan T, Matas AJ: Long-term (20-37 years) follow-up of living kidney donors. *Am J Transplant* 2(10):959–964, 2002.

Schnermann J: Concurrent activation of multiple vasoactive signaling pathways in vasoconstriction caused by tubuloglomerular feedback: a quantitative assessment. *Annu Rev Physiol* 77:301–322, 2015.

Stanton BA, Koeppen BM: Elements of renal function. In Koeppen BM, Stanton BA, editors: *Berne & Levy physiology*, ed 6, Philadelphia, 2010, Mosby, pp 557–577.

Stockand JD, Sansom SC: Glomerular mesangial cells: electrophysiology and regulation of contraction. *Physiol Rev* 78:723–744, 1998.

27

Intrarenal Disorders

Jacquelyn L. Banasik and Roberta J. Emerson

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- How are the locations of renal pain, findings on urinalysis, and results of other diagnostic tests used to differentiate the causes of kidney disease?
- How are renal tumors differentiated, detected, and managed?
- How do autosomal-dominant and autosomal-recessive forms of polycystic kidney disease differ?
- What risk factors and clinical findings are associated with pyelonephritis?
- What physiologic and pathophysiologic disorders predispose to the formation of renal calculi of differing compositions?
- How are the various forms of glomerulonephritis differentiated?
- What laboratory and clinical findings suggest a diagnosis of nephrotic syndrome?

CHAPTER OUTLINE

Common Manifestations of Kidney Disease, 575
Pain, 575
Abnormal Urinalysis Findings, 576
Other Diagnostic Tests, 576
Congenital Abnormalities, 577
Renal Agenesis and Hypoplasia, 577
Cystic Kidney Diseases, 578
Autosomal-Recessive Polycystic Kidney Disease, 578
Autosomal-Dominant Polycystic Kidney Disease, 578
Neoplasms, 579
Benign Renal Neoplasms, 579
Renal Cell Carcinoma, 579
Nephroblastoma (Wilms Tumor), 580
Infection, 581
Acute Pyelonephritis, 582
Chronic Pyelonephritis, 582
Obstruction, 583
Renal Calculi (Nephrolithiasis), 584
Glomerular Disorders (Glomerulopathies), 586
Glomerulonephritis, 587
Acute Glomerulonephritis, 588
Crescentic Glomerulonephritis/Rapidly Progressive Glomerulonephritis (RPGN), 589
Chronic Glomerulonephritis, 589
Nephrotic Syndrome, 589
Membranous Nephropathy (MN), 590
Focal Segmental Glomerulosclerosis (FSGS), 590
Minimal Change Disease (MCD), 591

Functional kidneys are necessary for the removal of waste products from the blood and the maintenance of fluid, electrolyte, and acid–base balance despite wide variations in intake and losses. Systemic disorders that alter the delivery of blood flow to the kidney may adversely affect the kidney's ability to perform its filtering and homeostatic functions. In addition, many disorders occur primarily within the kidney and have the potential to result in chronic kidney disease or end-stage renal disease (ESRD). In general, these disorders can be categorized as (1) congenital, (2) neoplastic, (3) infectious, (4) obstructive, and (5) glomerular.

COMMON MANIFESTATIONS OF KIDNEY DISEASE

Pain

Thorough pain assessment is an essential component of the history and physical examination of any patient. The results can be useful in localizing the etiology of the pain, but assessment is also challenging because pain perceived as coming from the abdomen can originate from many varied organs and tissues within the abdomen or extraabdominally (see Chapter 47). Pain associated with the urinary tract may originate from the lower urinary tract (ureters, bladder, or urethra) or

the kidney itself. Renal or kidney pain is also referred to as *nephralgia* (*-algia* is from the Greek *algos,* meaning pain). Extensive damage to a kidney can occur without nephralgia because most of the kidney lacks pain receptors. However, the renal capsule is innervated by nociceptors, and when a disease process causes it to be distended, inflamed, or punctured, a dull to sharp pain is felt. Pain may be caused by intrarenal fluid accumulation, infected or bleeding cysts, hemorrhage from blunt trauma, or neoplastic expansion. In addition, whenever the renal capsule is penetrated (e.g., during biopsy or trauma), a dull pain or intense pressure may be felt. The renal pelvis and the rest of the urinary tract are innervated by many pain receptors. Obstruction of the intrarenal collecting system causes pain if the obstruction leads to distention of the renal pelvis or capsule. Large calculi, however, can develop insidiously in the renal pelvis or calices and may be painless until they start to move into the ureteral junction. Ischemia caused by the occlusion of renal blood vessels (e.g., from an embolus, atherosclerotic disease, or neoplasm) results in a constant dull or sharp pain.

Pain associated with intrarenal disorders affecting the capsule is classically assessed by palpation or light percussion over the costovertebral angle (CVA) posteriorly and is recorded as *CVA tenderness.* Sympathetic nerves transmit signals from renal and ureteral nociceptors to the spinal cord between the T10 and L1 levels, and the pain may be felt throughout the corresponding dermatomes. A dermatome is an area of skin innervated by a specific spinal cord segment (Fig. 27.1). Visceral and cutaneous afferent fibers enter the spinal cord in close proximity and converge on some of the same neurons at the spinal, thalamic, and cortical levels of the central nervous system. When visceral pain fibers are stimulated, concurrent stimulation of cutaneous fibers occurs and the visceral pain is perceived as though it had originated in the skin. Nerve fibers from the renal plexus communicate with the spermatic plexus, and because of this association, scrotal pain in males and labial pain in females may accompany renal pain.

Abnormal Urinalysis Findings

Urinalysis is an essential laboratory test for all suspected problems of the genitourinary system. After history taking and a physical examination, urinalysis generally serves as a starting point for the differential diagnosis. First, urine is examined grossly, encompassing both the solvent and the solutes. The color, odor, and turbidity of the urine offer the first clues. Dark, strong-smelling urine may be an indicator of decreased renal function. Cloudy, pungent urine generally indicates an infectious process, with the turbidity being a result of leukocytes in the urine. Dipstick tests and microscopic analysis provide a great deal of additional information. Microscopic examination entails the assessment of the urine sediment, the portion that remains after the urine specimen is centrifuged. Kidney disorders and the associated abnormalities identified by urine dipstick testing are shown in Table 27.1. Abnormal microscopic urinalysis results indicative of kidney disorders are shown in Table 27.2.

Other Diagnostic Tests

Many of the diagnostic tests presented in Chapter 26 are applicable to intrarenal disorders. The simple kidney, ureter, and bladder radiograph identifies gross abnormalities of the kidney related to position, size, and shape, as well as renal calculi that are radiopaque. Renal vasculature can be examined by renogram or renal scan; renal scans will also identify neoplasms in the kidney. Ultrasonography differentiates the solid mass of a neoplasm from fluid-filled cysts. Computerized tomography (CT) and magnetic resonance imaging (MRI) provide detailed information regarding multiple pathologies, including thrombi or other vascular occlusions, masses, and obstructions involving the kidney. In situations

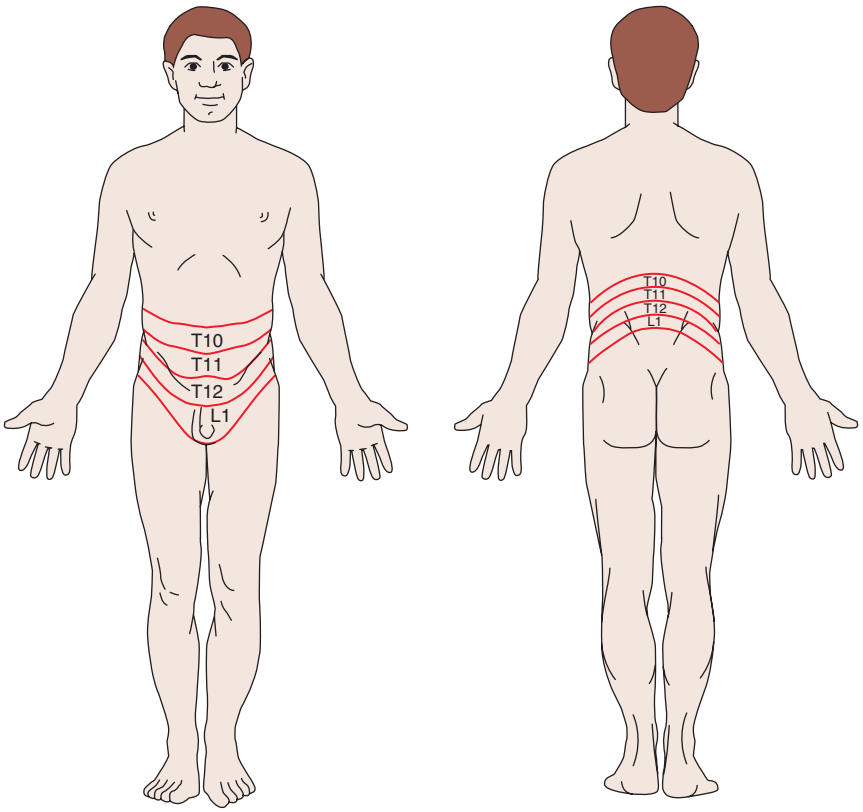


FIG 27.1 Dermatomes T10 (thoracic) to L1 (lumbar) correspond to areas that innervate the renal structures.

TABLE 27.1 Urine Dipstick Findings Associated With Kidney Disorders

Urine Dipstick Finding	Associated Kidney Disorders
Specific Gravity	
Decreased	Chronic kidney disease (decreased concentrating ability)
	Diabetes insipidus
Increased	Diabetes mellitus
	Syndrome of inappropriate secretion of antidiuretic hormone
pH	
Increased (6.5–8.0)	>7.5 Urinary tract infection with urea-splitting bacteria (e.g., *Proteus*)
	Renal tubular acidosis
	Calcium or struvite calculi
Decreased (4.5–5.5)	Uric acid or cystine calculi
Hematuria	Renal cell carcinoma
	Tubulointerstitial renal disease
	Urinary tract infection
	Trauma
	Glomerulonephritis
	Obstructive processes (e.g., calculi, neoplasms)
Proteinuria	Diabetic nephropathy
	Renal hypertension
	Glomerulopathies
	Nephrotic syndrome
	Renal arterial or venous obstruction
Glycosuria and Ketones	Diabetes mellitus
	Protein-energy malnutrition
White Blood Cells/Leukocyte Esterase	Urinary tract infection
Nitrites	Urinary tract infection (especially with gram-negative bacteria)

TABLE 27.2 Microscopic Urinalysis Findings and Associated Kidney Disorders

Microscopic Urinalysis Finding	Associated Kidney Disorders
Cells	
RBCs (circular shaped)	Urinary tract infection
	Trauma
	Obstructive disorders (e.g., calculi, neoplasm)
RBCs (irregularly shaped)	Glomerulopathies
WBCs (>1–2/HPF)	Infection or inflammation (e.g., urinary tract infection, neoplasm, calculi)
Irregular transitional epithelial cells	Malignancy
Casts	
Hyaline	Usually no pathologic significance
	Chronic kidney disease
	Pyelonephritis
Fatty	Nephrotic syndrome
	Nephrosis
RBC	Glomerulonephritis
WBC	Acute glomerulonephritis
	Acute pyelonephritis (sometimes)
	Acute tubulointerstitial nephritis
Waxy	Chronic kidney disease
	Diabetic nephropathy
	Glomerulonephritis
Crystals	
Cystine	Cystinuria
	Acidic urine
Uric acid	Uric acid calculi
	Acidic urine
Calcium oxalate	Calcium calculi
	Acidic urine
Calcium phosphate	Calcium calculi
	Alkaline urine
Triple phosphate/struvite	Struvite calculi
	Alkaline urine
Bacteria ≥5/HPF	Bacterial urinary tract infection
Yeast	Yeast urinary tract infection
Parasites (e.g., *Trichomonas vaginalis*)	Vaginitis in women
	Urethritis in men

HPF, High-power field; *RBC,* red blood cell; *WBC,* white blood cell.

in which these diagnostic tests are insufficient and actual tissue examination is necessary (e.g., neoplasm assessment), a renal biopsy may be required.

KEY POINTS

- Renal pain is generally perceived at the costovertebral angle (CVA). Pain is transmitted to the spinal cord between T10 and L1 by sympathetic afferent neurons. Pain may be felt throughout the dermatomes corresponding to T10-L1. Renal pain is usually due to distention and inflammation of the renal capsule and has a dull, constant character.
- Urinalysis provides a foundation for the differential diagnosis of renal dysfunction. Dipstick and microscopic urinalysis results provide clues to intrarenal pathologies.
- Other diagnostic tests provide information related to abnormal kidney anatomy and function.

CONGENITAL ABNORMALITIES

A wide variety of anomalies in the development of the kidneys have been documented in the literature. One or both kidneys may be involved. In some cases, findings are associated with other abnormalities, whereas in others the renal abnormalities are isolated. About 10% of live births have a significant urinary system malformation. Abnormalities may be identified prenatally during ultrasound assessment of the developing fetus, noted at birth, or manifest only in later life. Common congenital anomalies of the kidney include renal agenesis and cystic disease. Congenital disorders of the urinary collecting system are discussed in Chapter 29.

Renal Agenesis and Hypoplasia

Renal agenesis means a failure of one or both kidneys to embryonically develop. On the other hand, renal hypoplasia describes a condition in which some fetal development of the kidneys has occurred but they are smaller than normal.

Bilateral renal agenesis results from failure of the metanephros (renal buds) to develop in the fetus. It is incompatible with extrauterine life and results in stillbirth or death shortly thereafter. Unilateral renal agenesis (URA) is a rare disorder that is compatible with live birth. URA is often associated with concurrent urologic or other congenital anomalies. Nonurologic anomalies are usually cardiac or gastrointestinal in nature. Sometimes renal agenesis has been found to be familial and inherited as a dominant trait; screening by ultrasound of parents and siblings has been recommended when infants with agenesis or dysgenesis are diagnosed. Poorly controlled diabetes and exposure to certain drugs (e.g., those affecting angiotensin II) and chemicals have been implicated as teratogens. In URA, the remaining kidney usually enlarges as a compensatory mechanism. Lifelong monitoring of renal function is recommended.

Congenital renal hypoplasia increases the risk of developing chronic kidney disease. Gene mutation is likely responsible for the incomplete development of the kidney. Hypoplasia may be insufficient for extrauterine life if both kidneys are involved or may not affect renal functioning until later in life. When renal hypoplasia has been identified, regular monitoring of kidney function is recommended.

Cystic Kidney Diseases

Cystic disease of the kidneys incorporates a wide range of hereditary, developmental, and acquired conditions. Depending on the classification, these fluid-filled dilations may be present at birth or only visible later in life. They may involve one or both kidneys and be accompanied by other anomalies, or they may be the only pathology present. More commonly found in men, and increasing in prevalence with aging, renal cysts have been reportedly identified in more than half of patients over the age of 50. Cysts may be found in other organs or limited to the kidneys, depending on the disorder. Within the kidney, cysts may be diffuse or confined to one anatomic area. Renal cysts of significant size may produce flank pain and hemorrhage.

The two most common forms of cystic kidney disease are the autosomal-recessive and autosomal-dominant polycystic diseases. Autosomal-recessive polycystic kidney disease (ARPKD) is usually diagnosed in infants and young children, whereas autosomal-dominant polycystic kidney disease (ADPKD) may not be apparent until adulthood. Different genes are involved in ARPKD and ADPKD; however, both types of genes code for proteins associated with the primary cilium on the surface of tubule cells. The primary cilium is a mechanoreceptor that senses urine flow and is linked to regulatory responses, including a Ca^{2+} channel. The exact mechanisms are not known, but mutations of these proteins appear to cause growth dysregulation and abnormal remodeling of the tissue. Although the ARPKD and ADPKD forms of the disease are genetically different, their pathophysiology is similar. The onset and clinical courses are distinct (Table 27.3).

Autosomal-Recessive Polycystic Kidney Disease

ARPKD is often identified in the neonatal period, and when accompanied by pulmonary hypoplasia, it may result in death. In ARPKD, the kidneys retain their shape but are uniformly enlarged, and collecting ducts are dilated from the medulla to the cortex. The liver is commonly cystic as well as the kidneys. The most common clinical signs in the neonatal period are respiratory distress or palpable kidneys on physical examination. Systemic hypertension is frequently severe. Those who live to reach adulthood typically retain some renal function but experience a progressive decline in liver function. A recessive pattern of inheritance within the family and pathology on liver biopsy are important diagnostic data because both the clinical presentation and the results of ultrasounds, CT, or MRI can be similar to those of the autosomal-dominant form of polycystic kidney disease.

TABLE 27.3 Comparison of Autosomal-Recessive and Autosomal-Dominant Polycystic Kidney Disease

Feature	Autosomal Recessive	Autosomal Dominant
Gene defect	Chromosome 6p	Gene *PKD1* on chromosome 16, or gene *PKD2* on chromosome 4
Incidence	1:20,000	1–2:1000
Age at diagnosis	Usually neonate to childhood	Usually fourth to fifth decade
Imaging findings	Symmetrically enlarged kidneys	Enlarged kidneys, often asymmetric
Histologic findings	Cysts derived from epithelial cells of collecting ducts	Entire nephron involved
Liver involvement	Abnormal portal ducts progressing to fibrosis	Multiple cysts
Other systemic findings	Usually none	Cysts in other abdominal organs, aneurysms, abnormal cardiac valves, hernias, and diverticuli

Autosomal-Dominant Polycystic Kidney Disease

ADPKD is the most common of all the hereditary cystic kidney diseases. Typically, both kidneys are involved, but they may not progress at the same rate. ADPKD progresses at a variable rate and eventually results in end-stage kidney disease. The disease process appears to advance more rapidly in men than in women. Although prenatal and neonatal cases have been reported and ADPKD can present at any age, it usually manifests in patients who are 40 to 59 years old.

There are two genetically distinct but phenotypically similar forms of ADPKD: *PKD1* and *PKD2*. Distribution of these genotypes is such that 85% of the cases are associated with *PDK1* and just less than 15% with *PDK2*. Approximately 10% of the cases of ADPKD are not familial in origin and appear sporadically. Because of significant variability in clinical presentation within families, environmental modifying factors have been suggested.

While the cysts multiply and expand, the overall size of the kidneys increases and there is a progressive decline in glomerular filtration rate (GFR). A comparison of normal and polycystic kidneys is shown in Fig. 27.2. Expansion of the cysts compacts and distorts the vascular system, and the resulting local ischemia activates the intrarenal renin–angiotensin system. The progressive reduction in renal function is associated with an increase in the size of the kidney and the overall volume of the cysts.

Other tissues are affected in ADPKD, with the liver being the most common. Additional extrarenal sites include the spleen, pancreas, lung, seminal vesicles, circle of Willis, skin, and heart.

Involvement of these other organs and tissues results in additional clinical manifestations. In the early phases of the disease, the ability to concentrate urine is decreased. Hypertension is often diagnosed late in the disease process and increases the likelihood of escalated loss of renal function, proteinuria, and hematuria. In many adult patients with ADPKD, pain is a frequent complaint. Pain may be due to bleeding within the kidney, movement of kidney stones, or the development of urinary tract infections. Stasis of urine because of cysts predisposes to the development of kidney stones and infections.

FIG 27.2 Comparison of normal and polycystic kidneys. (From Brundage DJ: *Renal disorders,* St Louis, MO, 1992, Mosby.)

Diagnosis is based on family history, genetic testing and imaging techniques. When there is no family history of ADPKD, a presumption of the disease is made if imaging either identifies cysts and bilaterally enlarged kidneys or shows cysts in the liver and both kidneys. Genetic testing is then performed to substantiate the diagnosis. Treatment of ADPKD is primarily supportive, emphasizing the control of blood pressure and the management of any associated pathologic conditions. Once ESRD is reached, dialysis or kidney transplantation is required.

KEY POINTS

- Renal agenesis is relatively rare, and its presence is often associated with other congenital malformations. Bilateral renal agenesis is not compatible with life. Unilateral renal agenesis results in compensatory hypertrophy of the functional kidney. A single normal kidney is sufficient to maintain relatively normal renal function.
- Polycystic kidney diseases are genetically transmitted kidney disorders. Autosomal-recessive forms are evident at birth. In the autosomal-dominant type, symptoms generally occur later in life. Expanding cysts disrupt urine formation and flow. The inevitability of renal failure necessitates dialysis or transplantation.

NEOPLASMS

Neoplasms found in the kidney may be benign or malignant primary tumors, or they may result from metastases from extrarenal sites. Because the kidney is encased in a tough, fibrous capsule, growing renal neoplasms will distort the architecture of the kidney and ultimately hinder kidney function. Malignant renal neoplasms also carry the threat of metastasis to distant sites.

Benign Renal Neoplasms

Several benign neoplasms may be found in the kidney, developing from the renal cortex, medulla, or capsule. The etiology and pathogenesis of some of these benign neoplasms are summarized in Table 27.4.

Diagnosis and treatment. Benign renal neoplasms are typically detected incidentally during abdominal imaging for other reasons. The neoplasm may attain sufficient size to be detected with abdominal palpation, produce flank pain, and cause hematuria. Generally, benign renal neoplasms are treated by removal of some or all of the kidney (nephrectomy) because they are space-occupying lesions and because of their propensity to undergo malignant changes.

TABLE 27.4 Summary of Benign Renal Neoplasms

Benign Neoplasm	Etiology and Pathogenesis
Renal cortical adenomas	Small, solid growths that develop from cortical tissue; incidences of less than 1% up to 23%; typically <1 cm; patient is asymptomatic
Metanephric adenoma	Histologically related to Wilms tumor
Oncocytoma	Most difficult to differentiate from renal cell carcinoma; light brown to tan in color with well-defined borders
Angiomyolipoma	Composed of adipose, smooth muscle tissue, and blood vessels; believed to be hormone dependent because it is primarily found in postpubescent women
Nephroma	Cystic neoplasm not reliably differentiated from renal cell carcinoma in adults or nephroblastoma in children
Mixed epithelial stromal neoplasm	Most often found in perimenopausal women, most of whom are taking estrogen replacement therapy
Leiomyomas	Evolve from renal capsule, pelvis, or renal vein
Hemangiomas, fibromas, lipomas, lymphangiomas, and reninomas	Rare benign neoplasms

From Campbell SC, Novick AC, Bukowski RM: Neoplasms of the upper urinary tract. In Wein JA et al, editors: *Campbell-Walsh urology,* ed 11, Philadelphia, 2016, Saunders, pp 1575–1582.

Renal Cell Carcinoma

The great majority of kidney cancers are renal cell carcinomas (RCC). The remaining 10% to 15% of cases are primarily urothelial cell cancers of the pelvis. A cross-sectional view of RCC is shown in Fig. 27.3. Approximately 65,000 cases of renal cell and renal pelvis cancers are diagnosed annually in the United States, representing 3% of all cancers. Rates are higher in men, African Americans, and older adults.

Etiology and pathogenesis. Although only a small portion of the cases of RCC are attributed to known genetic factors, the risk of its development is higher when there is a first-degree relative with the disease. Several specific genes have been identified in familial cases and are associated with specific subtypes, including clear cell carcinoma (*VHL* gene) and papillary carcinoma (*MET* gene).

Several risk factors for RCC have been identified, and these are listed in Box 27.1. Obesity, cigarette smoking, and hypertension are important modifiable risk factors. Estrogen therapy and exposure to asbestos or heavy metals also increases risk.

Clear cell carcinomas originate in the renal cortex from cells of the proximal tubule and are usually unilateral and random in occurrence, with some cases of familial patterns of inheritance. Clear cell RCC frequently is associated with metastatic disease. Activity of the tumor suppressor gene protein *VHL* is lost, allowing excessive growth stimulation. Papillary RCC, representing 10% to 15% of RCC diagnoses, has been associated with excessive activity of the proto-oncogene *MET* that

codes for a growth factor receptor. (See Chapter 7 for a discussion of protooncogenes and tumor suppressor genes.)

Clinical manifestations. RCC is commonly asymptomatic until it is quite advanced. When it does present with signs and symptoms, the most common are CVA tenderness, hematuria, and a palpable abdominal mass. Dyspnea, cough, and bone pain develop secondary to metastasis.

Diagnosis and treatment. In many cases RCC has metastasized at the time of initial diagnosis. The staging system used with RCC is shown in Fig. 27.4. Radical nephrectomy is indicated for advanced-stage cancer; early-stage cancer may be managed with partial nephrectomy More targeted therapies, specifically angiogenic inhibitors and tyrosine kinase inhibitors, may be used in selected patient situations. The overall 5-year survival for RCC is approximately 70%.

Nephroblastoma (Wilms Tumor)

Nephroblastoma is the fifth most frequently occurring pediatric malignancy and the most common childhood kidney cancer. Seventy-five percent of the cases occur in children less than 5 years old. Incidence is equally distributed between genders, and it is usually unilateral and sporadic; only 1% to 2% of patients diagnosed with Wilms tumor have a relative who has been diagnosed with the disease.

Etiology and pathogenesis. Nephroblastomas develop from embryonic pluripotent kidney precursor cells called *nephrogenic rests.* The first gene found to be associated with Wilms tumor was *WT1*, located on the short arm of chromosome 11. A second gene called *WT2* has also been implicated.

Nephroblastomas are typically large, well-encapsulated tumors that grow rapidly. They undergo hemorrhage and cystic changes; necrosis that occurs because of hemorrhaging is responsible for the development of cysts. As the malignant cells proliferate, the normal architecture of the kidney is altered. The renal pelvis becomes compressed, and locally the tumor infiltrates into the renal veins and hilar nodes. Metastasis is typically via the bloodstream to the brain, liver, adrenal glands, and bone.

Clinical manifestations. A palpable abdominal mass is evident in about 80% of cases. Abdominal pain, hypertension, and hematuria are also common local manifestations. Nephroblastoma may produce a tumor thrombus in the inferior vena cava, which can lead to decreased venous return and lower extremity edema. Additional signs or symptoms may be noted reflective of the sites of metastasis.

Diagnosis and treatment. Most often the tumor is identified by the parents or during a routine physical examination. Renal ultrasound and/or CT scan will usually identify anatomic abnormalities associated with the tumor. Treatment protocols have improved to the point that this previously lethal disease now has a cure rate of more than 90%. Key to treatment is removal of the involved kidney (nephrectomy). Radiation and chemotherapy are routine after surgery. Radiation therapy may also be administered preoperatively to shrink tumors and reduce their vascularity.

BOX 27.1 Risk Factors for Renal Cell Carcinoma

- Cigarette smoking
- Obesity
- Hypertension
- Diabetes mellitus
- Asbestos exposure
- Multiparous women
- Long-term renal dialysis
- Kidney transplantation

KEY POINTS

- A number of benign and malignant primary neoplasms may develop in the kidney. Symptoms depend on the size of the neoplasm and the presence of metastasis. Neoplasms may be asymptomatic until quite large. At that time, a palpable abdominal mass, hematuria, and flank pain may be noted.
- Neoplasms are usually detected with renal ultrasound and/or CT scan. Nephrectomy remains the initial treatment of choice for both benign and

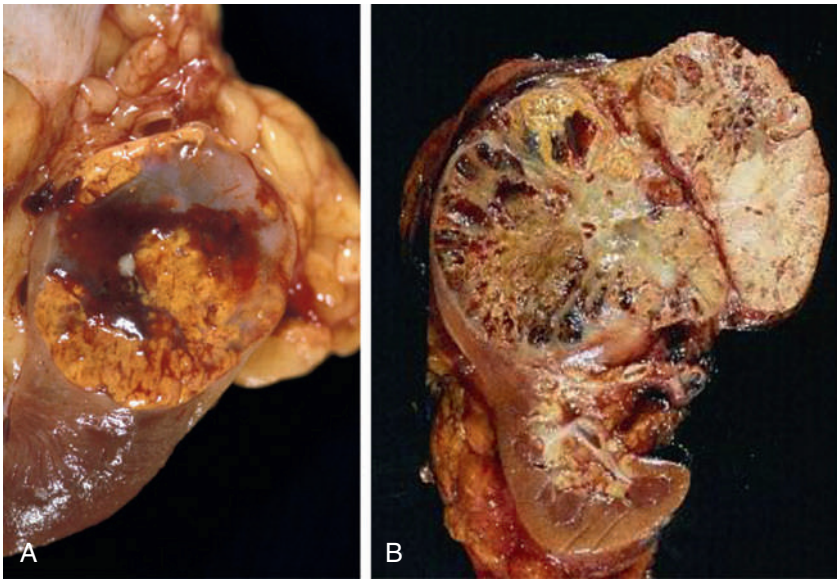


FIG 27.3 Clear cell renal cell carcinoma (CCRCC). The gross appearance of a CCRCC is characteristically golden yellow, and it may be associated with variable amounts of hemorrhage and necrosis. Tumors arise from cortical tubules and can be **(A)** confined to the renal parenchyma or **(B)** extend beyond the capsule into perinephric soft tissue (From Hirsch MS, Signoretti S, Dal Cin P: Adult renal carcinoma. *Surg Pathol Clin* 2015;8(4):587–621.)

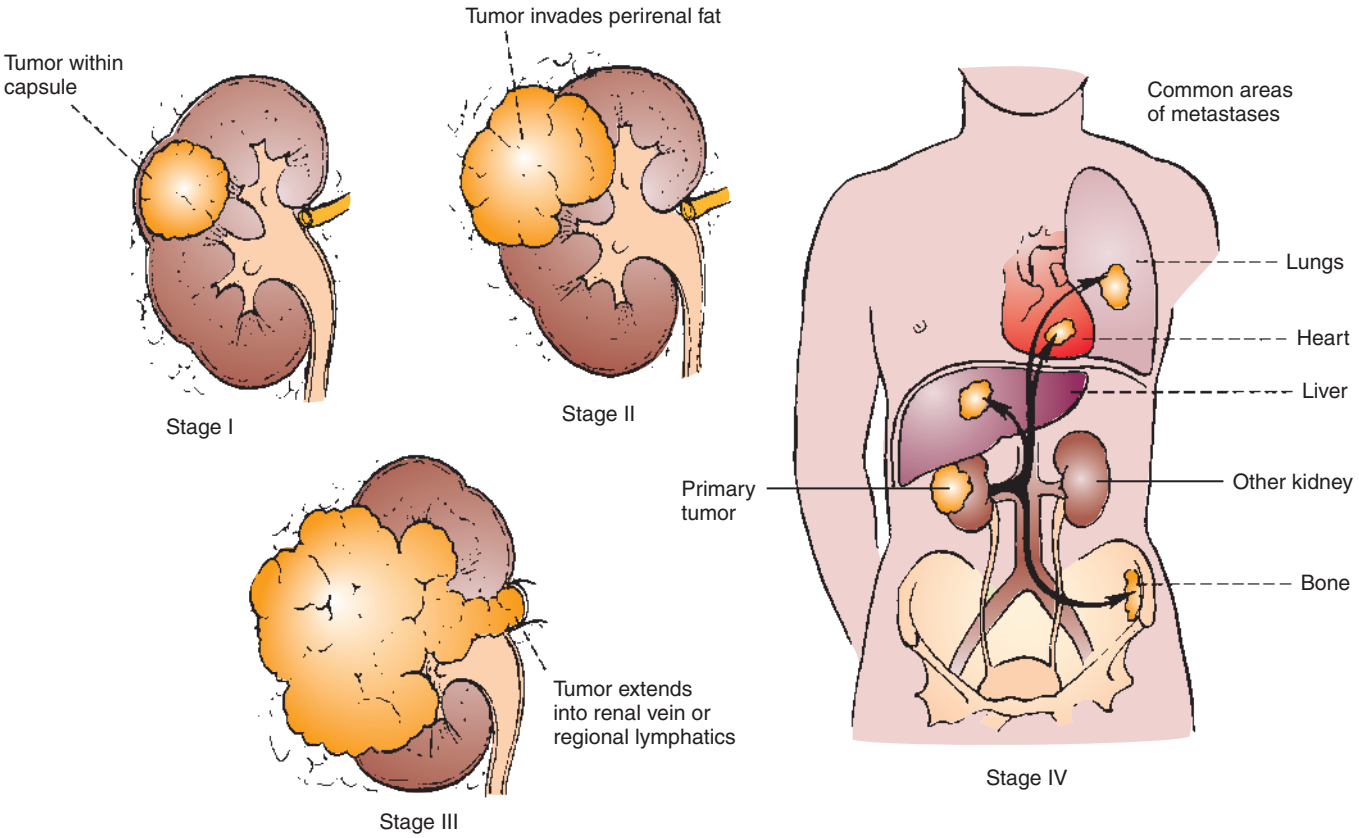


FIG 27.4 Staging system for renal cell carcinoma. (From Black JM, Matassarin-Jacobs E: *Medical-surgical nursing: clinical management for continuity of care,* ed 8, Philadelphia, 2009, Saunders, p 923.)

malignant renal neoplasms. Nephron-sparing partial nephrectomy is utilized when appropriate.

- Renal cell carcinomas (RCC) is often metastasized at the time of diagnosis because it remains asymptomatic. Two gene defects have been associated with the development of RCC *(VHL, MET)*. The overall survival rate of RCC is approximately 70%, higher at early stages and lower at late stages.
- Nephroblastoma, or Wilms tumor, is the most common kidney cancer in children. Nephrectomy, radiation therapy, and chemotherapy are used in the management of nephroblastomas. Cure rates are high.

INFECTION

Normally, a number of host defense mechanisms serve to protect the renal system from infection. Chemically, the acidic pH and the presence of urea in the urine produce a relatively hostile environment for bacterial growth. Bacteriostatic prostatic secretions in men also act as a protective mechanism against bacterial invasion. In women, glands in the distal urethra secrete mucus that captures bacteria, preventing progression to the bladder. Small numbers of bacteria that may enter the system are washed out by micturition. Normal unidirectional flow prevents reflux of urine from the bladder to the kidney via the ureter by contraction of the vesicoureteral junction that occurs with bladder filling. In children, urinary tract infections may be associated with vesicoureteral reflux or other anatomic malformations of the urinary tract (see discussion of vesicoureteral reflux in Chapter 29). Epithelial cells of the urinary tract provide a physical barrier to infectious organisms, and indigenous proteins trap bacteria or block their adhesion to epithelial cells.

Infection of the kidney is known as *pyelonephritis,* or upper urinary tract infection, and affects the renal tubules, pelvis, and calices. Although infectious organisms responsible for upper urinary tract infections may be delivered to the kidney via the bloodstream or lymphatic system, they most commonly reach the kidneys as an ascending infection from the lower urinary tract (urethra, bladder, and ureters) (Fig. 27.5). Infection of the lower urinary tract is discussed in Chapter 29.

The most common causative agents of renal system infection are serogroups of *Escherichia coli,* whereas *Enterobacter, Enterococcus, Proteus mirabilis,* and strains of *Klebsiella* are responsible for most of the other infections. Most often, infections with these other organisms are linked to calculi or to anatomic abnormalities of the urinary tract. In certain populations, *Staphylococcus saprophyticus* and group B *Streptococcus* have been identified. Fungal infections, mycoplasmas, and other anaerobic bacteria are occasionally responsible. Infections attributable to *Neisseria gonorrheae* or *Chlamydia trachomatis* are associated with sexually transmitted diseases and are typically limited to the urethra. Uropathogenic bacteria possess adhesins that allow the bacteria to bind to epithelial cells of the urinary tract.

Risk factors for urinary tract infections are listed in Box 27.2. One of the most significant preventive interventions is the early removal of urinary catheters. Anatomically, women are at increased risk because of their shortened urethras, but the resulting infections are most often confined to the lower urinary tract. Urinary tract infections are common in women and elderly men. Children with congenital anomalies that allow reflux are also at high risk. Male children who are uncircumcised experience urinary tract infections more frequently than those who are circumcised.

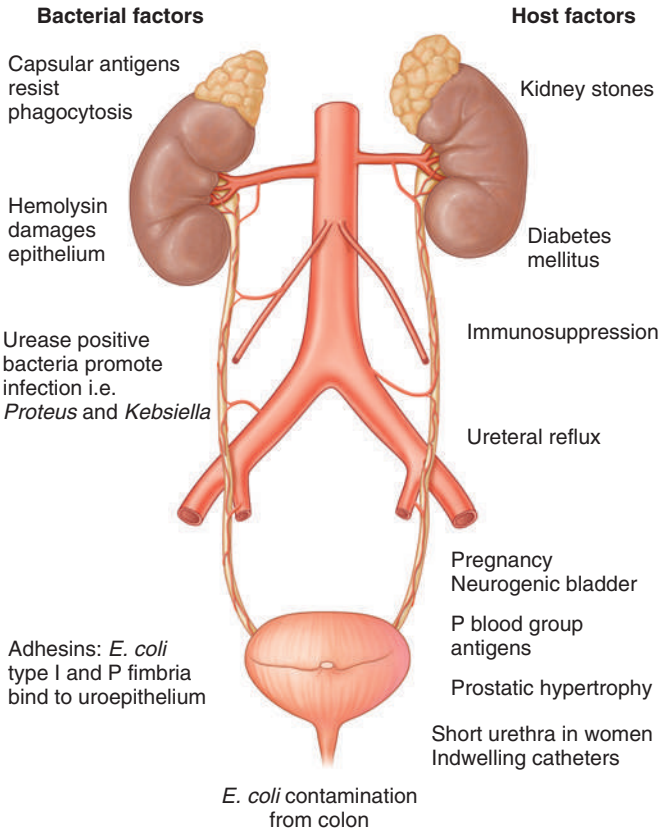


FIG 27.5 Mechanisms of urinary tract infection. (From Huether S, McCance K: *Understanding pathophysiology,* ed 6, St Louis, MO, 2016, Mosby.)

BOX 27.2 Risk Factors for Urinary Tract Infections

- Increasing age
- Vesicoureteral reflux
- Congenital anatomic anomalies of the urinary tract
- Female gender
- Pregnancy
- Diaphragms with spermicidal agents for contraception
- Neurogenic bladder
- Instrumentation of the urinary tract (catheterization, cystoscopy)
- Urinary obstruction (calculi, benign prostatic hyperplasia)
- Glucocorticoids, radiation, or cytotoxic chemotherapy
- Immunodeficiency conditions such as AIDS
- Diabetes mellitus
- Obesity
- Sickle cell trait

Acute Pyelonephritis

Etiology and pathogenesis. Although bladder infections are quite common, especially in women, pyelonephritis is unlikely to occur unless there is obstruction or ureteral reflux that allows contaminated urine to enter the kidney. Infection of the kidney from the bloodstream is rare in comparison to infections that ascend from the lower urinary tract. Pregnancy is a risk factor for acute pyelonephritis in women because of the physiologic alterations that occur in the urinary tract. In nonpregnant women, men, and children, the most common risk factors for acute pyelonephritis are diabetes mellitus, anatomic abnormalities of the urinary tract (especially vesicoureteral reflux in children), and obstructive causes. Populations with the highest incidence are young women, infants, and the elderly. *E. coli* is the causative organism in the large majority of acute pyelonephritis cases. Pyelonephritis may be unilateral or bilateral. Acute infection of the kidney usually originates as an ascending infection, but may arrive at the kidney via the bloodstream. Once in the kidney, bacteria bind to tubule epithelial cells, initiating an inflammatory response. Inflammatory mediators and bacterial toxins are responsible for the damage to the kidney tubules.

Clinical manifestations. The onset of acute pyelonephritis is sudden; patients are usually acutely ill and present with fever, chills, and CVA tenderness, as well as symptoms of lower urinary tract infection (dysuria, urgency, and frequency). Nausea, vomiting, and anorexia are frequent accompaniments, increasing the fever-induced dehydration.

Complications of acute pyelonephritis include abscesses, sepsis, acute respiratory distress syndrome, recurrent/chronic pyelonephritis, and chronic kidney disease attributable to scarring produced by recurrent infections. Renal scarring and the subsequent development of chronic kidney disease are more likely when there are preexisting anatomic or functional urinary tract abnormalities. Chronic kidney disease has been found to progress more rapidly after acute pyelonephritis. Preexisting chronic kidney disease can also increase the severity of an infection. Overall, however, the risk of decreased renal function subsequent to acute pyelonephritis is low, occurring in only 3% to 4% of cases. The term *urosepsis* describes organisms in the bloodstream originating from a urinary tract infection. Urinary tract infections are the cause of 20% to 30% of all sepsis cases. Fig. 27.6 illustrates acute pyelonephritis complicated by abscess formation.

Diagnosis and treatment. In addition to the clinical manifestations, diagnosis of acute pyelonephritis is suggested by urinalysis results illustrating the presence of significant amounts of bacteria and white blood cell (WBC) casts. WBC casts differentiate pyelonephritis from bladder infections because they are formed in the kidney tubules and not in the lower urinary tract. Infections commonly are treated with antimicrobial therapy on an outpatient basis for 7 to 10 days. Further assessment for urinary obstruction or an extraurinary focus of infection may be appropriate in some situations. In cases of complicated acute pyelonephritis, such as septicemia, patients are hospitalized and treated with intravenous antimicrobials and fluids.

Chronic Pyelonephritis

Chronic pyelonephritis is characterized by small atrophied kidneys with diffuse scarring and blunting of the calices secondary to persistent or recurrent infection of the kidney.

Etiology and pathogenesis. Chronic reflux of infected urine into the renal pelvis is the typical cause of chronic pyelonephritis. The kidneys are usually smaller than normal with caliceal deformity, chronic inflammation, and parenchymal scarring. Chronic pyelonephritis causes about 2% to 3% of the cases of ESRD. Individuals at risk for developing chronic pyelonephritis have bacteriuria associated with obstructive disorders such as renal calculi, neurogenic bladder, vesicoureteral reflux, or underlying intrarenal disease. Chronic or recurrent pyelonephritis is one potential cause of chronic kidney disease (see Chapter 28).

Clinical manifestations. The symptoms of chronic pyelonephritis may be vague, inconsistent, or similar to those of acute pyelonephritis. Patients may have flank or abdominal pain, fever, malaise or anorexia.

Diagnosis and treatment. Urinalysis results typically parallel the findings of acute pyelonephritis, but may not be as profound. Diagnostic testing includes renal ultrasound and other imaging tests that show one or both kidneys to be smaller than normal with distorted architecture and significant scarring. Renal tubules may be dilated or atrophied. Other diagnostic tests may be performed to determine underlying pathologies such as vesicoureteral reflux or obstruction caused by renal

FIG 27.6 Acute pyelonephritis with abscess formation. (From Klatt E: *Robbins and Cotran atlas of pathology,* ed 3, Philadelphia, 2015, Elsevier, Saunders.)

calculi. Treatment is based on the correction of these underlying causes, administration of antimicrobial therapy that may continue for several months, and support of compromised renal function, if it exists.

KEY POINTS

- Pyelonephritis is an infection of the renal pelvis and kidney tubules that is usually due to an ascending urinary tract infection. Costovertebral angle (CVA) tenderness is the classic symptom. It is frequently accompanied by fever, chills, nausea, vomiting, and anorexia. Urinalysis generally shows evidence of an infective process. The presence of WBC casts is specifically indicative of an upper urinary tract infection as opposed to a lower urinary tract infection. When managed promptly and effectively, acute pyelonephritis does not generally result in decreased renal function.
- Chronic pyelonephritis can potentially result in chronic kidney disease. It is usually associated with vesicoureteral reflux or obstructive processes leading to persistent urine stasis. Ongoing inflammation causes fibrosis and scarring and loss of functional nephrons. The diagnosis is confirmed by renal imaging. Urinalysis results mirror those of acute pyelonephritis, but may not be as significant. Treatment includes correction of the underlying processes and often extended antimicrobial therapy.

OBSTRUCTION

Obstructive disorders of the urinary tract interfere with the flow of urine. Obstruction can occur at any point in the system from the renal pelvis to the urethral meatus (Fig. 27.7). In general, it causes dilation of the tract proximal to the obstruction. Stasis of urine occurs and predisposes to urinary tract infection and structural damage. Prolonged obstruction results in postrenal acute kidney injury and can ultimately lead to acute tubular necrosis (intrarenal acute kidney injury) (Chapter 28).

Disorders resulting in urinary tract obstruction may be congenital or acquired. In children, urinary tract obstruction is usually due to anatomic abnormalities such as with ureteral valves, strictures of the urethral meatus, and stenosis at the ureterovesical or ureteropelvic junction. Obstruction in adults predominantly occurs as a result of acquired disorders and may be either intraluminal (e.g., renal calculi) or secondary to extrinsic compression (e.g., tumors, prostatic hyperplasia). Acquired obstructions are due to anatomic anomalies of the urinary tract (Table 27.5). Complications of urinary obstruction include infection, sepsis, acute kidney injury, and, potentially, chronic kidney disease.

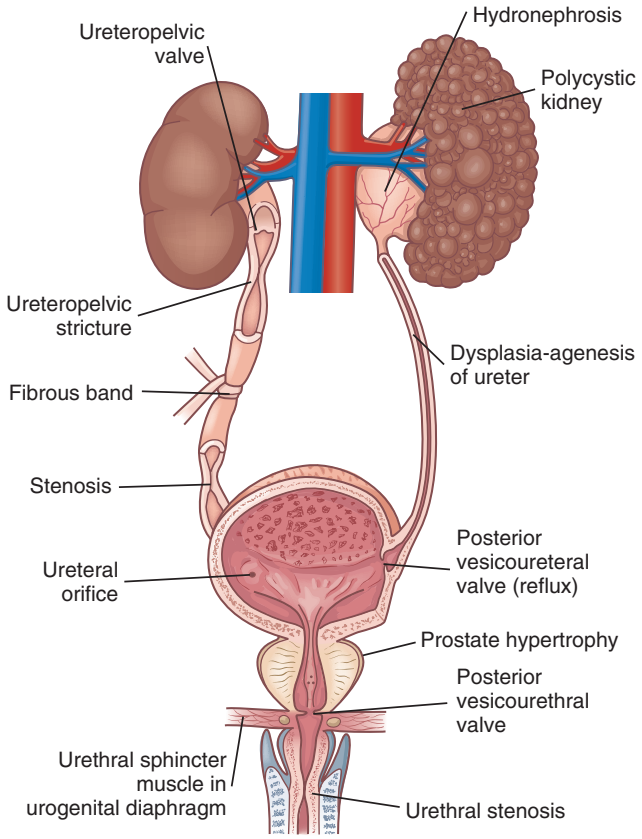


FIG 27.7 Major sites of urinary tract obstruction. (From Huether S, McCance K: *Understanding pathophysiology,* ed 6, St Louis, MO, 2016, Mosby.)

Changes that occur within the urinary tract as a result of obstruction are dependent on (1) location and degree of obstruction (i.e., partial or complete, unilateral or bilateral) and (2) the duration and timing (acute onset or chronic) of the obstruction. Initially, in complete or significant partial obstruction, hydrostatic pressure increases proximal to the obstruction as a consequence of continued glomerular filtration

TABLE 27.5 Causes of Renal System Obstruction

Type of Obstruction	Cause
Intraluminal	Calculi, clot
	Tumor: bladder, urethra, kidney
	Papillary necrosis
Extrinsic	Prostatic hypertrophy
	Retroperitoneal fibrosis
	Tumor: pelvic, retroperitoneal
Acquired	Neurogenic bladder
	Ureteral stricture
	Urethral stricture

and simultaneous obstruction to the flow of urine. Other structures proximal to the obstruction then begin to dilate. The more distal to the kidney the obstruction is located, the less dilation is seen because the pressure is distributed over a greater surface area. Complete obstruction of a ureter results in its dilation (hydroureter). The renal pelvis and tubules enlarge with the increased pressure, and the renal papillae flatten. The enlarged kidney is called *hydronephrosis.* GFR and renal perfusion decline, and eventually portions of the kidney become ischemic. Over the course of several weeks, if the obstruction is not corrected, tubular atrophy and destruction of the medulla result in scar tissue and nonfunctioning or poorly functioning glomeruli. Postrenal acute kidney injury is discussed in detail in Chapter 28.

Partial obstruction is much more common. In this situation, the renal pelvis may become very dilated but the structural or functional disruption of the kidney may be minimal. If the obstruction is bilateral, however, clinical manifestations of fluid retention will be present. Patients often complain of symptoms such as weight gain, nausea, anorexia, malaise, headaches, increased abdominal girth, and ankle edema. Functionally, partial obstruction can produce a slight to moderate decrease in blood flow and GFR and an inability to concentrate urine or secrete potassium and hydrogen ions. Compensatory hypertrophy occurs in the unaffected kidney. Renal calculi and benign prostatic hyperplasia (BPH) in men are common causes of urinary tract obstruction. BPH is discussed in Chapter 29.

Renal Calculi (Nephrolithiasis)

Renal calculi, or lithiasis, are crystal aggregates composed of organic and inorganic materials located within the urinary tract. These calculi can form within the kidney, in the tubules, or in the collecting system and may then migrate to more distal structures. Occasionally, terminology is used to identify where these calculi are found (e.g., *ureterolithiasis,* located in the ureters), but more commonly, the terms *renal calculi* and *nephrolithiasis* are used to describe the formation and passage of calculi anywhere within the urinary tract. The Greek term *lithiasis* means *stone,* so renal calculi are also referred to as *kidney stones.*

Nephrolithiasis affects individuals throughout the world and has been traced back as far as Egyptian mummies. It is estimated that 10% to 15% of Americans will experience a kidney stone during their lifetimes. The risk of nephrolithiasis is significantly lower among African Americans and Mexican Americans than among Caucasians. A family history of kidney stones increases the risk of developing some types of nephrolithiasis. Reduced fluid intake, dehydration, increased urine concentration, and decreased urine volume increase the risk for stone formation. Occupations where there is increased exposure to high temperatures or sedentary levels of activity are also associated with an increased incidence of stone formation. Nephrolithiasis is uncommon before age 20, with a peak onset between 20 and 30 years.

BOX 27.3 General Contributing Factors to Nephrolithiasis

- Hyperparathyroidism
- Gout
- Certain medications
- Hypertension
- Urinary tract infections
- Chronic inflammatory bowel disease; irritable bowel disease; chronic diarrhea
- Excess dietary meat
- Excess dietary sodium
- Excess dietary oxalate
- Past medical history or family history of nephrolithiasis
- Obesity
- Insulin resistance/type 2 diabetes mellitus
- Dehydration
- Prolonged immobility
- Congenital kidney defects/anatomic alterations (e.g., abnormal kidney shape)
- Vesicoureteral reflux

Etiology and pathogenesis. Crystallization and stone formation occur with several solutes found in urine and may be promoted or inhibited by a number of factors. Some of these factors are inherent to the urinary tract and the characteristics of urine, whereas others are unique to the individual and the type of renal calculi formed. Box 27.3 identifies general factors that contribute to the formation of renal calculi.

The pathogenesis of nephrolithiasis begins with urine becoming supersaturated with the specific solute. Urine is a solution of solvent (water) and solutes (particles). Certain of these solutes have a tendency to form crystals if their concentration within the urine becomes great enough; this is the meaning of *supersaturation.* These crystals usually begin their development in the concentrating areas of the nephron and kidney pelvis. Crystallization is enhanced when a person is dehydrated (decreased solvent) or has higher-than-normal levels of solute in the urine from excessive excretion (calcium, uric acid). Crystals are unable to aggregate into a calculus of sufficient size to obstruct the urinary tract if urine is moving freely through the system. Adequate fluid intake is therefore an inhibitor to stone formation. Stasis of urine flow because of obstruction, immobility, and a sedentary lifestyle facilitates calculus formation.

In addition to a crystalline component, nephrolithiasis consists of a noncrystalline, organic component termed the calculus *matrix.* The matrix itself consists of protein, sugar, glucosamine, bound water, and organic ash. It represents a small percentage of the overall weight of the calculus, but it is present in all types of stones.

The types of renal calculi are shown in Table 27.6, along with their relative incidence in adults and primary risk factors. Calcium oxalate–based stones are by far the most common in both children and adults. Calcium-based nephrolithiasis is primarily idiopathic in etiology (unknown cause), but a family history is common. Primary risk factors for calcium-based calculi are hypercalciuria and/or hypercalcemia, hyperoxaluria, and hyperuricosuria. Hypercalciuria and/or hypercalcemia is commonly due to increased gastrointestinal absorption, impaired renal tubular reabsorption, and primary hyperparathyroidism and less often to prolonged immobility, metastatic bone cancer, multiple myeloma, prolonged metabolic acidosis, hypocitraturia, and excessive amounts of vitamin D. Hyperoxaluria can increase the urinary saturation of calcium oxalate. It may be due to a genetic defect that increases urinary oxalate excretion, a defect in liver metabolism, or increased gastrointestinal absorption of an oxalate-rich diet (rhubarb, spinach, chocolate, nuts,

TABLE 27.6 Types of Nephrolithiasis, Relative Incidence, and Specific Risk Factors

Constituent of Stone	Incidence	Specific Risks
Calcium oxalate (primarily)	75%	Genetic predisposition; idiopathic; hypercalcemia and/or hypercalciuria (hyperparathyroidism, prolonged immobility, increased gastrointestinal absorption, impaired renal reabsorption); hyperuricosuria (see risks later); hyperoxaluria
Struvite	7%–10%	Urinary tract infection with urea-splitting bacteria (e.g., *Proteus, Klebsiella, Pseudomonas*)
Uric acid	7%–10%	Genetic predisposition; hyperuricosuria associated with gout, acute leukemia, glycogen storage disease, malignancy; excessive intake of meat, fish, poultry; obesity; type 2 diabetes; urine pH <5.5
Cystine	1%–3%	Genetic defect in renal cystine reabsorption
Triamterene, guaifenesin, indinavir, xanthine	All <1%	Secondary to specific medications

BOX 27.4 General Interventions for Nephrolithiasis

- Narcotic analgesics if needed for pain management
- Increased fluid intake (>2 L/day; oral and/or IV)
- Antimicrobials if needed for urinary tract infection
- Shockwave lithotripsy, ureteral stenting, ureteroscopy for removal of large stones unable to pass spontaneously
- Dietary modifications unique to stone composition
- Medications unique to stone composition (e.g., allopurinol for uric acid stones)

BOX 27.5 Dietary Modifications for Nephrolithiasis

- Increase water intake (minimum of 2 L/day).
- Ensure adequate dietary calcium intake.
- Avoid calcium supplementation.
- Avoid foods with additional vitamin D (and vitamin C, if recommended).
- Avoid calcium-based antacids.
- Limit intake of coffee, tea, or colas to ≤2 per day.
- Reduce protein intake from meat, fish, and poultry.
- Limit dietary sodium and oxalate intake.
- Avoid high-purine foods for uric acid stones (and, if recommended, for calcium).

beer, coffee, tea, cocoa, or excessive vitamin C). Hyperuricosuria facilitates precipitation of calcium oxalate stones or may produce uric acid stones. There is a genetic predisposition for hyperuricosuria, which may be found in association with gouty arthritis or secondary purine excess attributable to overproduction in myeloproliferative disorders, excess ingestion of purine-rich foods, or errors of purine metabolism. Foods high in purine include organ meats such as liver and kidney, sardines, anchovies, salmon, and foods high in yeast such as beer and bread.

Uric acid and struvite calculi are the second most frequently occurring stones after calcium oxalate stones. In addition to the risk factors previously identified for hyperuricosuria, an increased body mass index and glucose intolerance or type 2 diabetes is frequently associated with uric acid stones. The basis of struvite stones is not metabolic, but rather the changes that occur within the urinary tract with some infectious processes, and is often assisted by some underlying anatomic or functional abnormalities that facilitate urinary stasis. Struvite stones are composed of magnesium, ammonium, and phosphate. Urinary tract infections with certain bacterial species capable of splitting urea into two ammonium ions and one bicarbonate ion neutralize the urine and support bacterial proliferation as well as stone formation. Struvite stones form around a bacterial nucleus, producing an antimicrobial barrier. They may grow rapidly and assume a jagged formation known as *staghorn.* Other types of nephrolithiasis are less common and are associated with familial defects in renal transport or are produced as adverse effects of certain drugs (e.g., indinavir, triamterene, and xanthine).

Clinical manifestations. Signs and symptoms of renal calculi differ with their size and location and may mimic any number of other causes of abdominal pain. Pain may be vague or, more commonly, acute renal colic or flank pain. Stones within the kidney are responsible for flank pain that may be dull and localized. While the stone moves to the ureteropelvic junction and down into the ureter, spasmodic, intermittent sharp pain known as *renal colic* develops. The pain may radiate into the ipsilateral groin area, testicle, or labia. The pain induced by nephrolithiasis makes the patient acutely uncomfortable and is often accompanied by nausea and vomiting, diaphoresis, tachycardia, and tachypnea. When the stone reaches the bladder, there is often a noticeable reduction in the pain. If the stone results in a partial obstruction at the urethra, dysuria, urgency, and frequency are common findings. If stones injure the linings of the ureters or urethra as they pass, microscopic or macroscopic hematuria occurs. Prolonged or repeated nephrolithiasis may result in scarring of the kidney and, if bilateral, chronic kidney disease.

Diagnosis and treatment. In addition to the clinical manifestations, a thorough family history should be obtained, risk factors for nephrolithiasis should be assessed, and any anatomic or functional urinary tract anomalies should be identified. Urinalysis permits the identification of a concomitant infection, levels of specific stone-forming constituents (e.g., hypercalciuria), and urinary pH, which affects crystallization. Hematuria (gross or microscopic) may be present persistently or intermittently. If a stone has been passed and is available, it can be sent for analysis of its composition. Determination of a complete blood count; measurement of levels of serum electrolytes, serum creatinine, and blood urea nitrogen (BUN); and assessment of parathyroid hormone level provide information related to risk factors and renal function.

The traditional gold standard for the assessment of renal calculi has been intravenous pyelogram, but it has been effectively replaced by CT scans. Basic abdominal x-rays identify only those stones that are radiopaque. CT scanning can identify renal calculi, other potential sources of flank pain, and anatomic anomalies.

Treatment of nephrolithiasis may be medical or surgical and is determined by the size, position, and composition of the calculus; the presence or absence of urinary tract infection; and the involvement of one or both kidneys. Interventions are summarized in Box 27.4, and dietary changes are listed in Box 27.5. The goal is not only treatment of existing stones, but also prevention of their recurrence. Acute pain

management may be necessary. Opiates are commonly used because they reduce both renal colic and pain perception. With increased fluid intake, most smaller stones pass out of the urinary tract spontaneously. Stones less than 5 mm in diameter have a high chance of being passed, whereas those that are 5 to 7 mm in diameter have a 50% chance, and those >7 mm almost always require urologic intervention. Other interventions that may be required include extracorporeal shockwave lithotripsy, where the stone is broken into smaller pieces for passage, or percutaneous approaches to stone removal by endoscopy with basket retrieval or ultrasonic or laser lithotripsy. Open surgical approaches are used when no other intervention is successful, but are usually avoidable and therefore uncommon. Unfortunately, depending on the type of renal calculus, 30% to 50% of adults can expect a recurrence within 5 to 10 years. Dietary modifications can decrease this risk in some cases. Recommendations are designed for each individual patient, based upon the type of stone and specific risk factors.

KEY POINTS

- Obstructive processes result in urine stasis, which predisposes to infection and structural damage. Common causes of obstruction include stones, tumors, prostatic hyperplasia, and strictures of the ureters or urethra.
- Complete obstruction results in hydronephrosis, decreased glomerular filtration rate (GFR), and ischemic kidney damage because of increased intraluminal pressure. Prolonged postrenal acute kidney injury resulting from the obstruction may result in acute tubular necrosis (intrarenal acute kidney injury) and chronic kidney disease.
- Stones tend to form in the urinary tract under conditions of solute supersaturation, low urine volume, and abnormal urine pH. Certain factors increase the risk of stone formation, whereas others act as inhibitors.
- Most stones are composed of calcium crystals. Other forms include uric acid, struvite, cystine, and stones that form in association with certain medications.
- Stationary stones in the renal pelvis are generally asymptomatic. When the stone migrates to the junction with the ureters and beyond, intense renal colic pain ensues. Pain is usually abrupt in onset and may radiate. Additional signs and symptoms include nausea, vomiting, and diaphoresis.
- Most stones pass spontaneously, but other interventions may be necessary, including lithotripsy or endoscopic approaches. Stones tend to recur, and prevention is enhanced by a high fluid intake to dilute the urine and dietary changes based upon the type of stone.

GLOMERULAR DISORDERS (GLOMERULOPATHIES)

Glomerular disorders result from alterations in the structure and function of the glomerular capillary circulation and are broadly delineated as *glomerulopathies.* The membranes of glomerular capillaries have three layers: endothelium, basement membrane, and a layer of specialized epithelial cells with footlike projections (podocytes) that encircle the basement membrane (Fig. 27.8). Together, this triple layer comprises the filtration barrier of the glomerulus. The glomerular filtrate passes through gaps (slit pores) between these podocytes, enters the space in Bowman capsule, and progresses into the proximal tubule (see Chapter 26). Glomerulopathies are responsible for a great majority of cases of ESRD.

Pathologic changes to glomeruli may occur insidiously, altering function over the course of months or years, or they may have an acute onset, with rapidly developing impairment. In some cases, there are no apparent signs or symptoms, and glomerular dysfunction is identified serendipitously during routine urine examination or evaluation of some other health concern. The specific causes of glomerulopathies are often unknown. Hereditary and environmental factors are implicated; metabolic, infectious, hemodynamic, toxic, immune, genetic, and other mechanisms of injury are involved. Deposition of antibodies into the glomeruli, where they attach to local antigens, is the usual inciting event, followed by inflammation and immune injury. Deposition of immune complexes (antigen–antibody complexes) that are already formed in the bloodstream also occurs in cases of systemic immune disorders (Box 27.6).

The challenge in any discussion of glomerulopathies lies in the often-confusing range of clinical and pathologic approaches to their classification. One approach is to classify glomerular disorders according to primary and secondary etiologies. *Primary glomerulopathies* are disease states in which the kidney is the only or the predominant organ involved, and *secondary glomerulopathies* result from a variety of systemic disorders. Autoimmune and immune-complex diseases such as Goodpasture syndrome and systemic lupus erythematosus are examples of secondary glomerulopathies. Diabetic nephropathy and renal damage attributable to undiagnosed or inadequately managed hypertension are common factors in secondary glomerulopathies.

The specific location and features of glomerular injury can also be used to classify the disease. Glomerular involvement may be characterized as *diffuse* (all glomeruli) or *focal* (some but not all glomeruli). Within the affected glomeruli, lesions may be *global,* affecting all parts of the glomerulus, or present as patches *(segmental)* when only specific parts of the glomerulus are involved. If thickening of glomerular capillary walls is present, the glomerulopathy is *membranous.* Membranous changes involve the basement membrane and are usually due to accumulated deposition of immunoglobulins (IgG) and complement components

BOX 27.6 Glomerular Diseases

Primary Glomerulopathies
Acute proliferative glomerulonephritis
Postinfectious
Other
Rapidly progressive (crescentic) glomerulonephritis
Membranous nephropathy
Minimal-change disease
Focal segmental glomerulosclerosis
Membranoproliferative glomerulonephritis
Dense deposit disease
IgA nephropathy
Chronic glomerulonephritis

Systemic Diseases With Glomerular Involvement
Systemic lupus erythematosus
Diabetes mellitus
Amyloidosis
Goodpasture syndrome
Microscopic polyarteritis/polyangiitis
Wegener granulomatosis
Henoch–Schönlein purpura
Bacterial endocarditis

Hereditary Disorders
Alport syndrome
Thin basement membrane disease
Fabry disease

Modified from Kumar V, et al, editors: *Pathologic basis of disease,* ed 9, Philadelphia, 2015, Saunders.

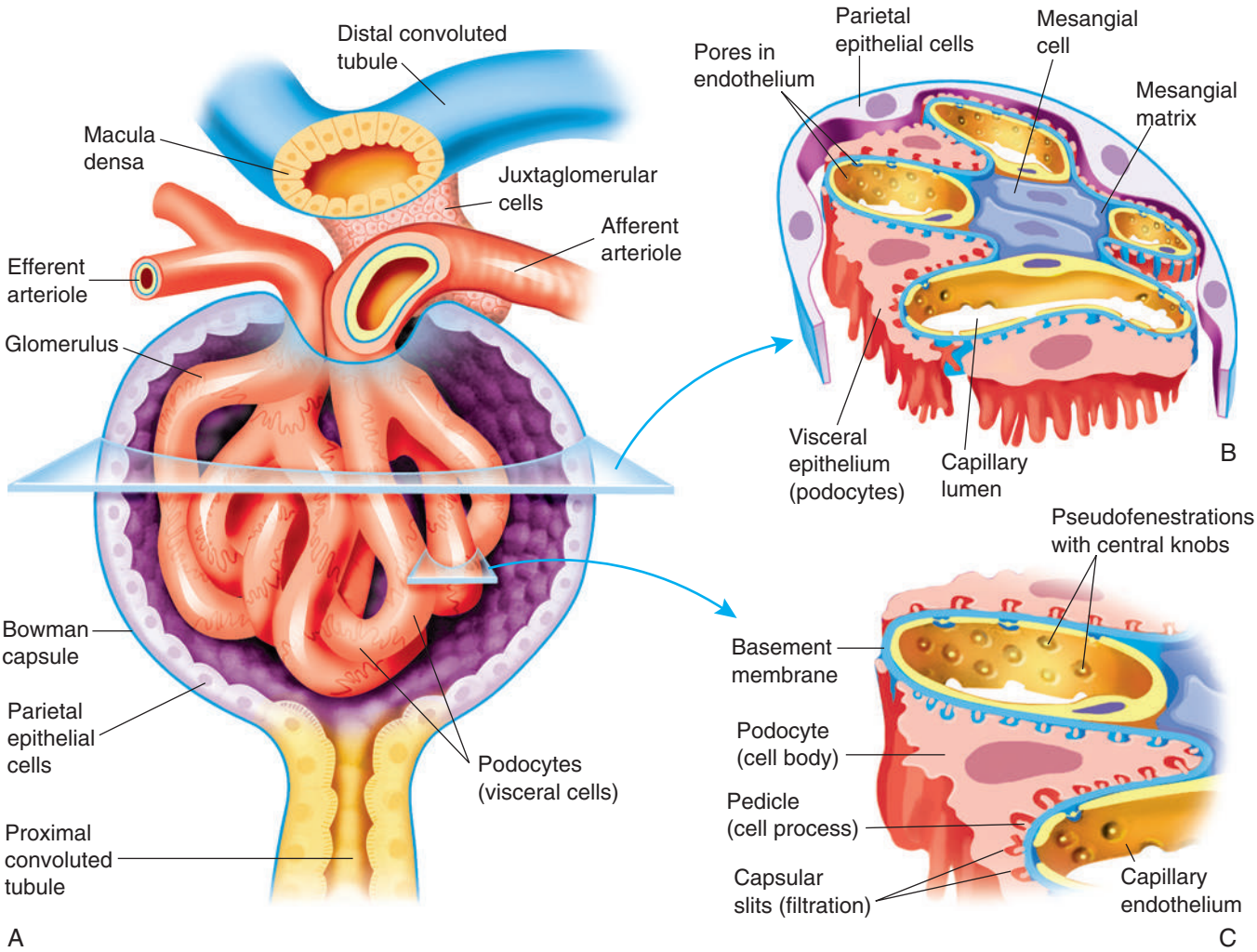


FIG 27.8 Anatomy of the glomerulus and juxtaglomerular apparatus. **A,** Longitudinal cross-section of glomerulus and juxtaglomerular apparatus. **B,** Horizontal cross-section of glomerulus. **C,** Enlargement of glomerular capillary filtration membrane. (From Huether S, McCance K: *Understanding pathophysiology,* ed 6, St Louis, MO, 2016, Mosby.)

(C5). *Sclerotic* changes refer to scarring attributable to persistent or recurrent injury and may occur within the capillaries, in the capillary space, or at the point of initiation of the proximal tubule. The site of deposition of noncellular materials as part of the glomerulopathic process may be described specifically as *mesangial, subendothelial,* or *subepithelial.* Renal biopsy and histologic examination often are required to determine a specific type of glomerular pathology.

Glomerulopathies commonly result in proteinuria, hematuria, red blood cell (RBC) urinary casts, decreased GFR, and hypertension; however, these manifestations may not all be present in a single pathology, and/or they may present along a continuum of degree. Proteinuria is a classic clinical manifestation associated with glomerulopathy. Glomerulonephritis (inflammatory) presents with hematuria, RBC casts, reduced GFR, generalized edema, or hypertension. Under normal circumstances, RBCs, albumin, and other large protein molecules are too large to pass through the membranes of the glomerulus, so they are retained in the bloodstream. Small proteins, including amino acids, are filtered through but actively reabsorbed along the nephron tubule. Transient proteinuria may be seen with diets high in protein or as a result of excessive exercise or emotional stress. Normally, protein lost in the urine amounts to less than 100 mg in 24 hours. Excessive protein in the urine occurs with damage to the glomerular basement membrane and slit pore proteins that normally block protein filtration. *Nephrotic syndrome* is a condition that can develop when there is excessive protein loss over time, characterized by protein loss ≥3 to 3.5 g in 24 hours. It is most commonly associated with minimal change disease (MCD)/lipoid nephrosis, focal segmental glomerulosclerosis, and membranous nephropathy, but can occur with any glomerulopathy if the injury is sufficient to allow significant protein loss. *Nephritic syndrome* is also a reflection of glomerular inflammation and leakiness, resulting in gross hematuria and RBC casts in the urine sediment.

The most common types of glomerular disease are acute postinfectious glomerulonephritis, IgA nephropathy, rapidly progressive/crescentic glomerulonephritis, nephrotic syndrome (includes membranous glomerulonephritis, focal segmental glomerulosclerosis, MCD/lipoid nephrosis), and chronic glomerulonephritis. The most common secondary glomerulopathy is diabetic nephropathy (see Chapter 41).

Glomerulonephritis

Glomerulonephritis includes an assortment of immune-mediated conditions that produce inflammation of the glomeruli. It may have a primary etiology, which is commonly believed to be autoimmune, or a secondary one, in which it is associated with a number of autoimmune, metabolic, malignant, or infectious systemic disorders.

Acute Glomerulonephritis

The term *acute glomerulonephritis* encompasses a constellation of inflammatory glomerulopathies that are characterized by the abrupt onset of varying degrees of hematuria, proteinuria, oliguria, azotemia, edema, and hypertension. A wide variety of triggers may initiate the inflammatory process (Fig. 27.9).

Postinfectious acute glomerulonephritis is historically most well known to follow skin (impetigo) and throat infections with specific strains of group A β-hemolytic streptococci. This etiology is less common today because of early diagnosis and treatment of streptococcal infections. Viruses are also commonly associated with postinfectious glomerulonephritis. The infectious organism stimulates the production of antibodies that bind to microbial antigens, initiating formation of antibody–antigen complexes. The onset of glomerular inflammation varies; in the case of a streptococcal etiology, damage to the glomerulus is initiated 1 to 3 weeks after the initial infection.

As with most other forms of glomerulonephritis, antibody deposition leading to inflammation is the key pathologic event. Proliferation of mesangial cells is initiated after immune complex deposition, resulting in characteristic proliferative lesions. Complement is activated within the glomerulus, resulting in release of chemotaxic factors and attracting macrophages, neutrophils, and T-helper cells. Lysosomal enzymes are released and attack glomerular walls. The changes in the structure of the glomerular membranes result in a decrease in the surface area available for filtration and allow substances that were previously restricted to enter the glomerular space. Locally acting vasoactive compounds such as angiotensin II and leukotrienes contract mesangial cells and reduce perfusion to glomerular capillaries, decreasing the GFR. In severe cases, the coagulation cascade is activated and fibrin is deposited within capillaries, further impairing perfusion and filtration.

Clinical manifestations of postinfectious acute glomerulonephritis vary in severity. Smoky or coffee-colored urine attributable to hematuria and RBC casts (nephritic syndrome) are the most common findings. The degree of proteinuria is variable and may reach nephrotic syndrome level. Reduced GFR presents as increased circulating volume, producing edema, hypertension, and oliguria with increasing serum levels of creatinine and nitrogenous wastes (BUN).

Diagnosis is based on the patient's history, clinical manifestations, and urinalysis results. Renal function is evaluated by BUN and creatinine levels. A renal biopsy may be indicated to identify a cause. Care is supportive and symptomatic. In severe cases dialysis may be needed to support renal function. Especially in children, resolution of the inflammatory process may occur within about 2 weeks (poststreptococcal) with no sequelae or complications. In some cases deterioration of renal function may continue, resulting in chronic glomerulonephritis, nephrotic syndrome (glomerulosclerosis), and chronic kidney disease.

IgA nephropathy (Berger disease) is the most commonly diagnosed type of primary glomerulonephritis worldwide. Upper respiratory tract or gastrointestinal viral infections appear to be the triggers, initiating the deposition of IgA in the glomerular mesangium. Increased abnormal IgA production and formation of IgA immune complexes cause mesangial injury. Hematuria is usually apparent within 1 to 2 days, but the proteinuria, edema, and hypertension common with other glomerulopathies are generally not evident. Prognosis is variable; as many as half the patients develop glomerulosclerosis and ESRD.

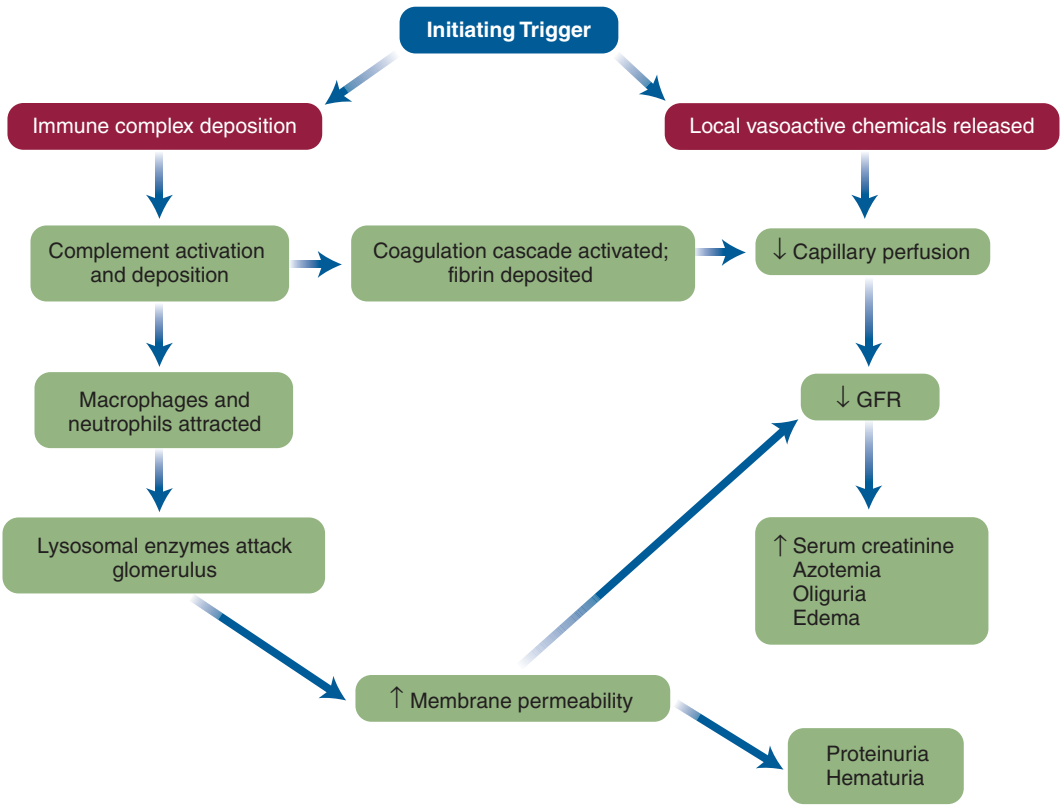


FIG 27.9 Summary of the pathophysiologic process of acute glomerulonephritis and associated clinical manifestations.

Crescentic Glomerulonephritis/Rapidly Progressive Glomerulonephritis (RPGN)

The condition previously known as *rapidly progressive glomerulonephritis* is now more often called *crescentic glomerulonephritis* because of its characteristic lesions. The lesions are proliferative in nature and composed of crescent-shaped depositions of accumulated epithelial cells, fibrin, and macrophages that are found in Bowman space. Podocytes that are progressively losing their markers of differentiation are also found. This form of glomerulonephritis may have an acute onset, with hematuria, proteinuria, and red cell casts, followed by a swift decline in renal function within a few months. It occurs predominantly in patients 20 to 50 years old. When not idiopathic, causes of crescentic glomerulonephritis fall into four general categories: (1) complication of an acute or subacute infection; (2) complication of a multisystem disease; (3) drug exposure; and (4) a primary disorder in the absence of other systemic disease. The most common infections associated with this disorder are post-streptococcal glomerulonephritis and infective endocarditis. Multisystem diseases associated with it are systemic lupus erythematosus, Henoch–Schönlein purpura, systemic necrotizing vasculitis, and Goodpasture syndrome. Examples of drugs identified as causative are penicillamine, hydralazine, allopurinol in the presence of vasculitis, and rifampin.

Goodpasture syndrome is an autoimmune disorder that results from the combination of glomerulonephritis with pulmonary alveolar hemorrhage and the presence of anti–glomerular basement membrane antibodies. Patients suffer with the clinical manifestations of glomerulonephritis and also shortness of breath and hemoptysis because of pulmonary involvement. Manifestations of pulmonary involvement may appear before those associated with glomerulonephritis.

Chronic Glomerulonephritis

Glomerular diseases that assume a progressive course ultimately develop into chronic kidney disease. These patients present with persistent proteinuria, with or without hematuria, and slowly declining renal function. In many cases, patients appear to have recovered from the initial insult. Proteinuria and hypertension are both capable of advancing renal damage. The pathophysiology is an extension of that seen in acute glomerulonephritis. Proliferative and membranous lesions are present, but sclerotic injury dominates, resulting in ongoing fibrotic changes. Tubulointerstitial damage contributes to the reduction in renal function. Nephrons atrophy, and ultimately the kidneys become small, scarred, and nonfunctional. The chronic kidney disease progresses to ESRD, at which time dialysis or transplantation is required (Chapter 28).

Nephrotic Syndrome

Nephrotic syndrome is a glomerulopathy in which there is a urinary elimination of >3 to 3.5 grams of protein per day due to glomerular leakiness. Ultimately, most patients will develop the full constellation of nephrotic syndrome manifestations: hypoalbuminemia, hyperlipidemia, generalized edema, and a propensity for thrombus formation. The most common primary causes are MCD (lipoid nephrosis), idiopathic focal segmental glomerulosclerosis, and membranous nephropathy. Several systemic diseases, such as systemic lupus erythematosus and Henoch–Schönlein purpura, as well as infections, malignancies, and vasculitis, have been found to be associated with nephrotic syndrome, but the most common cause in adults is diabetes mellitus.

The pathophysiology of nephrotic syndrome is shown in Fig. 27.10. The increased permeability of the glomerular membrane allows large

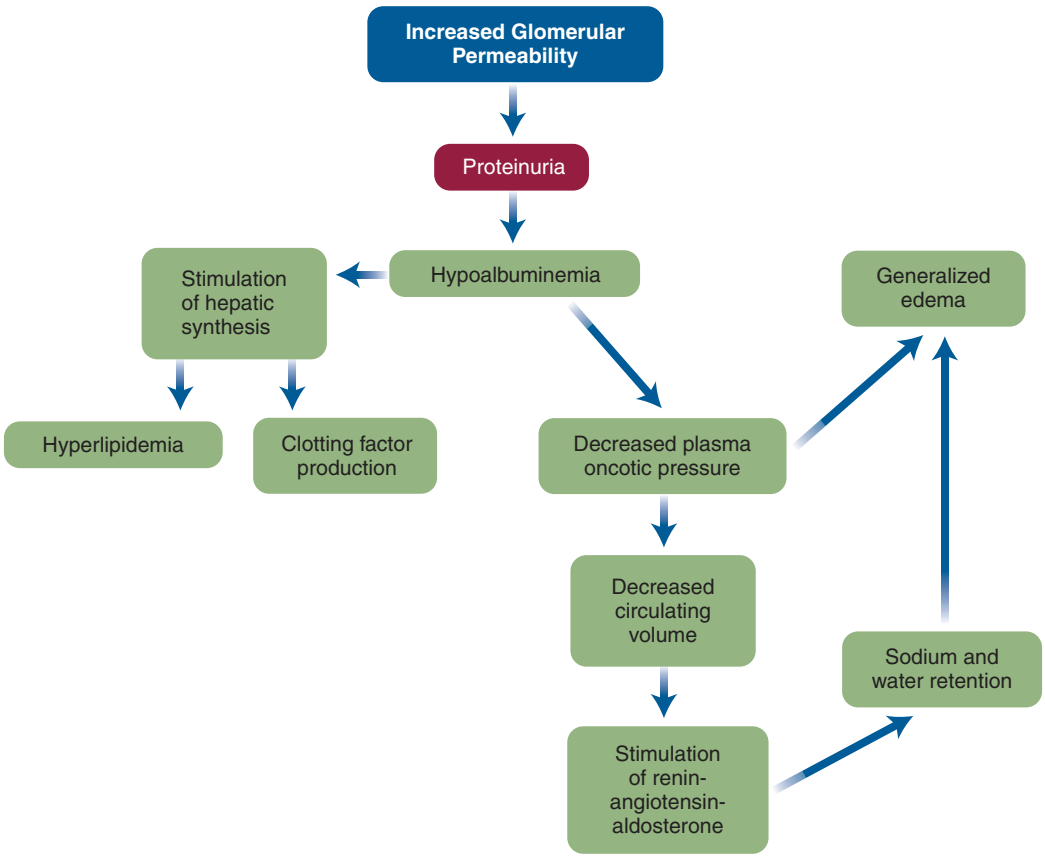


FIG 27.10 Pathophysiologic process of nephrotic syndrome.

quantities of protein to leave the bloodstream and exit the body in the urine. This produces hypoalbuminemia and proteinuria. The low serum albumin concentration serves to stimulate the liver to increase production of various substances, including lipoproteins, producing hyperlipidemia. Levels of total cholesterol and low-density lipoproteins are elevated. Lipid casts or fat droplets may appear in the urine. Oval fat bodies, epithelial cells from the tubules, may also be sloughed. There is an increased risk of thrombotic events in nephrotic syndrome related to stimulation of hepatic synthesis of clotting factors. Edema is a common clinical manifestation of nephrotic syndrome. Two related processes contribute to edema formation. As the serum albumin is lost, the oncotic pressure within the blood vessels declines, allowing excessive fluid to filter into the interstitial spaces and causing a generalized edema. Reduction in circulating volume results in the activation of the renin–angiotensin–aldosterone system, causing sodium and then water retention and more edema.

Nephrotic syndrome is generally treated with drugs to reduce immune injury and inflammation such as corticosteroids. Elevated lipid levels are routinely treated with lipid-lowering therapy if rapid recovery from nephrotic syndrome is not anticipated. Angiotensin II receptor blockers and/or angiotensin-converting enzyme inhibitors may reduce intraglomerular pressure and reduce protein loss, as well as treat systemic hypertension.

Membranous Nephropathy (MN)

MN accounts for about one-third of the cases of adult-onset nephrotic syndrome. Twenty percent of these cases have secondary causes such as infections, autoimmune diseases, and malignancies. For lack of any concrete evidence, the remaining 80% are said to be idiopathic. Fig. 27.11 illustrates the histologic findings with MN and the pathologic changes in the glomerulus. Immune deposits and as-yet-unidentified antigens collect in the subepithelium of the distal portion of the basement membrane and produce a membranous thickening. Spontaneous remission occurs in approximately 25% of cases, 25% develop persistent nonnephrotic-range proteinuria, 25% develop persistent nephrotic syndrome, and 25% of the individuals with MN progress to ESRD.

Focal Segmental Glomerulosclerosis (FSGS)

FSGS is now the most common cause of idiopathic (or primary) nephrotic syndrome among adults in the United States. It may develop secondarily to a number of immunologic, metabolic, interstitial, hemodynamic, or other conditions. Only a portion of the glomeruli are affected (segmental), and glomeruli within the kidney are not uniformly affected (focal). This is a disease of glomeruli podocytes that over time progresses to scarred lesions. Epithelial injury and endothelial

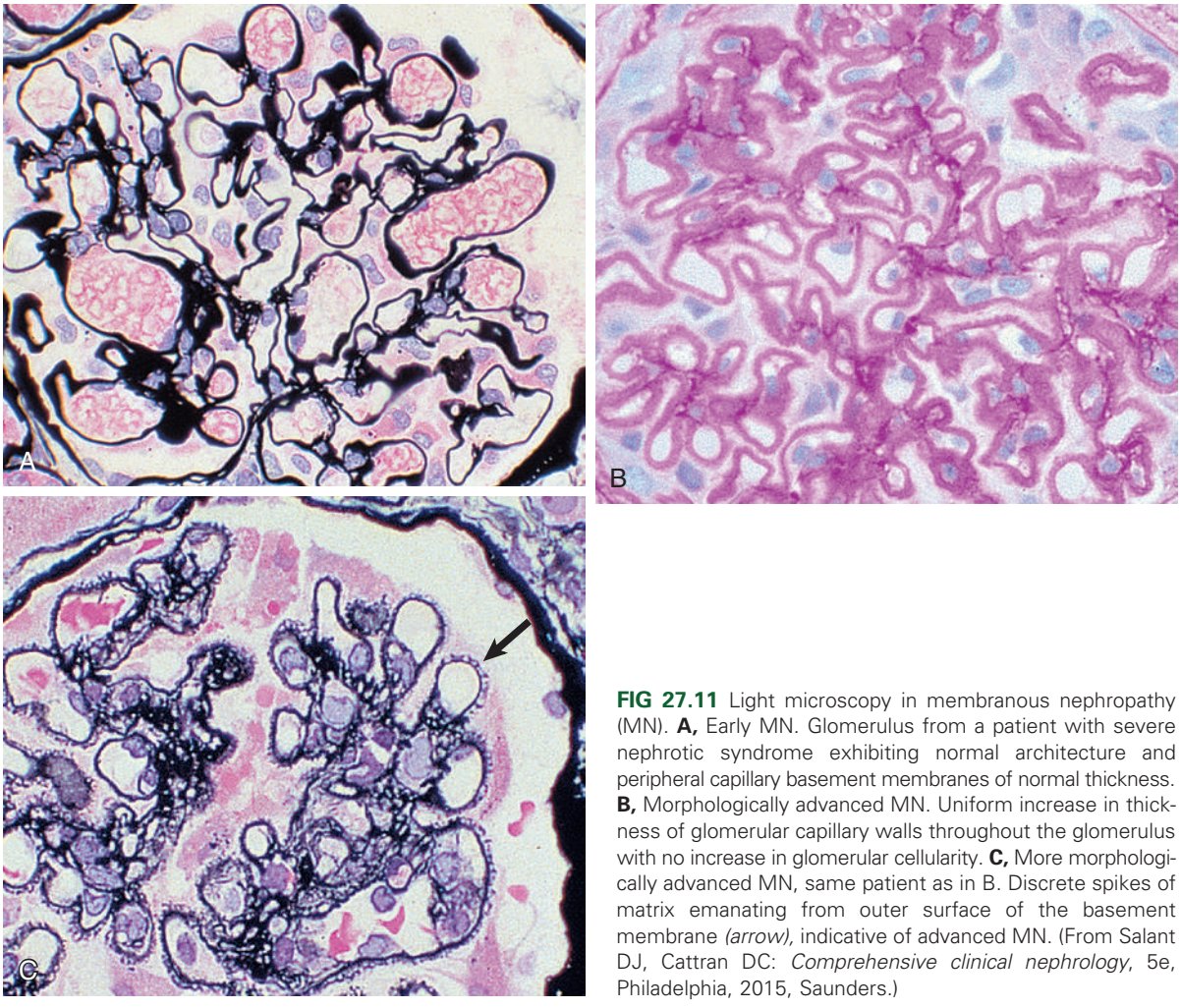


FIG 27.11 Light microscopy in membranous nephropathy (MN). **A,** Early MN. Glomerulus from a patient with severe nephrotic syndrome exhibiting normal architecture and peripheral capillary basement membranes of normal thickness. **B,** Morphologically advanced MN. Uniform increase in thickness of glomerular capillary walls throughout the glomerulus with no increase in glomerular cellularity. **C,** More morphologically advanced MN, same patient as in B. Discrete spikes of matrix emanating from outer surface of the basement membrane *(arrow),* indicative of advanced MN. (From Salant DJ, Cattran DC: *Comprehensive clinical nephrology*, 5e, Philadelphia, 2015, Saunders.)

cellular injury initiate the protein loss across the glomerular membrane. In addition to proteinuria, patients can develop mild hematuria, hypertension, and azotemia. Few patients will experience complete remission. Once proteinuria reaches nephrotic syndrome levels, half of those with FSGS will reach ESRD within a few years.

Minimal Change Disease (MCD)

Previously called *lipoid nephrosis,* MCD is the primary cause of idiopathic nephrotic syndrome in children. It is usually initiated by an allergic or immune disorder, including lymphomas. The result of the minimally altered structure of the podocytes is a less effective glomerular filtration barrier and the loss of large amounts of albumin in the urine. Onset of edema, heavy proteinuria, and hypoalbuminemia is sudden. Although a longer course is usually needed in children, MCD responds well to treatment with corticosteroids. Fewer patients progress to ESRD with MCD than the other common forms of nephrotic syndrome.

KEY POINTS

- Glomerulopathies alter glomerular capillary and podocyte structure and function. Damage usually is mediated by immune processes. The glomerular damage may result in some combination of hematuria, proteinuria, red blood cell (RBC), casts, decreased glomerular filtration rate (GFR), edema, and hypertension.
- Glomerulonephritis is due to an immune response to a variety of potential triggers and may have a primary or secondary etiology. Attraction of immune cells to the area of inflammation results in lysosomal degradation of the basement membrane. The GFR may fall, in part because of contraction of mesangial cells, resulting in decreased surface area for filtration.
- Glomerulonephritis may be classified as acute, crescentic, or chronic. Acute forms are usually triggered by infection. The cause of the crescentic form is often unknown, but it may be secondary to autoimmune processes such as Goodpasture syndrome. Chronic forms are those that progress to chronic kidney disease.
- Treatment of glomerulonephritis may include corticosteroids, immune suppressants, and supportive measures such as dietary and fluid management and management of systemic and renal hypertension. End-stage renal disease (ESRD) is a common outcome of chronic glomerulonephritis, necessitating dialysis or kidney transplantation.
- Nephrotic syndrome occurs because of increased glomerular permeability to proteins, which results in a urinary loss of 3 to 3.5 g of protein or more per day. Proteinuria leads to hypoalbuminemia and generalized edema as a result of decreased blood colloid osmotic pressure. Hyperlipidemia and hypercoagulability are thought to occur because of a generalized increase in liver activity stimulated by hypoalbuminemia.
- Treatment of nephrotic syndrome is conservative, consisting of management of symptoms and the underlying process that initiated the syndrome, when possible. Although many cases resolve spontaneously, others progress to ESRD.

SUMMARY

Many diseases can cause damage to the kidney. Any process that disrupts the normal architecture of the kidney will cause altered function, whether in the glomeruli, the vascular tree, or the collecting/draining system. Despite the kidney's resiliency and capacity to respond to treatment, severe acute or chronic damage may progress to chronic kidney disease. The stages of acute renal failure and chronic kidney disease and their management are discussed in Chapter 28.

RESOURCES

Congenital and Polycystic Kidney Disease

Balcells T, Ars Criach A: Molecular diagnosis of autosomal dominant polycystic kidney disease. *Nefrologia* 31:35–43, 2011.

Dillman JR, et al: Hereditary renal cystic disorders: imaging of the kidneys and beyond. *Radiographics* 37(3):924–946, 2017.

Halvorson CR, Bremmer MS, Jacobs SC: Polycystic kidney disease: inheritance, pathophysiology, prognosis, and treatment. *Int J Nephrol Renovasc Dis* 3:69–83, 2010.

Peinter L, Borner C: Role of apoptosis in the development of autosomal dominant polycystic kidney disease (ADPKD). *Cell Tissue Res* 2017 May 30. doi:10.1007/s00441-017-2628-2636.

Sweeney WE, Jr, Avner ED: Emerging therapies for childhood polycystic kidney disease. *Front Pediatr* 5:77, 2017. doi:10.3389/fped.2017.00077.

Renal Cancer

Arai E, Kanai Y: Genetic and epigenetic alterations during renal carcinogenesis. *Int J Clin Exp Pathol* 4:58–73, 2011.

Bahrami A, et al: The genetic factors contributing to the development of Wilm's tumor and their clinical utility in its diagnosis and prognosis. *J Cell Physiol* 2017 May 20. doi:10.1002/jcp.26021.

Brok J, et al: Biology and treatment of renal tumours in childhood. *Eur J Cancer* 68:179–195, 2016.

National Cancer Institute: *Genetics of Kidney Cancer (Renal Cell Cancer) (PDQ®)—Health Professional Version*. https://www.cancer.gov/types/kidney/hp/kidney-genetics-pdq.

Tazi EM, et al: Advanced treatments in non-clear renal cell carcinoma. *Urol J* 8:1–11, 2011.

Wein A, et al: *Campbell-Walsh urology*, ed 11, Philadelphia, 2016, Elsevier Saunders.

Pyelonephritis and Urinary Tract Infection

Alevizopoulos A, et al: UROLITHIASIS managing small ureteral stones: a retrospective study on follow-up, clinical outcomes and cost-effectiveness of conservative management vs. early surgery. *Curr Urol.* 9(1):36–43, 2016.

Chung A, et al: Bacterial cystitis in women. *Aust Fam Physician* 39:295–298, 2010.

Dawson-Hahn EE, et al: Short-course versus long-course oral antibiotic treatment for infections treated in outpatient settings: a review of systematic reviews. *Fam Pract* 2017 May 9. doi:10.1093/fampra/cmx037.

Evan AP: Physiopathology and etiology of stone formation in the kidney and the urinary tract. *Pediatr Nephrol* 25:831–841, 2010.

Lane MC, Mobley HLT: Role of p-fimbrial-mediated adherence in pyelonephritis and persistence of uropathogenic *Escherichia coli* (UPEC) in the mammalian kidney. *Kidney Int* 72:19–25, 2007.

Lundstedt A, et al: Inherited susceptibility to acute pyelonephritis: a family study of urinary tract infection. *J Infect Dis* 195(8):1227–1234, 2007.

Samplaski MK: Less invasive ways to remove stones from the kidneys and ureters. *Cleve Clin J Med* 76:592–598, 2009.

Tolkoff-Rubin NE, Cotran RS, et al: Urinary tract infection, pyelonephritis, and reflex nephropathy. In Taal M, editor: *Bremmer and Rector's the kidney*, ed 9, Philadelphia, 2012, Saunders.

Wein A, et al: *Campbell-Walsh urology*, ed 11, Philadelphia, 2016, Elsevier Saunders.

Glomerulopathy

Cattran DC, Brenchley PE: Membranous nephropathy: integrating basic science into improved clinical management. *Kidney Int* 91(3):566–574, 2017.

Colucci M, et al: Immunology of idiopathic nephrotic syndrome. *Pediatr Nephrol* 2017 Apr 27. doi:10.1007/s00467-017-3677-5.

Hall JE: *Guyton and Hall textbook of medical physiology*, ed 13, Philadelphia, 2016, Elsevier Saunders.

Hinkes BG, et al: Nephrotic syndrome in the first year of life: two thirds of cases are caused by mutations in 4 genes *(NPHS1, NPHS2, WT1, and LAMB2)*. *Pediatrics* 119:907–919, 2007.

Mizuno M, et al: Complement regulation and kidney diseases: recent knowledge of the double-edged roles of complement activation in nephrology. *Clin Exp Nephrol* 2017 Mar 24. doi:10.1007/s10157-017-1405-x.

Rodriques JC, et al: IgA Nephropathy. *Clin J Am Soc Nephrol* 12(4):677–686, 2017.

Zhou Y, et al: Steroids in the treatment of IgA nephropathy to the improvement of renal survival: a systematic review and meta analysis. *PLoS ONE* 2011, published online 4/12/11. http://dx.doi.org/10.1371/journal.pone.0018788.

28

Acute Kidney Injury and Chronic Kidney Disease

Cheryl Rockwell and Robin Y. Beeman

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- How do the prerenal, intrinsic, and postrenal types of acute kidney injury differ in etiology, prognosis, clinical manifestations, and management?
- What are the characteristic clinical and laboratory findings in each of the three stages of acute tubular necrosis (acute intrarenal renal failure)?
- What is the relationship between the degree of nephron loss, reductions in glomerular filtration rate, and the stages of chronic kidney disease?
- What are the similarities and differences between acute kidney injury and end-stage chronic kidney disease?
- What are the characteristic findings of uremic syndrome?
- How can acute kidney injury be prevented, and how is it treated?
- How is the progression of chronic kidney disease slowed?
- How is end-stage renal disease treated?

CHAPTER OUTLINE

Acute Kidney Injury, 593
Etiology and Pathophysiology, 594
Prerenal Kidney Injury, 594
Postrenal Kidney Injury, 595
Intrinsic/Intrarenal Kidney Injury, 595
Clinical Presentation of Acute Kidney Injury, 596
Prodromal Phase, 599
Oliguric Phase, 599
Postoliguric Phase, 600
Chronic Kidney Disease, 601
Risk Factors, 601
Pathophysiology of Progression of Chronic Kidney Disease, 602
Stages of Chronic Kidney Disease, 602
Complications of Chronic Kidney Disease, 602
Hypertension and Cardiovascular Disease, 603
Uremic Syndrome, 603
Metabolic Acidosis, 603
Electrolyte Imbalances, 603
Mineral and Bone Disorders, 603
Malnutrition, 603
Anemia, 603
Pain, 603
Depression, 604
Clinical Management, 604
Acute Kidney Injury, 604
Chronic Kidney Disease, 604
Hypertension and Cardiovascular Disease, 605
Metabolic Acidosis, 605
Fluid and Electrolyte Imbalances, 605
Bone and Mineral Disorders, 605
Malnutrition, 605
Anemia, 605
Pain, 605
Depression, 606
Acute-on-Chronic Kidney Disease, 606
Dialysis, 606
Kidney Transplant, 606
Chronic Kidney Disease in Older Adults, 607

The kidneys have a number of regulatory roles within the body. These key functions include the regulation of body fluid volume and osmolality, electrolyte balance, and acid–base balance. Additionally, the kidneys produce and secrete hormones and excrete metabolic waste products and foreign materials. When the kidneys are unable to carry out these functions on a temporary or permanent basis, the ramifications are significant to each body system.

ACUTE KIDNEY INJURY

Acute kidney injury (AKI), also known as *acute renal failure,* represents a broad spectrum of kidney diseases ranging from minor changes in renal function to complete renal failure requiring renal replacement therapy. AKI is the sudden reduction of kidney function causing disruptions in fluid, electrolyte, and acid–base balances; retention of nitrogenous

waste products; increased serum creatinine level; and decreased glomerular filtration rate (GFR). The Kidney Disease Improving Global Outcomes (KDIGO) 2012 guidelines define AKI as any of the following:

- Increase in Serum creatine by ≥0.3 mg/dL (≥26.5 µmol/l) within 48 hours; or
- Increase in Serum creatine to ≥1.5 times baseline, which is known or presumed to have occurred within the prior 7 days; or
- Urine volume <0.5 mL/kg/h for 6 hours.

Classification criteria for AKI have been developed to identify kidney injury and improve patient outcomes. The 5-point system is known as the *RIFLE classification system* (R = Risk of injury, I = Injury, F = Failure, L = Loss of function, and E = End-stage kidney disease) (Table 28.1). The first three stages indicate severity of kidney injury, and the last two stages represent patient outcomes.

The incidence of AKI in hospitalized patients ranges from 15.5% to 22.7%, with higher rates in elderly patients. Mortality is fourfold higher among patients with AKI compared with those without AKI. Additionally, length of hospital stay is twice as long for patients with AKI compared with those without, and mortality is higher in patients with more severe AKI.

Etiology and Pathophysiology

The risk of developing AKI is increased by certain preexisting conditions. These comorbidities include preexisting kidney impairment, cardiovascular and peripheral vascular disease, hypertension, diabetes mellitus, heart failure, malignancies, benign prostatic hypertrophy, and exposure to potentially nephrotoxic drugs. Not only are the elderly more likely to have one or more of these conditions, but also aging itself results in changes within the kidney that make it more susceptible to damage. The average decrease in renal blood flow is approximately 10% per decade starting at age 40 years, with an average decrease in the GFR by about 0.8 to 1.0 mL/min/1.73 m^2 per year. These alterations in function increase the risk of AKI and can negatively affect overall prognosis. The aging kidney is less capable of concentrating and diluting urine, conserving sodium, producing prostaglandin, and maintaining renin and aldosterone levels. See Chapter 26 for a discussion of the effects of aging on renal function.

Acute loss of renal function is attributed to conditions that affect renal perfusion (prerenal); factors that obstruct urine flow distal to the kidney (postrenal); or circumstances within the kidney blood vessels, tubules, glomeruli, or interstitium (intrinsic). These anatomic delineations are broadly seen as the types or causes of AKI, but the specific etiology must also be identified. Determining the specific etiology, as well as the type of AKI, is essential for effective management. Box 28.1 shows the types of AKI and some of their etiologies.

Prerenal Kidney Injury

When AKI develops because of diminished perfusion of the kidney, it is termed *prerenal* kidney injury because the etiology occurs before the kidney itself. As seen in Box 28.1, this can be due to an absolute or relative decrease in circulating volume or abnormalities of renal hemodynamics. Actual and relative depletion of volume are the most common etiologies. Fever, vomiting, diarrhea, burns, hemorrhage, and overuse of diuretic therapy produce fluid volume deficits that can lead to prerenal kidney injury. Decreased renal perfusion also results if large volumes of fluid collect in extravascular spaces as in edema (interstitial space) or ascites (peritoneal space). Any number of conditions reduces the ability of the heart to generate a cardiac output sufficient to meet the needs of body organ systems. Even though the kidney receives 20% to 25% of the cardiac output, that volume may be inadequate when the cardiac output is markedly decreased by cardiogenic shock, heart failure, or lethal ventricular dysrhythmias.

Although most patients who develop prerenal kidney injury have an episode of decreased blood pressure (BP) that results in decreased perfusion to the kidney, in some cases perfusion drops without the BP falling below normal. These "normotensive" cases of prerenal kidney injury arise in susceptible individuals with very modest reductions in BP who have preexisting impairments in renal autoregulation. Use of nonsteroidal antiinflammatory drugs (NSAIDs), angiotensin-converting enzyme inhibitors (ACEIs), and angiotensin II (AII) receptor blockers is known to interfere with renal vascular autoregulation and can precipitate prerenal kidney injury in certain populations of patients. This includes those who are older than 60 years of age with atherosclerotic cardiovascular disease or who have preexisting renal insufficiency (elevated serum creatinine level), heart failure, advanced liver disease, or nephrotic syndrome. These drugs cause either vasoconstriction of afferent arterioles (NSAIDs) or vasodilation of efferent arterioles (ACE inhibitors and AII blockers); either of these actions results in a decrease in glomerular perfusion pressure. Thrombus, embolus, dissection, or stenosis of the renal arteries will also result in prerenal kidney injury, and the risk increases significantly if ACE inhibitors or AII blockers are being used.

Prerenal oliguria is the kidney's normal physiologic response to a decrease in perfusion, and at least for a time the renal tissue is unharmed. Neurohumoral mechanisms of local autoregulation are activated as the kidneys attempt to autoregulate perfusion and maintain GFR, and systemic mechanisms such as the renin–angiotensin–aldosterone system (RAAS) act to increase the total circulating volume. The sensed decrease in renal blood flow results in a decrease in GFR and urine output. Because of the kidney's ability to tolerate significant reduction in perfusion (as long as it is not greater than 20% to 25% of normal), and as

TABLE 28.1 RIFLE Classification for Staging Acute Kidney Injury

Stage	GFR Criteria	Urine Output Criteria
Risk	Increased SCr to >1.5 × baseline	Urine output <0.5 mg/kg/h for >6 h
Injury	Increased SCr to >2 × baseline	Urine output <0.5 mg/kg/h for >12h
Failure	Increased SCr to >3 × baseline; or an increase of ≥0.5 mg/dL to a value of ≥4 mg/dL	Urine output <0.3 mg/kg/h for >12h or anuria for >12h
Loss	Need for RRT for >4 wk	
End-stage kidney disease	Need for RRT for >3 mo	

From Palevsky et al: KDOQI US commentary on the 2012 KDIGO clinical practice guideline for acute kidney injury, *Am J Kidney Dis* 61(5):649–672, 2013.
RRT, Renal replacement therapy; *SCr,* serum creatinine.

BOX 28.1 Types of Acute Kidney Injury

Prerenal

- Absolute decrease in circulating volume
  - Hemorrhage
  - Dehydration
  - Burns
- Relative decrease in circulating volume
  - Distributive shock (neurogenic, anaphylactic, septic)
  - Third-spacing and edema
  - Decreased cardiac output
    - Cardiogenic shock
    - Dysrhythmias
    - Cardiac tamponade
    - Heart failure
    - Myocardial infarction
- Primary renal hemodynamic abnormalities
  - Occlusion or stenosis of renal artery*
  - Drug-induced impairment of renal autoregulation in susceptible persons†

Postrenal

- Benign prostatic hyperplasia
- Kinked or obstructed catheters
- Intraabdominal tumors
- Strictures
- Calculi

Intrarenal/Intrinsic

- Tubular (acute tubular necrosis)
  - Ischemic
    - Prolonged prerenal failure
    - Transfusion reactions
    - Rhabdomyolysis
  - Nephrotoxic
    - Prolonged postrenal failure
    - Certain antimicrobials (antibiotics; antifungal and antiviral drugs)
    - Radiographic contrast media
    - Certain cytotoxic chemotherapy agents
    - Recreational drugs (amphetamines, heroin)
    - Environmental agents (heavy metals, carbon tetrachloride, insecticides)
    - Snake and insect venom
- Glomerular
  - Acute glomerulonephritis
- Interstitial
  - Acute allergic interstitial nephritis
  - Acute pyelonephritis
- Vascular
  - Vasculitis
  - Emboli
  - Nephrosclerosis (due to primary hypertension, hypertensive emergencies, and urgency)

*Use of ACE inhibitors or AII receptor blockers increases the risk.
†Preexisting chronic renal insufficiency, cirrhosis, heart failure, or elderly persons (>60 years) with atherosclerotic cardiovascular disease, hypotension, diuretic use, or nephritic syndrome.

long as the hypoperfusion etiology is identified and corrected, prerenal oliguria will not affect the parenchyma of the kidney. Efforts to restore adequate perfusion should be fully effective in restoring normal renal function within 1 to 2 days. In the case of normotensive patients who have impaired perfusion, by far the smallest subset of prerenal kidney injury cases, interventions must be targeted to the specific etiology. Regardless of the etiology, persistent prerenal kidney injury will result in hypoxic renal cells. If hypoxia continues and ischemia lasts more than a few hours, prerenal kidney injury will progress to intrinsic kidney injury.

Postrenal Kidney Injury

Obstruction of the normal outflow of urine from the kidneys can result in postrenal kidney injury. Box 28.1 lists the most common etiologies of this type of AKI. If only one kidney is affected, the activity of the remaining kidney will increase to maintain fluid and electrolyte balance. Obstruction of the renal pelvis or ureters of both kidneys, of the bladder outlet, or of the urethra will result in discernible postrenal kidney disease. This type of AKI is more common in the elderly and the most amenable to intervention. Normalization of renal function depends on the length of time the obstruction persists. Should obstruction persist, the increasing retrograde pressure of urine will result in acute tubular necrosis (intrinsic AKI), and if the obstruction continues over several days or weeks, irreversible damage to the kidney may result.

Intrinsic/Intrarenal Kidney Injury

AKI intrinsic to the kidney itself is further classified by the specific anatomic area involved: vascular, interstitial, glomerular, or tubular (see Box 28.1). Some references incorporate the vascular and tubular classifications together because damage to one ultimately leads to damage to the other. All of these etiologies are capable of producing the potentially reversible rapid decline in renal function that is AKI. When the small vessels within the kidney are inflamed, obstructed, or damaged by an acute hypertensive episode, the injury may be sufficient to impair nephron functioning. Acute glomerulonephritis is due to an abnormal immune reaction, whereby immune complexes are deposited in the basement membrane of the glomerulus, thereby damaging the glomeruli. Normal renal function is disrupted to some degree during this acute inflammatory process, usually lasting about 2 weeks. Acute glomerulonephritis is discussed in detail in Chapter 27. Inflammation of interstitial tissues may be sufficient to result in intrinsic kidney injury. This is usually due to an infection of the kidney (pyelonephritis), an allergic reaction to medications, or an autoimmune disease. Long-standing pyelonephritis causes damage to the renal medulla and progressive loss of functional renal tissue.

Intrinsic kidney injury is commonly associated with acute tubular necrosis (ATN), which itself has many potential etiologies. ATN is the result of tubular cell injury, primarily attributable to ischemia or exposure to nephrotoxic substances. It accounts for nearly half of all cases of AKI in hospitalized patients and is the most prevalent form of intrinsic AKI in the elderly. Sepsis is the most common cause of ischemic ATN and may develop in about 50% of critically ill patients. It causes vasodilation leading to hypoperfusion within the kidney. Ischemic ATN is often identified in the elderly surgical patient, and septic ATN is the cause for about 33% of cases in the elderly.

Nephrotoxins are another leading cause of ATN. The list of medications and chemicals toxic to the kidney is expansive, each one inducing a specific toxic reaction in the tubular cells and causing the death of many of them. Examples of commonly used nephrotoxic medications include aminoglycosides, NSAIDs, amphotericin B, cisplatin, and tetracycline.

Of all of these nephrotoxins, contrast medium is often the most common offending agent. Contrast-induced AKI (CI-AKI) (also known as *contrast-induced nephropathy*) is a major cause of AKI in hospitalized elderly patients and can develop within 12 to 24 hours of contrast administration. By definition, CI-AKI is defined as a rise in serum creatinine of ≥0.5 mg/dL (≥44 mmol/l) or a 25% increase from baseline

value, assessed at 48 hours after a radiologic procedure. Risk factors for developing CI-AKI are underlying kidney insufficiency, age greater than 70, volume depletion, repeated exposures to contrast media in a short time, and coexisting heart failure or diabetes mellitus. Prevention is aimed at avoiding unnecessary contrast administration to high-risk patients, avoiding multiple procedures over a 24- to 48-hour period, minimizing contrast-media volume, using low-osmolar and isoosmolar contrast media, and receiving adequate administration of hypotonic and isotonic intravenous (IV) fluids before and after contrast administration. ATN caused by contrast media results in prolonged hospitalization, increased health care costs, and an increased risk of death. Prolonged prerenal kidney injury, perioperative and postoperative hypotension, hemorrhage, gastrointestinal drainage, and preoperative cardiac complications also contribute to many cases of ATN.

In ATN, two pathophysiologic processes result in the rapid decrease in GFR: a vascular process and a tubular process. The two processes are interrelated, and the severity of one contributes to the severity of the other. Renal blood flow is decreased by 30% to 50% in ATN, and blood is shunted from the medulla to the cortex, further compromising the medullary cells. Local vasoconstrictors such as prostaglandins and leukotrienes are released, and the effects of sympathetic nervous system (SNS) stimulation contribute further to the vasoconstriction. Hypoxia or direct tubular damage, attributable to toxins, initiates an inflammatory response, activating the cascade of inflammatory mediators. When perfusion is restored, more inflammatory cells are enlisted and reperfusion injury perpetuates damage in some areas. Cells in part of the proximal tubule and outer cortex begin the repair process when perfusion is returned, but endothelial cells and those in the ascending limb continue to be injured, become necrotic, and commit apoptosis, resulting in a further decline in GFR.

The pathogenesis of the tubular process is a reflection of the ischemia and the inflammatory process (Fig. 28.1). Damaged tubular epithelial cells, both viable and nonviable, are shed from the basement membrane and accumulate in the tubular filtrate, where they obstruct filtrate flow. These cells combine with inflammatory cells and debris to form casts that contribute further to the urinary obstruction. Obstructed urinary flow produces an increased pressure within the nephron that is communicated backward to the glomerulus, further reducing the GFR. The increasing pressure generated by the tubular obstruction forces filtrate through the partially denuded tubular basement membrane into the interstitial space and even into the bloodstream, a process known as *tubular backleak*. Half of the already limited quantity of glomerular filtrate may be lost to the interstitium by this process.

Recovery after ATN is highly dependent on the extent of injury and slower than in the other two types of AKI. If sufficient destruction of the basement membrane occurs, there may be no recovery and the patient develops end-stage renal disease (ESRD), the final stage of chronic kidney disease (CKD). But the tubules can repair themselves within 10 to 20 days when the basement membrane is intact and new epithelial cells are produced on that surface. As with the other types of renal failure, the clinical presentation of ATN is primarily a reflection of the loss of the normal functions performed by the kidney.

Clinical Presentation of Acute Kidney Injury

Prerenal kidney injury can be reversed if treated before perfusion drops to below 20% of normal and ischemia occurs. It is at this point that ATN develops. Prerenal or postrenal kidney injury will ultimately progress to intrinsic kidney injury if not corrected within a few hours. The course of AKI is roughly divided into three phases, and the clinical presentation varies with the phase (Fig. 28.2). The laboratory findings that differentiate prerenal oliguria from intrinsic kidney injury are shown in Table 28.2. Table 28.3 provides the laboratory profile associated with renal failure; some of these findings are more likely to be noted in ESRD than in AKI. Though serum creatinine levels begin to increase within 12 hours to 2 days after injury, new biomarkers are being investigated to detect AKI earlier than the rise in serum creatinine level, with the promise of leading to earlier detection and treatment. Examples of

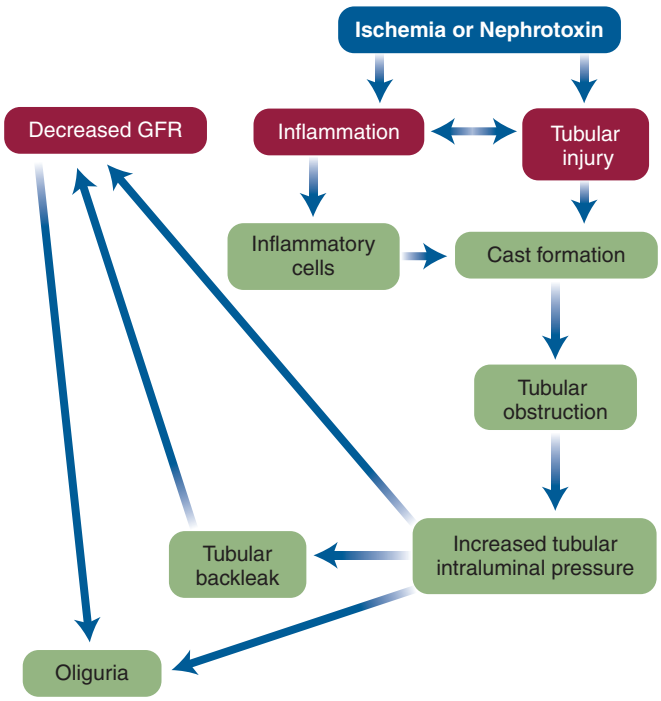


FIG 28.1 Pathogenesis of acute tubular necrosis.

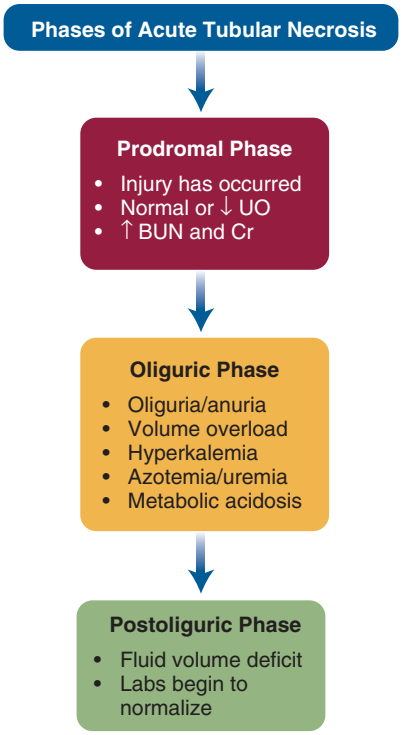


FIG 28.2 Phases of acute tubular necrosis and primary clinical issues. *BUN,* Blood urea nitrogen; *Cr,* creatinine; *UO,* urine output.

TABLE 28.2 Laboratory Value Differences in Prerenal Oliguria and Intrinsic AKI

Laboratory Test	Prerenal Findings	Intrarenal Findings
FENa %	<1	>2
Proteinuria	Absent	Possible
Urine specific gravity	>1.020	1.010–1.020
Urine osmolality (mOsm/kg)	>500	300–500
BUN/creatinine ratio	>20:1	10–20:1
Urine sodium concentration (mmol/L)	<10	>20
Urinary sediment	Few hyaline casts	Tubular, RBC, and WBC casts

Derived from Gammill HS, Jeyabalan A: Acute renal failure in pregnancy, *Crit Care Med* 33(10):S372–S384, 2005; Needham E: Management of acute renal failure, *Am Fam Physician* 72(9):1739–1746, 2005; Lameire N, Van Biesen W, Vanholder R: Acute renal failure, *Lancet* 365:417–430, 2005.
AKI Acute kidney injury; *FENa %,* fraction of excreted sodium, percent.

TABLE 28.3 Laboratory Profile for Renal Disease

Test	Normal Range for Adults	Values in Renal Disease	Comments
Test to Evaluate Removal of Nitrogenous Wastes			
Serum creatinine	*Male:* 0.6–1.2 mg/dL *Female:* 0.5–1.1 mg/dl *Older adults:* Decreased	**In Chronic Kidney Disease** May increase by 0.5–1.0 mg/dL every 1–2 yr May be as high as 15–30 mg/dL *before* symptoms of CKD are present **In Acute Kidney Injury** Gradual increase of 1–2 mg/dL every 24–48 hr May increase 1–6 mg/dl in 1 wk or less	Consistently elevated levels indicate decreased renal function. Serum creatinine levels are used to evaluate effectiveness of dialysis treatments.
Blood urea nitrogen	10–20 mg/dL *Older adults:* May be slightly increased	**In Chronic Kidney Disease** May reach 180–200 mg/dL *before* symptoms develop **In Acute Kidney Injury** Often increases by 1–20 mg/dL at same pace as serum creatinine level May reach 80–100 mg/dL within 1 wk	Increases depend on protein intake and other factors. Rate of increase is controlled by limiting protein intake. This intervention is believed to decrease the rate of onset of systemic symptoms, such as anorexia, nausea, and vomiting. Elevations have multiple causes, including diminished renal function, excessive protein intake, sepsis, GI bleeding, dehydration, and tissue catabolism.
Electrolyte Studies			
Serum sodium	136–145 mEq/L; 136–145 mmol/L (SI units)	Normal or decreased	Clients with renal disease retain sodium. With associated water retention, serum sodium levels remain normal. With excessive water retention, serum sodium levels are decreased due to hemodilution. Assess client for evidence of fluid volume excess: edema, weight increase, or elevation of diastolic blood pressure. Limit fluid intake as directed. Avoid excessive sodium intake. Monitor for signs of hypernatremia: dry skin, excessive thirst, dry mucous membranes, elevated body temperature, and flushed skin. Client may need diuretics or dialysis.

Continued

TABLE 28.3 Laboratory Profile for Renal Disease—cont'd

Test	Normal Range for Adults	Values in Renal Disease	Comments
Serum potassium	3.5–5.0 mmol/L (SI units)	Increased	Advise client to avoid salt substitutes and to limit potassium-containing foods. Monitor for rapidly increasing serum potassium levels in AKI. ECG changes occur with serum potassium levels ≥6.5. Monitor for signs of hyperkalemia: dizziness, weakness, cardiac irregularities, muscle cramps, diarrhea, and nausea. May require administration of sodium polystyrene sulfonate (Kayexalate) or other treatment.
Serum phosphorus (phosphate)	3.0–4.5 mg/dL, 0.97–1.45 mmol/L (SI units) *Older adults:* May be slightly decreased	Increased	Short-term increases have potential to cause rapid decrease in serum calcium level and cardiac rhythm disturbances. Long-term increases demineralize bones of calcium and enhance fracture potential. Phosphate-binding medications help control hyperphosphatemia and prevent calcium depletion from bones.
Serum calcium	Total calcium: 9.0–10.5 mg/dL; 2.25–2.75 mmol/L (SI units) Ionized calcium: 4.5–5.6 mg/dL; 1.05–1.3 mmol/L (SI units) *Older adults:* Slightly decreased	Decreased	Decreases in AKI may necessitate replacement. Decreases in CRF may only be slight and may or may not necessitate replacement. As serum phosphate level increases, serum calcium level decreases. Chronic calcium deficiency leads to renal osteodystrophy. Control of phosphate excess is usually essential before calcium replacement is initiated. Monitor for manifestations of hypocalcemia: abdominal cramps, hyperactive reflexes, tingling fingertips, and spasms in feet and wrists.
Serum magnesium	1.3–2.1 mEq/L; 0.65–1.05 mmol/L (SI units)	Increased	Advise patient to avoid compounds containing magnesium (e.g., laxatives).
Serum bicarbonate	23–30 mEq/L (venous); 23–30 mmol/L (SI units)	Decreased	Replace bicarbonate. Monitor respiratory rate and depth. Monitor for decreased orientation.
Arterial blood pH	7.35–7.45	Decreased (in metabolic acidosis) or normal	Respiratory system attempts to compensate by hyperventilation (increased rate and depth of respiration). Values are within normal range if blood buffers and lungs can compensate. Monitor breathing rate and depth. Monitor level of consciousness.
Arterial blood bicarbonate (HCO_3^-)	21–38 mEq/L	Decreased	Provide replacement oral, IV, or by hemodialysis or peritoneal dialysis.
Arterial blood $Paco_2$	35–45 mm Hg	Decreased	Monitor for respiratory fatigue (client breathes more rapidly and deeply to "blow off" carbon dioxide).
Other Blood Studies			
Hemoglobin	*Female:* 12–16 g/dL, 7.4–9.9 mmol/L (SI units) *Male:* 14–18 g/dL, 8.7–11.2 mmol/L (SI units) *Older adults:* Slightly decreased	Decreased	Decreased levels indicate anemia. Monitor for pallor, weakness, lethargy, dizziness, possible shortness of breath, and activity intolerance.
Hematocrit	*Female:* 37%–47% *Male:* 42%–52% *Older adults:* May be slightly decreased	Decreased to 20%	Same as for hemoglobin. With erythropoietin therapy, anemia improves.

TABLE 28.3 Laboratory Profile for Renal Disease—cont'd

Test	Normal Range for Adults	Values in Renal Disease	Comments
Urinalysis*			
Specific gravity	Usually 1.010–1.025 Possible range: 1.005–1.030	Usually decreased and fixed	Reflects inability of tubules to produce concentrated or diluted urine in response to changes in plasma osmolarity. Monitor for fluid volume deficit or excess.
pH	Average: 5.5–6 Possible range: 4.6–8	May be fixed; pH does not change with dietary changes	Collect a freshly voided specimen for testing.
Glucose	None or <15 mg/dL Usually detectable in urine of nondiabetic clients when blood level is 160–180 mg/dL	Increased	Renal threshold is often increased; therefore blood glucose level may be >160–180 mg/dL before glucose is detectable in urine. Monitor *blood* glucose levels.
Protein	0–8 mg/dL	Increased when there is glomerular damage or disease	Increases may be an incidental and benign finding. Transient increases occur with extreme exercise, fever, stress, or infection. Persistent proteinuria requires 24-hr collection for determination of total quantity excreted. Persistent proteinuria may indicate a serious renal problem. Instruct client about need for follow-up. Instruct client in correct procedure for collection of 24-hr specimen.
Occult blood	No RBCs or occasionally 2 or 3 RBCs per HPF No hemoglobin	More than 2 or 3 RBCs per HPF Detectable hemoglobin	Hemoglobin is detectable when hemolysis of RBCs has occurred. Intact RBCs are detectable only with microscopic examination. Collect a freshly voided specimen for testing.
WBCs	0–5 per HPF	Increased in urinary tract infection	Often indicates need for urine culture.
Bacteria	Fewer than 1000 colonies/ml	Increased in presence of infection, with or without an increase in WBCs	Obtain urine culture.
Casts	None or a few; composed of RBCs, WBCs, protein, or tubular cell casts such as hyaline	Casts present	Casts may be a benign occurrence or may signify that some renal injury or disease is present. Collect a freshly voided specimen for direct microscopic examination.
Creatinine clearance	*Male:* 107–109 ml/min *Female:* 87–107 ml/min *Older adults:* Progressively decreased with advancing age	Decreased	Change reflects decreases in GFR. Creatinine clearance is determined from a 24-hr urine collection and a serum creatinine value.

Adapted from Ignatavicius DD, Workman ML: *Medical-surgical nursing: critical thinking for collaborative care,* ed 5, Philadelphia, 2006, Saunders, pp 1733–1735.

AKI, Acute kidney injury; *CKD,* chronic kidney disease; *CRF,* chronic renal failure; *ECG,* electrocardiogram; *GFR,* glomerular filtration rate; *GI,* gastrointestinal; *HPF,* high-power field; *IV,* intravenous; *RBCs,* red blood cells; *SI,* Système International d'Unités; *WBCs,* white blood cells.

*Urine may become cloudy with heavy sediment. Urine output and appearance vary, depending on remaining renal function.

these new biomarkers currently being tested are neutrophil gelatinase–associated lipocalin, cystatin C, and tissue inhibitor metalloproteinase-2. As with other aspects of the clinical presentation, laboratory findings are affected by the phase of AKI. Although the naming of the phases of AKI varies among resources, the clinical progression is consistent.

Prodromal Phase

Patients during this phase will typically have a normal or declining urine output. Serum blood urea nitrogen (BUN) and creatinine levels begin to rise. The insult to the kidney has occurred, and the duration of this phase will vary depending on the cause of the injury, the amount of the toxin ingested, or the duration and severity of the hypotension.

Oliguric Phase

Most patients with AKI develop oliguria, though some will remain nonoliguric. Those with severe injury often become anuric. Renal function is monitored according to fluctuations in the serum creatinine level and the calculated GFR that is derived from it. As urine output drops, metabolic waste products are retained and fluid and electrolyte balance is disrupted. These alterations are reflected in the physical assessment and laboratory data. Because it cannot be excreted, body water is retained. This hypervolemia results in the development of edema and hypertension. Other signs and symptoms of fluid volume overload include distended neck veins (jugular venous distention), weight gain, crackles, and possibly

TABLE 28.4 Pathophysiology of Manifestations of Uremic Syndrome

System	Manifestation	Pathophysiologic Basis
Central nervous system	Uremic encephalopathy • Disorientation • Lethargy • Coma	Brain cells shrink because of osmotic gradient
Peripheral nervous system	Motor neuropathies • Weakness Sensory neuropathies • Numbness • Tingling	Toxin interference with nerve transmission
Cardiovascular system	Bleeding	Decreased platelet function and coagulation factor production
	Anemia	Decreased RBC life span; blood loss
	Decreased cardiac output	Negative inotropic effect
	Ischemic syndromes	Escalated coronary heart disease
	Pericarditis*	Uremic crystals deposited in pericardium
Immune system	Infections	Decreased immune cell production and immune response
Gastrointestinal system	Anorexia	Direct effect of toxins
	Nausea and vomiting	
Integumentary system	Impaired healing*	Decreased collagen production
	Pruritus; dermatitis*	Deposition of uremic crystals on skin
	Uremic frost*	
Acid-base balance	Metabolic acidosis	Accumulation of acidic metabolic wastes

*More likely to occur with chronic kidney injury.
RBC, Red blood cell.

additional manifestations of heart failure. Normally, as renal perfusion is reduced, the kidney responds by increasing the reabsorption of sodium and reducing its elimination. In AKI, this mechanism is impaired and sodium is lost in the reduced volume of urine that is produced. Tubular casts are present in the urine, and white blood cells, red blood cells (RBCs), and protein are also likely to appear. Other electrolytes are typically retained in the blood (potassium, magnesium, and phosphorus) and their levels in the urine are decreased (see Table 28.3). Of all the potential electrolyte problems, hyperkalemia is of greatest concern because the normal range is narrow (3.5 to 5 mEq/L) and levels of less than double the normal values can be fatal. Metabolic acidosis develops as a result of the kidney's decreased ability to excrete hydrogen ions (acid load). The hydrogen ions shift into the cells in exchange for intracellular potassium. Thus hyperkalemia and metabolic acidosis often occur together. Changes in the levels of other electrolytes may not be seen unless the oliguric phase is prolonged. Anorexia, nausea, vomiting, weakness, seizures, acidosis, confusion, and coma are also possible. Although these manifestations are classically much more severe in ESRD, there is no identifiable point at which uremia presents, so some symptoms may be present in AKI. A GFR that is only slightly less than 50% of the normal for age may produce mild manifestations.

As the GFR drops, organic metabolic waste products are retained. Urea represents the largest amount of these wastes. Other uremic solutes involved have been found to be responsible for many of the signs and symptoms that collectively are called *uremic syndrome* (Table 28.4). This phase typically lasts 10 to 14 days, but can range from 1 day to 8 weeks. Development of uremia in AKI necessitates temporarily replacing renal function with dialysis. Nonoliguric patients have lower mortality and morbidity and less need for dialysis.

Postoliguric Phase

Termination of the oliguric phase is marked by the beginning of renal recovery. The urine output gradually returns to normal. Unfortunately, not all patients recover. In about 5% of cases, AKI is not reversible. Tubular dysfunction may continue and is manifested by polyuria and sodium wasting, as well as depletion of electrolytes (especially potassium) previously retained. Typically, urine output is more than 1 L per day, and may be as much as 4 to 5 L, causing a fluid volume deficit if fluids are not rapidly replaced. The elevated BUN level contributes to an osmotic diuresis, as the large molecules are filtered and draw more water into the tubules. Renal perfusion begins to increase, slowly advancing filtration ability and tubular function.

If full recovery of renal function does occur, it may take only 1 week, but more often requires as much as 1 year. Full recovery is indicated when the serum creatinine level returns to within the normal range. This variation in recovery time is due to the diversity of AKI causes and degree of damage, as well as the quality of interventions provided and patients' concomitant conditions. Even after recovery, progressive loss of renal function after AKI is a frequent finding, especially in pediatric cases. The loss of renal function after AKI in childhood has been found to result in progressive deterioration by adolescence or early adulthood.

KEY POINTS

- Acute kidney injury (AKI) is an abrupt reduction in renal function producing an accumulation of waste materials in the blood. Oliguria is usually present. AKI is classified into three types according to the site of disruption: prerenal, postrenal, and intrinsic. Distinction between the types of AKI is necessary to determine appropriate therapy.
- Prerenal kidney injury is due to conditions that impair renal blood flow, such as hypovolemia, hypotension, cardiac failure, and renal artery obstruction. It is characterized by clinical manifestations of a low glomerular filtration rate (GFR), usually including oliguria, high urine specific gravity and osmolality, and low urinary sodium concentration. Signs and symptoms of fluid volume overload are present. Prolonged prerenal kidney injury results in intrinsic kidney injury.

- Postrenal kidney injury is due to obstruction within the urinary collecting system distal to the kidney. Obstruction results in elevated pressure in Bowman capsule, which impedes glomerular filtration. Clinical findings vary, based on the duration of the obstruction. Prolonged postrenal kidney injury results in intrinsic kidney injury.
- Intrinsic failure is due to a primary dysfunction of the nephrons. Although it may be due to glomerular, vascular, or interstitial etiologies, most often the problem is within the renal tubules, resulting in ATN. ATN may occur with nephrotoxic or ischemic insults. Clinical manifestations depend on the phase of acute tubular necrosis (ATN).
- AKI has three characteristic phases. The first phase, prodromal, is characterized by normal or declining urine output and varies in duration, depending on the causative factors. The oliguric phase may last up to 8 weeks with a usual urine output of 50 to 400 mL/day. Retention of nitrogenous wastes and certain electrolytes is likely. Fluid volume excess is expected. During the postoliguric phase, diuresis occurs, tubular function remains impaired, and azotemia continues. Fluid volume deficit is characteristic until the kidneys gradually recover. Recovery can last up to 12 months and is characterized by gradual normalization of serum creatinine and blood urea nitrogen (BUN) levels. Often a degree of renal insufficiency persists.
- AKI results in characteristic alterations in laboratory tests of the blood and urine. Renal function is monitored by serum creatinine level and calculated GFR. The retention of metabolic wastes (azotemia/uremia), which is monitored by the BUN level, produces widespread systemic effects (uremic syndrome).

CHRONIC KIDNEY DISEASE

Chronic renal failure is the outcome of the progressive and irrevocable loss of nephrons. The normal resiliency of the kidney means that more than 75% of the total number of nephrons must be lost before clinical manifestations appear. What is termed *chronic renal failure* is the final outcome of CKD. More commonly, this point is called *end-stage renal disease* (ESRD) in acknowledgment of the progressive process of deterioration of renal function. At some point in the CKD process, renal disease will proceed to ESRD progressively, irreversibly, and to the point where renal dialysis or transplantation is required for survival.

CKD is a health crisis worldwide. Ten percent of the population worldwide is affected by CKD. Over 2 million people worldwide currently receive treatment with dialysis or a kidney transplant to stay alive, and of those 2 million, the majority are treated in the United States, Japan, Germany, Brazil, and Italy. Because of the rapidly escalating incidence of the two primary risk factors, diabetes and hypertension, the number of people afflicted with CKD is likely to increase in the years to come.

Risk Factors

CKD is defined as either decreased kidney function or kidney damage of 3 months' or more duration based on blood tests, urinalysis, and imaging studies. Alternatively, it may be defined as a GFR <60 mL/min/1.73 m^2 for 3 months, with or without indication of damage to the kidney. CKD is a syndrome in that it can occur as a complication of many other conditions, such as diabetes and hypertension. This means that the etiologies and/or risk factors of these conditions are also risk factors for the development of CKD. Although some of the conditions and risk factors have been identified for many years, new research findings have added to the list and altered perceptions of their relative importance (Box 28.2).

Diabetes and hypertension account for more than 50% of the individuals in the United States with ESRD, followed by glomerulonephritis and cystic kidney disease. Type 2 diabetes is projected to double in incidence worldwide within the next 25 years. Diabetic nephropathy is a major complication of that disease. Of those who develop diabetic nephropathy, about 30% ultimately progress to ESRD, whereas the remainder most often die of cardiovascular disease before reaching that point.

Hypertension is another major risk factor for CKD, and it also develops as a complication of it. The majority of patients (70% to 80%) with CKD have hypertension, climbing to 90% in those treated with dialysis. The aging of the population in the United States is predicted to escalate the incidence of systolic hypertension and resulting renal disease. Some of the risk factors for CKD are not amenable to modification. These include a family history of CKD, a history of exposure to toxins (heavy metals, tobacco smoke, radiopaque dyes), age greater than 65 years, and ethnicity (non-Hispanic black, non-Hispanic white, and Mexican American). It has been recommended that high-risk groups be screened for CKD (those with family histories, diabetes mellitus, hypertension, recurrent urinary tract infections, urinary obstructions, or other systemic conditions affecting the kidney) and that screening be extended to those without any other identified risks other than being older than age 65. Research is under way to identify genes associated

BOX 28.2 Risk Factors for Chronic Kidney Disease

- Acute tubular necrosis (not progressing beyond the oliguric stage)
- Developmental/congenital conditions
  - Renal agenesis
  - Aplastic kidneys
  - Renal hypoplasia
  - Ectopic/displaced kidneys
  - Fused kidneys
- Cystic disorders
  - Polycystic kidney disease
  - Medullary cystic disease
- Neoplasms
  - Benign tumors of the kidney
  - Malignant tumors of the kidney (including Wilms tumor)
- Infections
  - Recurrent pyelonephritis
  - Renal tuberculosis
- Glomerulonephritis
- Systemic conditions
  - Diabetes mellitus*
  - Diabetes insipidus
  - Hypertension*
  - Hyperparathyroidism
  - Liver failure/cirrhosis
  - Gout
  - Amyloidosis
  - Scleroderma
  - Goodpasture syndrome
  - Systemic lupus erythematosus (produces glomerulonephritis)
- Other
  - Genetics
  - Increasing age
  - Race (blacks)
  - Overweight/obesity
  - Dyslipidemia
  - Family history of cardiovascular disease
  - Smoking

*Most common risk factors.

with CKD and diabetic nephropathy, as well as biomarkers to assess for risk and likelihood of a positive response to treatment.

Pathophysiology of Progression of Chronic Kidney Disease

Regardless of the origin of CKD, the decline in renal function is progressive and irreversible. The damaged nephrons are unable to function properly. The remaining nephrons initially compensate by enlarging and increasing their clearance capacity. Through this process, renal function remains relatively normal until 75% to 80% of the nephrons are damaged and nonfunctional. Over the course of several years, these compensatory changes may cause further injury to the remaining nephrons. Although the cause of this additional injury is unknown, some investigators believe that the increased workload of the remaining glomeruli and small arterioles may cause sclerosis of these vessels, leading to further decline in kidney function and ultimately ESRD. Progression of CKD is monitored by measuring changes in GFR, determining the presence and degree of proteinuria, examining urinary sediment for white or RBCs, measuring serum creatinine levels, and performing imaging studies with renal ultrasonography to document kidney size.

Stages of Chronic Kidney Disease

The progression of CKD is determined by monitoring GFRs. A five-stage system depicts severity of disease, with decreasing GFRs reflecting a higher stage of progression. The recommendation is that diagnosis of CKD not be based on a specific pathology (e.g., diabetes mellitus), but rather on the presence of damage to the kidney and the level of function according to the GFR measurement as shown in Table 28.5. Screening for the complications of CKD is recommended to begin in stage 3.

In stage 1 of the GFR staging system shown in Table 28.5, patients' BPs and laboratory values are usually normal and they are asymptomatic, though they have some form of kidney disease. An initial reduction in GFR, with or without documented kidney damage, depicts stage 2. This would include, for example, a diabetic patient with proteinuria or an older person with the predictable age-related diminishment of renal function. Again, patients may be asymptomatic without laboratory value changes. The aims in stages 1 and 2 are to identify individuals at risk for progressive renal disease and reduce those associated risks.

Beginning with stage 3, patients are classified as having CKD, according to GFR, regardless of the existence of kidney damage. Although they may be otherwise asymptomatic in stage 3, hypertension is nearly always noted. At stage 4, diagnosis is made because manifestations are usually very apparent with the significant decline in renal function. Planning for ESRD should occur. By stage 5, manifestations of ESRD are present and renal function is so impaired that dialysis or transplantation is required.

The staging system shown in Table 28.6 is useful because the terms *renal insufficiency* and *decreased renal reserve* are commonly used in clinical practice. Decreased renal reserve is not associated with signs or symptoms of renal failure, largely because the remaining nephrons accommodate the additional workload. Electrolyte and fluid levels are maintained within normal limits. Although not abnormally elevated, creatinine levels are usually at the high end of the normal range—a common finding in the elderly population. Interventions for those at risk for CKD (see Box 28.2) should be initiated to help slow disease progression. At this point, kidney function is already impaired and significant deterioration is possible if the kidney is stressed. There may be small amounts of protein in the urine. The stage of renal insufficiency is reached with further nephron damage. Although electrolyte levels remain within normal limits with the GFR decline, the metabolic wastes (creatinine and urea) are retained at levels proportional to nephron loss. Impairment of the kidney's ability to concentrate the filtrate often results in increased urinary output (polyuria), often leading to nocturia. ESRD develops when more than 90% of the kidneys' nephrons have been destroyed. At this point, the patient typically demonstrates the sequelae and complications of renal failure, seen as laboratory alterations and signs and symptoms associated with the inability of the kidneys to fulfill their multiple roles within the body. Without interventions, when the remaining nephrons number less than 5% to 10% of normal, death is inevitable.

Complications of Chronic Kidney Disease

The consequences of CKD are affected by the intake of food and water and the degree to which renal function is impaired. Some of the complications of CKD begin to appear when the GFR falls below 60 ml/min/1.73 m^2 (stage 3 of the GFR-based classification). Because of the widespread impact of the kidneys on nearly every part of the body, the effects of their failure are seen in every body organ system. Complications

TABLE 28.5 Stages of Chronic Kidney Disease According to Glomerular Filtration Rate*

Stage	Description
Stage 1	Kidney damage with normal or increased GFR GFR >90 mL/min/1.73 m^2
Stage 2	Mildly decreased GFR GFR 60–89 mL/min/1.73 m^2
Stage 3	Moderately decreased GFR GFR 30–59 mL/min/1.73 m^2
Stage 4	Severely decreased GFR GFR 15–29 mL/min/1.73 m^2
Stage 5	End-stage kidney disease GFR <15 mL/min/1.73 m^2 (or dialysis)

*From National Kidney Foundation: Kidney disease outcomes quality initiative, 2002.
GFR, Glomerular filtration rate.

TABLE 28.6 Stages of Chronic Kidney Disease According to Nephron Loss and Clinical Presentation

Stage	% Nephron Loss	Clinical Presentation
Decreased renal reserve	<75%	No signs or symptoms BUN and creatinine normal May not be diagnosed
Renal insufficiency	75–90%	Polyuria, nocturia Slight elevation in BUN and creatinine May be controlled by diet and medication
End-stage renal disease	>90%	Azotemia/uremia Fluid and electrolyte abnormalities Osteodystrophy Anemia Dialysis or transplantation essential

BUN, Blood urea nitrogen.

affect the gastrointestinal, neurologic, musculoskeletal, dermatologic, cardiovascular, and endocrine systems. Immune function, acid–base regulation, and the coagulation cascade are affected as well. In many cases, complications are interrelated, with one contributing to or exacerbating the development of others. Laboratory alterations become significant as CKD progresses; typical values are shown in Table 28.3. Major complications of CKD are discussed in the following sections.

Hypertension and Cardiovascular Disease

Cardiovascular disease and CKD have many risk factors in common: increasing age, black race, hypertension, diabetes mellitus, smoking, decreased GFR, proteinuria, obesity, and RAAS overactivity. If hypertension was not a precursor to CKD, it inevitably develops with decreasing renal function, appearing in stage 2 or 3. The excess fluid volume and escalated atherosclerotic process, along with heightened RAAS activity and evidence of increased SNS activity, are all involved in the development of hypertension in CKD. Most people with CKD do not die of kidney failure, but rather from the results of cardiovascular disease, which is both a risk factor for and a complication of CKD. The incidence of cardiovascular morbidity and mortality is markedly increased in CKD. In the elderly, death from cardiovascular disease is more common than progression to renal replacement therapy. Although the risk of cardiovascular disease is greatest for those with end-stage disease, it has been found to be an issue at all stages of CKD. A clear link exists between deterioration of kidney function and development of cardiovascular disease. Multiple factors have been identified that are conducive to this, including, but not limited to, dyslipidemias, anemia, electrolyte-induced dysrhythmias, hypertension, endothelial dysfunction (including inflammation, alterations in nitric oxide activity, oxidative stress), proteinuria, premature atherosclerosis, activity limitations, and volume overload.

Uremic Syndrome

As renal function declines, retention of metabolic wastes increases, creating a toxic environment within the body. Uremia produces signs and symptoms in many body systems (see Table 28.4). Compared with the presentation in AKI, uremic syndrome in CKD is far more severe and its effects are more widespread, even though the serum levels often must be much higher for manifestations to be apparent.

Metabolic Acidosis

Normally, the human body produces 50 to 80 more millimoles of acids than bases during daily metabolism. Metabolic acidosis occurs with the retention of acidic waste products of metabolism as part of uremia. It is also produced in association with hyperkalemia, when potassium ions in the blood are exchanged for intracellular hydrogen ions, lowering the pH of the blood (see Chapter 25). The kidneys lose their ability to secrete hydrogen ions or to produce bicarbonate, and with the limited capacity of the other buffers, pH can fall precipitously; coma develops and death will occur if the pH drops below 6.8. Metabolic acidosis depresses cardiac as well as central nervous system function. The respiratory system attempts to compensate for metabolic acidosis by increasing the rate and depth of respirations.

Electrolyte Imbalances

The loss of renal mechanisms involved in electrolyte balance result in the retention of potassium, phosphorus, and magnesium in the blood. Each of these imbalances is associated with specific manifestations (see Chapter 24). Hyperkalemia is of special concern, as discussed with AKI, because it can be responsible for fatal cardiac dysrhythmias. The inability to eliminate phosphorus and the loss of the renal mechanisms involved in maintaining calcium balance result in mineral and bone disorders, which are discussed next.

Mineral and Bone Disorders

The inability of the kidney to excrete phosphorus in CKD results in elevated phosphorous levels in the blood. Because phosphorus and calcium maintain a reciprocal relationship in the body, as hyperphosphatemia develops in CKD, hypocalcemia likewise develops. The body's system of maintaining calcium balance is sophisticated and involves the activation of parathyroid hormone (PTH) when the calcium level is low. When this occurs because of CKD, the result is secondary hyperparathyroidism. The normal response to PTH release is the increase in serum calcium levels by (1) increasing reabsorption by the kidney; (2) increasing absorption of dietary calcium in the presence of vitamin D, which must be activated by the kidney; and (3) mobilizing calcium release from bone. Vascular and soft tissue calcifications as well as osteoporosis are the end results of the prolonged elevation of PTH level and altered bone and mineral metabolism. The resulting bone pain, deformities, and fractures are significant long-term complications of CKD. Additionally, abnormal PTH, calcium, and phosphorous levels are reported to have been identified as independent risk factors of cardiac-caused mortality during dialysis.

Malnutrition

Protein-energy wasting (PEW) is the loss of muscle and visceral protein stores and is a common finding in advanced CKD. Decreased nutrient intake is due to the anorexia of uremic syndrome, changes in taste sensation, and the depression of chronic illness, as well as the dietary limitations imposed by disease. Additionally, a negative-nitrogen balance exists because of escalated protein catabolism and decreased protein synthesis. Medications and concomitant diseases such as diabetes may cause nausea, vomiting, and slowed gastric emptying. The best serum marker of PEW is albumin level. Hypoalbuminemia in dialysis patients has a strong association with increased mortality and morbidity. However, caution must be used in interpreting these levels because the serum level may be affected by the inflammatory response. Malnutrition has a negative impact on CKD prognosis.

Anemia

The development of anemia is an anticipated problem as CKD progresses. The production of RBCs by the bone marrow depends on numerous cofactors; perhaps the most significant of these is erythropoietin, which is produced by the kidneys. Erythropoiesis, with all essential ingredients, requires about 5 days. Lacking erythropoietin, fewer RBCs are produced and anemia is a persistent problem. The problem is often further escalated by malnutrition, due to nutritional deficits of iron, folate, and vitamin B12. Chronic inflammation and elevations of PTH level suppress bone marrow. At the same time, the uremia associated with CKD produces a toxic environment for RBCs, reducing their normal life expectancy of 120 days. Blacks appear to have both a greater prevalence of anemia and more severe reductions in RBC counts and hemoglobin concentrations. By stage 5 and the initiation of dialysis, approximately 66% of patients have hemoglobin levels <11 g/dL. The cardiovascular compensatory efforts for the chronic anemia of CKD can lead to myocardial dilatation, left ventricular hypertrophy, and remodeling. The combination of worsening CKD, anemia, and heart failure is referred to as *cardiorenal anemia syndrome* and becomes a chain of escalating pathologies.

Pain

The causes of pain in kidney disease are multifactorial. Surgery may be needed to create access for dialysis or to biopsy a kidney. Dialysis itself can be a painful experience with frequent needle sticks and accumulation of uremic toxins. The primary cause of kidney disease, such as cystic kidney or diabetes (due to neuropathies or ulcerations), may result in

pain. Finally, comorbidities such as gout, bone disease, and peripheral vascular disease may also contribute to pain in the person with CKD. The challenge lies in finding a balance between pain relief and the right analgesic and dose to avoid toxicity and drug-induced nephropathy.

Depression

The prevalence of depression in patients with CKD is reported to be between 20% and 30%. Reasons for depression stem from comorbid conditions, as well as the process of kidney disease itself. Many patients with renal disease experience a high rate of cardiovascular events, which are also associated with higher rates of depression. In addition, the disruption of social interactions and relationships, possibly attributable to dialysis and fatigue, contributes to depressive symptoms. In patients with CKD, as in the general population, depression may lead to a decrease in quality of life, functional impairment, and sexual dysfunction.

KEY POINTS

- Chronic kidney disease (CKD) is characterized by a gradual, irreversible loss of functional nephrons. The two most common causes are diabetes mellitus and hypertension, followed by recurrent pyelonephritis, glomerulonephritis, and polycystic kidney disease.
- Progression of CKD is monitored by a staging system based on increasing severity of disease. There are five stages of CKD progression, and with each higher stage, the glomerular filtration rate (GFR) and kidney function decline. In stage 1, kidney function may be normal, though some disease exists. The patient is asymptomatic. The focus in stages 1 and 2 is minimizing risk factors. By stage 3 symptoms may be starting to appear and treatment may be needed. In stage 4, planning for dialysis or transplant should begin, and in stage 5 renal replacement therapy is needed or death will ensue. Complications of CKD include hypertension and increased cardiovascular risks, uremic syndrome, metabolic acidosis, electrolyte disturbances, bone and mineral disorders, malnutrition, anemia, pain, and depression.

CLINICAL MANAGEMENT

Clinical management of AKI and CKD requires a multidisciplinary approach. Collaboration among the nurse, physician, clinical pharmacist, and dietitian is essential to attain optimal patient outcomes. Prevention of the development of AKI is the goal. With CKD, prevention of the myriad of causes is certainly desirable; however, prevention also focuses on retarding the inevitable progression of the disease and reducing cardiovascular risk factors.

Acute Kidney Injury

AKI is often entirely preventable. Prevention is highly dependent on recognizing patients who are at risk for the development of AKI and establishing prophylactic interventions to decrease this risk.

Etiologies and risk factors for prerenal kidney injury should be identified and, whenever possible, treated swiftly. Early consultation with a nephrologist is recommended. Hypotension attributable to hypovolemia should be addressed; medications that might be contributing (antihypertensives, opioids) should be decreased or discontinued. Nutrition should be supported, and indwelling catheters and other invasive equipment should be removed as soon as possible to decrease the risk of infection. Development of infection should be monitored and treated if it develops. Maintaining adequate circulating volumes and supporting cardiac function, especially in those at risk for prerenal kidney injury, are the primary foci for prevention.

Should prerenal oliguria develop, efforts should be initiated to enhance renal perfusion before ATN can occur. Aggressive management of conditions such as glomerulonephritis can prevent progression to intrinsic kidney injury. Prevention of intrinsic kidney injury caused by chemicals begins with an awareness of the most likely harmful agents; these are listed in Box 28.3. Whenever possible their use should be limited, and if these agents must be prescribed, their serum levels may need to be monitored. Single doses of aminoglycosides, rather than multiple doses, and liquid formulations of amphotericin B appear to be less nephrotoxic. Of all the chemicals listed in Box 28.3, the most common culprit is radiocontrast media.

Prevention of contrast-induced AKI is aimed at initiating intravenous volume expansion with saline both before and after the contrast administration, administering the antioxidant *N*-acetylcysteine (Mucomyst) to patients at risk, measuring the serum creatinine concentration of patients with suspected renal dysfunction, and using either low-osmolar or isoosmolar contrast media for all patients with renal insufficiency. It is important to note that the evidence to support the effectiveness of Mucomyst in preventing CI-AKI has demonstrated mixed results. More research is needed. Postrenal etiologies should be avoided whenever possible, but if they occur they should be rapidly identified and corrected. Postrenal kidney injury is the least common and most readily identifiable; promptly addressing it will prevent the progression to ATN.

Chronic Kidney Disease

Clinical management of CKD is complex and requires a multidisciplinary approach. Treatment is directed at slowing the progression to ESRD and managing the complications that are inevitable. This means that patient education regarding the trajectory of the illness, management of modifiable risk factors, and clinical management of the disease are essential. Because death is often due to cardiovascular pathologies, management of these risk factors assumes significant importance. At a community level, screening for hypertension, diabetes, and CKD is essential to stem the tide of these interrelated chronic diseases.

BOX 28.3 Selected Chemicals Toxic to the Kidneys

Acetaminophen
Acyclovir (Zovirax)
Allopurinol
Aminoglycosides
Amphotericin B (Fungizone)
Angiotensin-converting enzyme inhibitors
Certain cytotoxic chemotherapeutic agents
Cocaine
Cyclosporine (Sandimmune)
Foscarnet (Foscavir)
Heavy metals
Hemoglobin; myoglobin
Heroin
Lithium
Nonsteroidal antiinflammatory drugs
Pentamidine (Pentam 300 and others)
Radiocontrast media*
Uric acid
Vasopressors (norepinephrine, high-dose dopamine)

From McKenry L, Tessier E, Hogan M: *Mosby's pharmacology in nursing,* ed 22, St Louis, MO, 2006, Mosby; Needham E: Management of acute renal failure, *Am Fam Physician* 72(9):1739–1746, 2005.
*Most common.

Underdiagnosis and undertreatment mediate the problem, and lack of awareness is the enemy of early diagnosis and treatment.

The speed with which CKD progresses varies significantly among individuals. Complications are typically evident by stage 3 or 4, and attention is directed to therapeutic interventions designed to minimize and treat these complications. Therapeutic and pharmacologic interventions are presented within the context of each complication listed in the following paragraphs. An in-depth presentation is beyond the scope of this text; discussion is simply an overview. Goals of therapeutic interventions for CKD are summarized in Box 28.4.

Hypertension and Cardiovascular Disease

Evaluation of risk factors for cardiovascular disease is an important part of intervention, because the risk of morbidity and mortality rises with the decline of GFR. Hypertension is associated with a more rapid progression of CKD; therefore control of BP is critical. Lowering BP reduces the risk of proteinuria and prevents the development of cardiovascular and cerebrovascular events. The BP goal for most CKD patients is <140/90 mm Hg. An ACEI or an AII receptor blocker is considered the treatment of choice to delay progression, followed by a thiazide or loop diuretic. ACEIs and AIIs have an additional benefit of reducing proteinuria by 40% to 45%. However, it should be noted that effectively lowering BP does not necessarily improve survival.

The benefits of using statins to treat dyslipidemia in patients with CKD have not been clearly established and remain controversial. Although a reduction in total and low-density lipoprotein cholesterol levels occurs and may be effective in decreasing cardiovascular events, there is no improvement in the GFR or the rate of CKD progression. There is some evidence to indicate that there is a benefit from use of statins in stages 1 to 3, but not once patients start dialysis. This is an area of ongoing research.

Metabolic Acidosis

Mild acidosis with a pH of 7.30 to 7.35 requires no therapy. Patients with chronic metabolic acidosis (<7.30) may be prescribed sodium bicarbonate. In a single-center randomized trial, 134 adult patients with stage 4 CKD were treated with bicarbonate supplements. After 2 years of treatment, 6.5% of patients in the sodium bicarbonate group required dialysis compared with 33% in the control group. Replication of this study is needed.

Fluid and Electrolyte Imbalances

Fluid restrictions are implemented when the sodium level drops below 135 mmol/L. Patients with edema, heart failure, or hypertension may need a 2 g/day sodium restriction. Mild hyperkalemia (<6 mmol/L) can be treated by reducing potassium intake and correcting metabolic acidosis. Potassium levels >6 mmol/L require more urgent treatment, such as an IV infusion of calcium gluconate, 5% dextrose in water (D_5W), and insulin or oral or rectal administration of sodium polystyrene sulfonate.

BOX 28.4 Therapeutic Goals in Chronic Kidney Disease

- Maintain volume status.
- Prevent and treat acid–base and electrolyte disturbances.
- Prevent and treat uremia.
- Support nutritional needs.
- Prevent and treat infection.
- Prevent and treat anemia.
- Improve quality of life.
- Lower mortality and morbidity rates.
- Control pain.

Bone and Mineral Disorders

Hyperphosphatemia is addressed first nutritionally and later through drug management. Serum levels of PTH, calcium, and phosphorus should be monitored in patients with a GFR less than 60 mL/min/1.73 m^2. Calcium carbonate and calcium acetate effectively bind phosphorus, correct hypocalcemia, and are inexpensive, though they are associated with hypercalcemia and vascular calcifications. Lanthanum carbonate has potent phosphorous-binding ability and is generally well tolerated. Vitamin D therapies such as calcitriol, paricalcitol, and ergocalciferol are indicated for use in vitamin D deficiencies and to suppress PTH levels. Calcimimetics, such as cinacalcet, have also been shown to decrease secretion of PTH. Target ranges for both calcium and phosphorous levels are shown in Table 28.7.

Malnutrition

Nutrition plays a large role in the treatment of CKD. Advanced CKD produces gastrointestinal manifestations, anorexia, nausea, and changes in the sense of taste, which further complicate the necessary modifications in diet. Caloric requirements are increased. Sufficient carbohydrate and fat are needed to meet energy requirements. Aspects of nutritional management of CKD include limiting intake of dietary phosphorus, protein, sodium, potassium, and water, depending on laboratory values and other clinical manifestations. Avoiding malnutrition, preventing anemia, and countering disease- and drug-induced constipation are other aspects of nutritional management. Lists of foods high in sodium, potassium, and protein should be provided; patients can then be encouraged to identify their favorite foods and choose smaller portions or eat them less often. Additionally, diet-related risk factors for cardiovascular disease must be considered. This can all seem overwhelming to patients, so involvement of a dietitian and thorough education considering concomitant conditions, personal likes and dislikes, eating habits, and financial resources is essential.

Anemia

Erythropoiesis-stimulating agents, such as epoetin alfa and darbepoetin alfa, have dramatically improved the quality of life for patients with CKD, but have had little effect in reducing overall cardiovascular mortality or the rate of GFR loss. Target hemoglobin (Hgb) levels have been the subject of much debate. The National Kidney Foundation's *Kidney Disease Outcome Quality Initiative (KDOQI) Guidelines* recommend Hgb levels between 11 and 12 g/dL. In addition to pharmacologic interventions, strategies for reducing fatigue, such as balancing rest and activity throughout the day, should be employed.

Pain

Pain in patients with CKD may be acute or chronic. Choice of medication and dosage need to consider the patient's age, comorbidities, and degree

TABLE 28.7 Target Ranges for Calcium and Phosphorus in CKD

Stage	Calcium	Phosphorus
3	8.4–9.5 mg/dL	2.7–4.6 mg/dL
4	8.4–9.5 mg/dL	2.7–4.6 mg/dL
5	8.4–9.5 mg/dL	3.5–5.5 mg/dL

Data from Legg V: Complications of chronic kidney disease: a close look at renal osteodystrophy, nutritional disturbances, and inflammation, *Am J Nurs* 105(6):40-49, 2005.

of kidney failure; the pathway of elimination of the medication (i.e., through the liver, kidney, or dialysis); and the risk of drug-induced nephropathy. Nonpharmacologic methods of pain relief should also be explored.

Depression

Studies have explored the use of pharmacologic and nonpharmacologic approaches to treat depressive symptoms in patients with CKD. The safety of antidepressants in patients with decreased renal function is a significant concern. Psychotherapy, exercise therapy, cognitive behavioral therapy, and music therapy all have demonstrated varying degrees of success.

Acute-on-Chronic Kidney Disease

Acute-on-chronic kidney disease describes a scenario in which AKI occurs in someone with preexisting CKD. Incidence varies from 10% to greater than 30%, depending on the study population. Preexisting CKD is a strong risk factor for development of AKI (Table 28.8). The most common causes of AKI in someone with CKD include systemic infections, medications, dehydration, and urinary tract obstruction. Management is directed toward identifying and treating the underlying cause of the acute deterioration in order to prevent a possible irreversible drop in kidney function.

Dialysis

The procedures, advantages, disadvantages, and expected outcomes of dialysis should begin to be discussed when the patient is in stage 4 of CKD. When patients with CKD reach stage 5, dialysis is indicated. The primary reason for the initiation of dialysis is the development of uremia. Approximately two-thirds of the total body urea content is removed by each dialysis treatment. Dialysis may also be required if severe hyperkalemia is unresponsive to other interventions or in cases of severe volume overload. If these conditions develop during the oliguric phase of ATN, dialysis may be required temporarily. Dialysis supports all the treatment goals in Box 28.4 and is the only therapeutic option for those with ESRD who are unable to obtain a transplant. Dialysis may be accomplished by hemodialysis, peritoneal dialysis (PD), or continuous renal replacement therapy (CRRT).

Hemodialysis was first carried out on humans in 1924; before that time, patients with ESRD simply died. In 1972 the national ESRD program was signed into law, giving all Americans the right to treatment; ESRD is the sole recipient of this status. In hemodialysis, an artificial kidney serves as the dialyzing semipermeable membrane. The patient's blood passes through a bundle of hollow capillary tubules, and dialyzing fluid bathes these tubules. Solutes that are present in high concentration in the uremic blood (i.e., phosphate, urea, creatinine, potassium) diffuse across the dialyzing tubule membrane into the dialyzing fluid and are discarded. Excess water in the uremic blood is eliminated through osmosis across the membrane. Preferred access for hemodialysis is established by creating an arteriovenous (AV) fistula, most commonly in the arm. AV grafts and specialized central venous catheters are other means of access. Most patients with ESRD go through hemodialysis treatments three times per week, with each treatment lasting about 4 hours. Although a lifesaving intervention, dialysis treatments have complications, some of which are life threatening, and long-term morbidity remains quite high. Despite dialysis, cardiovascular disease remains the most common cause of death in ESRD patients.

In PD, the peritoneum serves as the dialyzing membrane. A dialysis catheter is surgically placed in the abdomen for access. During the treatment, the peritoneal cavity is slowly filled with dialysate through the catheter. Extra fluid and waste products are drawn out of the uremic blood and into the dialysate. There are two major types of PD: continuous ambulatory peritoneal dialysis (CAPD) and continuous cycling peritoneal dialysis (CCPD).

CAPD is carried out in the patient's home and without the use of machines. The patient instills about 2 quarts of dialysate into the peritoneum through the catheter. The dialysate remains there for 4 to 5 hours or longer before it is drained and discarded. This is called an *exchange*. While the dialysate resides in the peritoneal cavity, the patient has more freedom to continue his or her usual activities at work, at school, or at home. Peritonitis is a potentially serious complication.

CCPD can also be performed at home, but uses a special machine called a *cycler*. This is similar to CAPD except that a number of cycles (exchanges) occur. Each cycle usually lasts 1½ hours, and exchanges are done throughout the night while the patient sleeps.

CRRT is limited to in-hospital AKI patients. Continuous hemofiltration and hemodialysis procedures filter and dialyze the blood without interruption. CRRT removes fluid and wastes from patients who are not hemodyamically stable enough (e.g., shock, multiple organ system failure) to tolerate the larger quantity of blood removed from the body during typical hemodialysis. This "gentler" continuous removal of wastes and blood helps avoid the hypotensive episodes caused by intermittent hemodialysis and its intermittent removal of large volumes of fluid.

Kidney Transplant

Kidney transplantation is an alternative to dialysis for patients with ESRD. As with other conditions in which transplantation is indicated, the primary limiting factor is the availability of organs. Kidneys are obtained from deceased and living donors. For most of those who choose it, transplantation allows for increased independence, return to normal activities of daily living, and resumption of normal renal function. Thousands of patients receive kidney transplants each year in the United States, with a remarkable rise in transplant recipients over the last decades. The 5-year patient survival rate (2003–2008) was estimated to be about 82% for a deceased donor kidney and about 91% for a living donor transplant. Given the number of transplant recipients and the improving patient survival rates, many patients need follow-up health care outside of transplant centers. This means that general health care providers will be caring for transplant recipients and the chronic medical conditions that accompany transplantation surgery. Common medical complications include cardiovascular disease, obesity, hypertension, dyslipidemia, diabetes, cerebrovascular disease, anemia, gout,

TABLE 28.8 Staging of AKI

Stage	Serum Creatinine	Urine Output
1	1.5–1.9 times baseline OR ≥0.3 mg/dL (≥26.5 μmol/l) increase	<0.5 mL/kg/h for 6–12 hours
2	2.0–2.9 times baseline	<0.5 mL/kg/h for ≥12 hours
3	3.0 times baseline OR Increase in serum creatinine to ≥4.0 mg/dL (≥353.6 μmol/l) OR Initiation of renal replacement therapy OR, In patients <18 years, decrease in eGFR to <35 mL/min per 1.73 m^2	< 0.3 mL/kg/h for ≥24 hours OR Anuria for ≥12 hours

eGFR, Estimated glomerular filtration rate.

depression, bone disease, malignancies, and infections. Even with the best possible tissue matching, antirejection drug therapy is required, and the adverse effects of these medications may have a significant impact on health and quality of life. Transplant recipients will be prescribed a combination of immunosuppressants, usually tacrolimus or cyclosporine, mycophenolate, and prednisone.

Chronic Kidney Disease in Older Adults

A decline in kidney function as one ages is well understood. What is less clear is whether the CKD that develops in older adults is a manifestation of the aging kidney or the associated cardiovascular disease and life exposure to vascular risk factors such as hypertension, diabetes, and smoking. The majority of patients diagnosed with CKD are older adults, and the rates of treated ESRD among the elderly (>80 years) have risen by more than 50% in the last decade. Although progression of CKD to ESRD is costly and incurs significant health problems, it appears less frequently in older adults compared with cardiovascular mortality. The majority will die as a result of cardiovascular disease.

In the past 5 decades, increased numbers of older patients have initiated dialysis worldwide. Results from some studies indicate that elderly adults choosing PD have higher mortality rates than those receiving hemodialysis. With increased numbers of older adults on dialysis comes not only increased survival, but also increased morbidity. Elderly patients who are on dialysis seem to have a higher burden of age-related problems, such as frailty, falls, and cognitive impairment. There is also emerging evidence that dialysis initiation may be associated with accelerated rates of functional and/or cognitive decline. Primary care providers will be challenged with the complex care required of this population.

KEY POINTS

- Prevention of acute kidney injury (AKI) includes early identification of those at risk, maintenance of fluid volume status and cardiac output, avoidance of exposure to nephrotoxic chemicals as well as subsequent treatment if necessary, and avoidance and aggressive treatment of infections. Prerenal and postrenal kidney injuries are treated by addressing their specific etiologies. Intrinsic kidney injury (ATN) is treated with many of the same interventions used to support renal function in chronic kidney disease (CKD).
- Slowing the progression of CKD is the focus of interventions until stages 4 to 5. Appropriate management of acute tubular necrosis (ATN), blood glucose control in patients with diabetes, use of angiotensin-converting enzyme (ACE) inhibitors or AII blockers to reduce proteinuria, and aggressive management of hypertension are the primary foci. Because cardiovascular disease both is a risk factor for CKD and accelerates progression, interventions are also included to retard it.
- Nutritional needs for patients in renal failure include increased caloric intake as well as calcium and vitamin supplementation. Intake of fluids, phosphorus, potassium, sodium, and protein is usually restricted, depending on the underlying pathologic process and stage of the disease.
- Drug therapy in CKD is used to control hypertension, anemia, and some of the electrolyte and acid–base imbalances.
- Dialysis is used for some patients with ATN and for patients with CKD in stage 5 in order to remove metabolic wastes and correct fluid and electrolyte abnormalities.
- Kidney transplantation is a potential option for patients with ESRD. Kidney transplantation has been associated with a high degree of success.
- CKD is very prevalent in older adults. Although many are initiating dialysis in end-stage renal disease (ESRD), most will die of cardiovascular disease before reaching stage 5. The many comorbid conditions of this population require complex care.

SUMMARY

Renal failure can occur at any age. AKI has multiple causes that can be classified into one of three categories according to the physical location of the problem: prerenal, postrenal, or intrinsic. Each category has unique pathologic features and some variation in laboratory values. Intrinsic failure (ATN) is divided into three phases: prodromal, oliguric, and postoliguric. Interventions differ for each phase.

CKD is a progressive, irreversible process. It is characterized by stages of declining GFR producing increasing impairment in the ability of the kidney to maintain homeostasis. The clinical manifestations of CKD are determined by the degree of impairment of the kidneys' normal functions.

Key aspects of care include pharmacologic management of fluid overload, electrolyte abnormalities, and metabolic wastes; nutritional management; dialysis; and renal transplantation.

A clear understanding of the pathophysiology related to renal dysfunction is essential for any health care professional caring for patients in renal failure. Older adults with AKI, CKD, and ESRD add another layer of complexity to patient care. The impact that renal failure has on all other body systems presents many challenges.

RESOURCES

Chronic Kidney Disease

Abdelhafiz AH, Brown SHM, et al: Chronic kidney disease in older people: physiology, pathology or both? *Nephron Clin Pract* 116:c19–c24, 2010.

Ahmed AK, Brown SH, Abdelhafiz AH: Chronic kidney disease in older people: disease or dilemma? *Saudi J Kidney Dis Transpl* 21(5):835–841, 2010. Available at: www.sjkdt.org/text.asp?2010/21/5/835/68876.

Anderson S, et al: Prediction, progression, and outcomes of chronic kidney disease in older adults. *J Am Soc Nephrol* 20:1199–1209, 2009.

Baldea AJ: Effect of aging on renal function plus monitoring and support. *Surg Clin N Am* 95:71–83, 2015. http://dx.doi.org/10.1016/j.suc.2014.09.003.

Bastos MG, Kirsztajn GM: Chronic kidney disease: importance of early diagnosis, immediate referral and structured interdisciplinary approach to improve outcomes in patients not yet on dialysis. *Brazilian J Nephrol* 33:2011. Available at: http://dx.doi.org/10.1590/S0101-28002011000100013. (Accessed 28 November 2011).

Bonanni A, Mannucci I, et al: Protein energy wasting and mortality in chronic kidney disease. *Int J Environ Res Pub Health* 8(5):1631–1654, 2011.

Brosnahan G, Fraer M: Management of chronic kidney disease: what is the evidence? *Southern Med Assoc* 103(3):222–230, 2010.

Checherită IA, Turcu F, et al: Chronic complications in hemodialysis: correlations with primary renal disease. *Rom J Morphol Embryol* 51(1):21–26, 2010.

Chobanian AV, et al: Seventh Report of the Joint National Committee on prevention, detection, evaluation, and treatment of high blood pressure: the JNC 7 complete report, *Hypertension* 42:1206–1252, 2003.

Clark LE, Kahn I: Outcomes in chronic kidney disease: what we know and what we need to know. *Nephron Clin Pract* 114:c95–c103, 2010.

Dowling TC: Prevalence, etiology, and consequences of anemia and clinical and economic benefits of anemia correction in patients with chronic

kidney disease: an overview. *Am J Health Syst Pharm* 64(13 Suppl 8):S3–S7, 2007.
Eskridge M: Hypertension and chronic kidney disease: the role of lifestyle modification and medication management. *Nephrol Nurs J* 37(1):51–60, 2010.
Gupta G, Unruh ML, et al: Primary care of the renal transplant patient. *J Gen Intern Med* 25(7):731–740, 2010.
Hall JE: *Guyton and Hall textbook of medical physiology*, ed 13, Philadelphia, 2016, Saunders.
Hallan SI, Stevens P: Screening for chronic kidney disease: which strategy? *J Nephrol* 23(2):147–155, 2010.
Hedayati SS, Finkelstein FO: Epidemiology, diagnosis and management of depression in patients with CKD. *Am J Kidney Dis* 54(94):741–752, 2009.
Jassal SV, Watson D: Dialysis in late life: benefit or burden? *Clin J Am Soc Nephrol* 4(12):2008–2012, 2009.
Levey AS, Atkins R, Coresh J, et al: Chronic kidney disease as a global public health problem:approaches and initatives- a post statement from kidney disease improving global outcome. *Kidney Int* 72(3):247–259, 2007.
Marin E, Sessa WC: Role of endothelial-derived nitric oxide in hypertension and renal disease. *Curr Opin Nephrol Hypertens* 16:105–110, 2007.
Meyer TW, Hostetter TH: Uremia. *New Engl J Med* 357:1316–1325, 2007.
Nanayakkara1 PWB, Gaillard CAJM: Vascular disease and chronic renal failure: new insights. *Neth J Med* 68(1):5–14, 2010.
National Institute of Diabetes and Digestive and Kidney Disease (NIDDK), 2013: *Kidney failure: choosing a treatment that's right for you*. Available at http://kidney.niddk.nih.gov/KUDiseases/pubs/kustats/index.aspx: (Accessed 15 December 2015).
National Kidney Foundation Fact Sheet: *U.S. renal data system*, 2013. Available at www.kidney.org/news/newsroom/factsheets/end-stage-renal disease-in-the-US.
National Kidney Foundation: *Dialysis*. Available at www.kidney.org/atoz/content/dialysisinfo.cfm. (Accessed 29 January 2012).
National Kidney Foundation: *Global facts: about kidney disease*, 2015. Available at https://www.kidney.org/kidneydisease/global-facts-about-kidney disease. (Accessed 27 December 2015).
National Kidney Foundation: *KDOQI Guidelines: KDOQI clinical practice for CKD: evaluation, classification and stratification*. Available at www2.kidney.org/professionals/kdoqi/guidelines_ckdp4_class_g1htm. (Accessed 15 December 2015).
Nicola R, Schaqdan KW, Aran K, et al: Contrast-induced nephropathy: identifying the risks, choosing the right agent, and reviewing effective prevention and management methods. *Curr Probl Diagn Radiol* 44(6):501–504, 2015. doi:10.1067/j/cpradiol.2015.04.002.
Organ Procurement and Transplantation Network (OPTN) and Scientific Registry of Transplant Recipients (SRTR): *OPTN/SRTR 2010 annual data report*, 2011. Rockville, MD, U.S. Department of Health and Human Services, Health Resources and Services Administration, Healthcare Systems Bureau, Division of Transplantation.
Pagana KD, Pagana TJ: *Mosby's manual of diagnostic and laboratory tests*, ed 4, St Louis, MO, 2011, Mosby.
Porter RS, Kaplam JL: *The Merck manual*, ed 19, Whitehouse Station, NJ, 2011, Merck Sharp & Dohme.
The Renal Association: Available at www.renal.org/whatwedo/InformationResources/CKDeGUID:E/CKDstages.aspx. (Accessed 28 January 2012).
Rosner MH: Hemodialysis for the non-nephrologist. *South Med J* 98(8):785–791, 2005.
Stompór T, Olszewski A, Kierzkowska I: Can we prolong life of patients with advanced chronic kidney disease: what is the clinical evidence? *Polish Arch Int Med* 121(3):88–92, 2011.
Tamura MK: Incidence, management, and outcomes of end-stage renal disease in the elderly. *Curr Opin Nephrol Hypertens* 18(3):252–257, 2009.
Weiner DE: Causes and consequences of chronic kidney disease: implications for managed health care. *J Manag Care Pharm* 13(Suppl 3):S1–S9, 2007.
Williams A, Manias E: A structured review of pain assessment and management of patients with chronic kidney disease. *J Clin Nurs* 17(1):69–81, 2008.

Acute Kidney Injury

Abuelo JG: Normotensive ischemic acute renal failure. *New Engl J Med* 357:797–805, 2007.
Aitken C, Carruthers L, Gall L, et al: Acute kidney injury: outcomes and quality of care. *Q J Med* 2013. doi:10.1093/qjmed/hcs237.
Basile DP, Anderson MD, Sutton TA: Pathophysiology of acute kidney injury. *Compr Physiol* 2(2):1303–1353, 2012. doi:10.1002/cphy.c110041.
Bonventre JV, Lang L: Cellular pathophysiology of ischemic acute kidney injury. *J Clin Invest* 121(11):4210–4221, 2011.
Couser WG, Remuzzi G, Mendis S, Tonelli M: The contribution of CKD to the global burden of major noncommunicable disease. *Kidney Int* 80(12):1258-1270, 2011.
Dirkes S: Acute kidney injury: not just acute renal failure anymore? *Crit Care Nurse* 31(1):37–49, 2011.
Dirkes SM: Acute kidney injury: causes, phases and early detection. *American Nurse Today* 10(7):20–25, 2015.
Goldenberg I, Matetzky S: Nephropathy induced by contrast media: pathogenesis, risk factors and preventive strategies. *Can Med Assoc J* 172(11):1461–1471, 2005.
Kellum J, LeBlanc M, Venkataraman V: Clinical evidence concise: acute renal failure. *Am Fam Physician* 76(3):2007.
Kidney Disease: Improving Global Outcomes (KDIGO) Acute Kidney Injury Work Group: KDIGO Clinical Practice Guideline for Acute Kidney Injury. *Kidney Inter* 2(Suppl):1–138, 2012.
Lameire N, Van Biesen W, Vanholder R: Acute renal failure. *Lancet* 365:417–430, 2005.
Madala ND: Acute renal failure in patients with chronic kidney disease. *CME* 25(8):395–398, 2007. Available at: www.ajol.info/index.php/cme/article/viewFile/43804/27324. (Accessed 24 March 2012).
Needham E: Management of acute renal failure. *Am Fam Physician* 72(9):2005, 1739–1746.
Sabbahy ME, Vaidya VS: Ischemic kidney injury and mechanisms of tissue repair. *Wiley Interdisc Rev Syst Biol Med* 3(5):606–618, 2011.
Wang HE, Muntner P, Chertow GM, Warnock DG: Acute kidney injury and mortality in hospitalized patients. *Am J Nephrol* 35:349–355, 2012. doi:10.1159/000337487.
Willacy H: *Acute on chronic renal failure, Patient.co.uk*, 2010. Available at www.patient.co.uk/doctor/Acute-on-Chronic-Renal-Failure.htm. (Accessed 24 March 2012).
Yaklin KM: Acute kidney injury: an overview of pathophysiology and treatments. *Nephrol Nurs J* 38(1):13–17, 2011.

29

Disorders of the Lower Urinary Tract

Cheryl L. Brandt

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- How do the pathophysiologic characteristics and management of stress, urgency, and mixed incontinence differ?
- What are the manifestations and management of overactive bladder syndrome and painful bladder syndrome/interstitial cystitis?
- How are congenital abnormalities of the urinary collecting system detected and treated?
- What are the risk factors and clinical manifestations for bladder cancer?
- How are urethritis and cystitis prevented and managed?
- How do stones in the lower urinary tract present clinically?

CHAPTER OUTLINE

Lower Urinary Tract, 609
Functional Anatomy, 609
Physiology of Micturition, 610
Nervous System Innervation of the Lower Urinary Tract, 610
Mechanism of Micturition, 610
Diagnostic Tests, 611
Lower Urinary Tract Symptoms and Syndromes, 611
Incontinence, 611
Enuresis, 613
Overactive Bladder Syndrome, 613
Bladder Pain Syndrome/Interstitial Cystitis, 613
Neurogenic Bladder, 614
Congenital Disorders, 615
Primary Vesicoureteral Reflux, 615
Obstruction of the Ureteropelvic Junction, 616
Ureteral Ectopy, 616
Ureterocele, 616
Neoplasms, 617
Bladder Cancer, 617
Inflammation and Infection, 619
Urethritis, 619
Cystitis, 619
Obstruction, 622
Lower Urinary Tract Calculi, 622
Ureteral Calculi, 622
Bladder (Vesical) Calculi, 622

The lower urinary tract (LUT) encompasses several structures that together are responsible for the transport, storage, and elimination of urine from the body. These structures include the ureters, bladder, urethra, and associated urinary sphincters. The male prostate gland is anatomically located in this area, but is functionally involved in male reproduction; the role and disorders of the prostate gland are presented in Chapters 30 and 31, respectively.

In addition to LUT symptoms and syndromes, disorders of the LUT can generally be classified as congenital, neoplastic, infective, inflammatory, or obstructive. These disorders are often interrelated. For instance, stasis of urine, which occurs with a variety of disorders, often leads to urinary tract infection (UTI). Pathologies of the LUT can give rise to infection of the kidney, pyelonephritis (see Chapter 27), and acute postrenal kidney injury (see Chapter 28). LUT disorders can result in significant health problems that have tremendous physical, psychosocial, and economic ramifications for patients and their families.

LOWER URINARY TRACT

The role of the LUT is to transport and store urine and facilitate its elimination through micturition.

Functional Anatomy

Ureters collect the urine formed in the kidneys from the renal pelvises and transport it to the bladder. Urine movement from the kidneys to the bladder is due to the effect of gravity facilitated by peristaltic

movement of the ureters. The bladder is a hollow, muscular reservoir for urine that expands to store it and then contracts to expel urine through the urethra. Urine storage depends on intact spinal reflexes. The process of micturition (voiding) involves both reflex and voluntary mechanisms, mediated by the micturition center in the pons. Voiding is a result of coordinated function of bladder mechanoreceptors, neurologic impulse transmission, bladder muscle contraction, and urethral sphincter relaxation. Anatomic integrity of the ureters and bladder, competent urethral sphincters, and an appropriately functioning nervous system are required for the lower urinary system to properly carry out its role.

Physiology of Micturition

Micturition is often taken for granted but is actually a complicated process. Understanding the process begins with a review of the significant role the nervous system plays in controlling the functions of the LUT structures.

Nervous System Innervation of the Lower Urinary Tract

The central, autonomic, and peripheral nervous systems are all involved in urinary elimination. The pontine micturition center coordinates relaxation of the internal sphincter and contraction of the bladder to enable urination, whereas the cerebral cortex primarily inhibits the process through conscious control of the external sphincter. Any disease process affecting these areas can interfere with urination. This includes such pathologies as Parkinson disease, multiple sclerosis, traumatic brain injury, or stroke.

The smooth muscle of the bladder is innervated by both the sympathetic and parasympathetic branches of the autonomic nervous system, though sympathetic fibers mainly control blood flow to the bladder and have a role in pain sensation (Fig. 29.1). Parasympathetic innervation of the urinary tract is supplied via the pelvic nerves, which exit the spinal cord at S2-S4; sensory branches detect stretch of the bladder wall, and motor branches stimulate reflex detrusor contraction. In other words, the bladder itself does not respond to voluntary control, but rather to neurologic reflexes designed to initiate bladder emptying. Injury or disease of the spinal cord in this region profoundly affects LUT function. Finally, the somatic nerve fibers of the peripheral nervous system innervate the external bladder sphincter via the pudendal nerve. Motor fibers enable voluntary control of the external bladder sphincter, permitting selective timing of urination.

Mechanism of Micturition

The softness, pliability, mucosal secretions, and submucosal cushioning of the urethra help maintain a watertight seal, or "compression," of the urethra. The *internal sphincter* is located at the proximal portion of the urethra, where convergence of the detrusor muscle fibers provides pressure to keep it closed. When the internal sphincter has normal tone, the bladder is prevented from emptying until the pressure in the body of the bladder rises above a specific threshold. The *external sphincter* is located at the distal end of the urethra, surrounded by a ring of skeletal muscle from the pelvic floor. This musculature provides the tension needed to maintain continence at normal resting bladder pressures.

Bladder emptying occurs as a result of both reflex and voluntary activities. Urine filling the bladder raises the internal bladder pressure and triggers stretch receptors (mechanoreceptors) in the bladder wall. Activation of these receptors sends afferent signals to the spinal cord

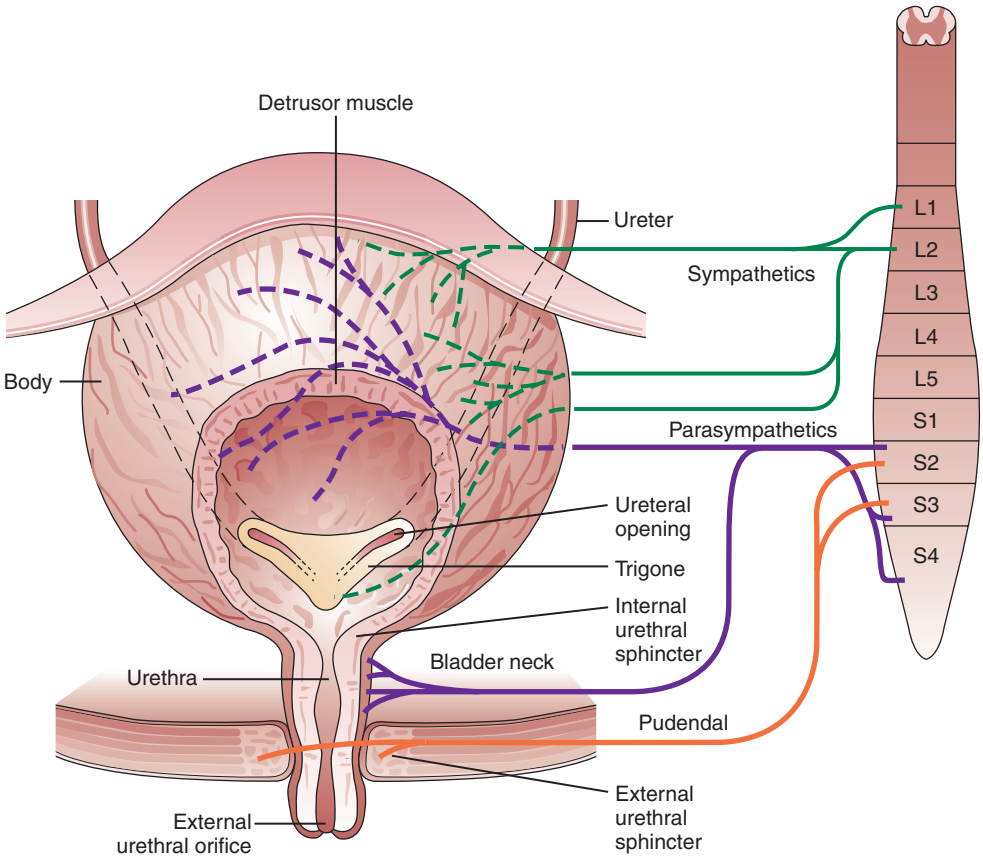


FIG 29.1 Bladder structure and innervation.

followed by reflex return signals via parasympathetic motor fibers that stimulate detrusor muscle contraction, called *micturition contractions.* Usually, enough sphincter pressure to hold the urine in the bladder can be voluntarily generated until about 350 to 400 mL of urine has collected in the bladder. At this point an urgent sensation to void occurs. However, activation of the voiding centers in the pons and cerebral cortex of the central nervous system (CNS) also help inhibit the bladder from emptying, primarily by partially inhibiting micturition reflexes and contracting the external sphincter. At the appropriate time, abdominal muscles are voluntarily, contracted which further raises internal bladder pressure and triggers a renewed micturition contraction; simultaneous relaxation of the external sphincter permits voiding. If an individual is able to coordinate urination at a suitable location and desirable time, the individual is said to be continent**.**

Normal adult bladder capacity is 300 to 500 mL; the urge to void usually develops when 150 to 250 mL are present. With an average fluid intake, this typically results in a voiding pattern of every 4 to 5 hours, with no need to awaken at night in order to void. With aging, bladder capacity declines to 200 to 350 mL; urination becomes more frequent (every 3 to 4 hours), and awakening at night to urinate (nocturia) is common. Under normal circumstances, the adult bladder contains less than 50 to 100 mL after voiding. This volume of postvoiding urine is called *residual urine*. Certain pathologies are associated with incomplete bladder emptying and increased residual urine volumes.

Diagnostic Tests

Several of the procedures presented in Chapter 26 are used to diagnose urologic disorders discussed in this chapter. Other, more specialized, diagnostic tests are often even more appropriate. A urinalysis is the simplest and least costly test that can provide a wealth of information. For the LUT, this test is primarily used in the diagnosis of infection.

Ultrasonography, which is painless, does not involve radiation, and provides excellent visualization of the urinary system, is a common nephrology diagnostic test. Cystography with contrast media instilled into the bladder may be required to yield more specific information about the bladder than can be obtained by ultrasound. Voiding cystourethrography (VCUG) involves placing a catheter in the bladder and then filling it with sterile, iodinated, dilute contrast material. The catheter is then removed, and the patient voids. Images of the bladder are taken before voiding to detect ureterocele or tumor, and images taken during voiding can identify reflux or urethral abnormalities. Radionuclide voiding cystography also requires catheterization but involves the use of a small amount of radioactive material. A technetium-99m–labeled radiopharmaceutical is instilled in the bladder through the catheter, followed by sterile normal saline to fill the bladder. Once the catheter is removed, images are taken with the bladder full and during voiding, but this procedure does not allow visualization of the urethra.

The term *urodynamic testing* is used for procedures associated with diagnosing voiding dysfunction. In urodynamic testing both the filling/storage and voiding phases of urinary elimination are studied. There are multiple urodynamic tests and procedures, and the choice of tests is based on clinical presentation, typically beginning with the least invasive of the desired tests. The most common tests are uroflowmetry (noninvasive method of measuring characteristics of urine flow during voiding), cystometry (measurement of intrabladder pressure during filling and postvoid residual), urethral pressure profilometry (measurement of intraluminal pressure along the length of the urethra), and voiding pressure flow studies (invasive measurement of urine flow during micturition). Pelvic floor electromyelogram may also be done to identify pathologic processes underlying voiding dysfunction.

Diagnostic tests for bladder cancer range from cystoscopy for the detection of bladder tumors to computed tomography (CT) scans or magnetic resonance imaging (MRI) to identify muscle-invasive cancer. Urine tests for bladder cancer markers, such as bladder tumor antigen and nuclear matrix protein (NMP22), are used to monitor for tumor recurrence after treatment.

KEY POINTS

- The lower urinary tract (LUT) transports and stores urine from the kidneys and facilitates its elimination.
- Bladder innervation is accomplished via autonomic and somatic nerves. Stimulation of parasympathetic nerves from S1-S4 results in bladder contraction and relaxation of the internal sphincter. The somatic pudendal nerve innervates the external bladder sphincter.
- Micturition requires central, autonomic, and peripheral nervous system functioning. It is a result of coordinated parasympathetic nervous system and voluntary actions.
- A variety of diagnostic tests are used in the diagnosis of LUT disorders.

LOWER URINARY TRACT SYMPTOMS AND SYNDROMES

Lower urinary tract symptoms (LUTS) are subjective indicators of a change in condition or disease of the LUT. LUTS, including incontinence and enuresis, may also be associated with pathologies affecting the central, autonomic, and peripheral nervous systems. Genitourinary syndromes marked by LUTS include overactive bladder and painful bladder syndrome.

Incontinence

The International Continence Society (ICS) defines the symptom of urinary incontinence (UI) as the report of any involuntary urine loss. UI is prevalent among adults, affecting an estimated 30% of younger adults. The prevalence is higher in U.S. adults over age 65; greater than 50% of a noninstitutionalized sample of women and greater than 25% of a noninstitutionalized sample of men reported urinary leakage. Prevalence of daytime incontinence in children in the United Kingdom has been reported at 15% for 4.5-year-olds and 5% for 9.5-year-olds.

Pathogenesis. Continence requires both physiologic and cognitive capabilities. Physiologic requirements include an appropriately functional nervous system and intact bladder and urethral function. Cognitively, individuals must have the ability to react to bladder urges and be motivated to be continent. Any disruption in these capabilities can result in an inability to attain or maintain continence. UI is not a part of normal aging, but it may be a result of age-related changes in the function or structure of the LUT. Age brings about changes in the bladder, including altered sensitivity of urothelial sensory afferents, increased collagen deposition, enlarged space between myocytes, and changes in gap junctions. Bladder function is also affected; bladder capacity and detrusor contractility decrease and detrusor overactivity and residual urine volumes increase. Finally, decreased estrogen in postmenopausal women has an impact on fibers of the detrusor and urethral muscles.

UI has been classified by the ICS as urge, stress, and mixed incontinence. The etiologies and treatments vary, so identification of the type of incontinence is clinically important.

Urgency urinary incontinence (UUI) involves the involuntary leakage of urine along with or immediately after the sudden sensation of a need to urinate (urgency). This condition is most often due to an overactive detrusor muscle that suddenly contracts. Aging is known to increase the frequency of spontaneous involuntary detrusor contractions. Other contributing factors may include bladder infections that irritate the

bladder lining and bladder outlet obstruction attributable to prostate enlargement. Additionally, urgency incontinence may result from CNS conditions (e.g., stroke, Parkinson disease, multiple sclerosis) in which damage to bladder contraction inhibitory pathways occurs. Finally, drugs that increase urine flow, such as diuretics and alcohol, can aggravate UUI. In cases in which a specific cause is not determined, urgency incontinence is said to be *idiopathic.*

Stress urinary incontinence (SUI) occurs when urine is involuntarily lost with increases in intraabdominal pressure. It is precipitated by effort or exertion, such as by lifting heavy objects, coughing, sneezing, or bending. Stress incontinence is thought to occur as a result of loss of pelvic muscle and/or fascial support of the bladder and urethra. Without this support, whenever there is an increase in intraabdominal pressure, the normal angle between the bladder and posterior urethra is disrupted, forces that support urethral closure are reduced, and urine is lost. Decreased estrogen availability with menopause also contributes by reducing urethral closing pressure. Age-related loss of pelvic floor muscle fibers decreases muscular support. Additional risk factors for stress incontinence include obesity, childbirth-related trauma, pelvic surgery and radiotherapy, and the presence of conditions such as diabetes or degenerative neurologic diseases that impair nerves innervating the structures involved in micturition.

Mixed urinary incontinence (MUI) is a combination of both stress and urge incontinence. It is common for these two types of incontinence to occur together, especially in older women.

The term *overflow incontinence* is applied when the bladder becomes so full that it leaks urine, or "overflows." This phenomenon is more common in men, resulting when an obstructed urethra prevents the bladder from emptying normally, such as with an enlarged prostate gland. Some clinicians describe *functional incontinence* as incontinence related to physical or environmental limitations resulting in an inability to access a toilet in time. The urinary system may work well, but inaccessible toilets, mobility disorders, cognitive dysfunction, or mental disabilities prevent normal or timely toilet usage. Finally, *transient incontinence* has a sudden onset and is due to potentially reversible conditions such as UTIs, constipation, or fecal impactions. Risk factors for UI are summarized in Box 29.1.

Diagnosis. Patients should be assessed for reversible issues affecting the function of the LUT. These include potential drug-induced effects, UTIs or obstructions, fecal impaction, overuse of alcohol or caffeine, and excess intake of fluid. Impaired cognitive function may be due to chronic illness, depression, or delirium. All disorders associated with cognitive or neurologic function (e.g., stroke, Parkinson disease) should be identified. Circumstances or conditions affecting mobility should also be acknowledged.

Patients may be asked to keep a bladder diary, recording the time, frequency, and volume of micturition, as well as incidents of incontinence. In addition to the physical examination, such diagnostic tests as residual urine measurement, filling cystometry studies, and pressure flow studies during voiding may be used to establish the diagnosis of incontinence.

Treatment. Transient incontinence is managed by finding and treating the cause. Health care workers are critical in managing functional incontinence by manipulating the environment to facilitate the patient's timely access to the toilet to maintain what has been termed *dependent continence.*

Management of UUI, SUI, and MUI begins by addressing reversible contributing factors. Lifestyle changes such as losing weight, reducing caffeine intake, and avoiding constipation may be useful. Additional treatment may include behavioral, pharmacologic, and surgical interventions, depending on the cause of incontinence. Pelvic floor muscle training (PFMT), a behavioral intervention, is recommended for urge incontinence in both men and women. PFMT may be as simple as the performance of exercises to strengthen pelvic muscles or more complex and include techniques with vaginal weights, pelvic floor electrical stimulation, and biofeedback. Bladder training incorporates education, scheduled voiding with systematic delay of voiding to conform with the schedule, and positive reinforcement.

Pharmacologic agents may be used to promote or inhibit physiologic activities associated with micturition, depending on the cause of incontinence. This may include anticholinergic agents (e.g., oxybutynin), vaginal or oral estrogen, and α-adrenergic blockers (e.g., prazosin, tamsulosin) or 5α-reductase inhibitors (e.g., finasteride) for men with bladder outlet obstruction. Mirabegron, a β3-adrenergic agonist, is useful in the management of UUI. The addition of drug therapy often increases the effectiveness of behavioral interventions. Injections with onabotulinumtoxinA are being used for detrusor overactivity; the toxin causes relaxation of the muscle.

Surgical procedures for incontinence, used if nonsurgical interventions are ineffective, vary depending on the underlying anatomic or physiologic problems. Several surgical options are available to treat urge incontinence in women. The anterior colporrhaphy involves the repair of a weakened anterior vaginal wall that has allowed the bladder to prolapse into the vagina. A colposuspension is performed by using sutures to tighten the muscles of the pelvic floor that support the uterus and bladder; the procedure may be performed via laparoscope. Sling procedures involve implanting strips of synthetic or natural material around the bladder neck and urethra to support urethral closure. Surgical options for men may include artificial urinary sphincter implantation for sphincter incompetence and placement of a perineal compression sling for postprostatectomy incontinence.

Incontinence that is not resolved by behavioral, pharmacologic, or surgical treatment may be managed by supportive approaches such as intermittent catheterization, indwelling catheterization, or use of incontinence undergarments. Each of these options creates the potential for further complications. UTIs are more likely with stasis of urine in

BOX 29.1 Risk Factors for Incontinence

Risk Factors

Immobility
Impaired cognition
Medications (e.g., diuretics)
Morbid obesity
Smoking
Fecal impaction
Delirium
Environmental barriers to toileting
High-impact physical activities
Estrogen depletion
Low fluid intake
Pelvic muscle weakness
Childhood nocturnal enuresis
Pregnancy, vaginal delivery, or episiotomy

Pathologies

Urinary tract infection
Diabetes mellitus
Diabetes insipidus
Stroke
Multiple sclerosis
Parkinson disease
Spinal cord injury
Spinal cord defects

the bladder, as well as with continuous or intermittent catheterization of the normally sterile bladder. Catheterization is never an appropriate solution solely for caregiver convenience. Stasis of urine also increases the risk for bladder and renal calculi. Management of incontinence with incontinence undergarments predisposes patients to skin breakdown.

Enuresis

Enuresis as defined by the International Children's Continence Society means intermittent (discrete amounts) incontinence while asleep. Enuresis is a common childhood condition, twice as common in boys as in girls. An estimated 14% of 5-year-olds have enuresis, though the prevalence decreases with advancing age; enuresis typically spontaneously resolves. In *monosymptomatic enuresis,* the child has nocturnal incontinence but no other LUTS such as urgency, altered voiding frequency, and daytime incontinence. *Nonmonosymptomatic nocturnal enuresis* is diagnosed when the child has symptoms of urgency, frequency, or daytime incontinence in addition to nighttime enuresis. *Primary nocturnal enuresis* describes a child who has never achieved consistent nighttime continence, whereas *secondary enuresis* refers to enuresis that develops after a period of at least 6 months of dryness. Primary enuresis is most common.

Pathogenesis. Several pathophysiologic mechanisms are thought to contribute to enuresis. Some children exhibit nocturnal polyuria, linked in many cases to a decreased response to or deficiency in production of vasopressin (antidiuretic hormone [ADH]). Other children exhibit nocturnal overactivity of the detrusor muscle, which contributes to incontinence. Reduced nocturnal bladder function has been found to result in decreased nighttime bladder capacity. Finally, immature or abnormal sleep arousal mechanisms, postulated as a dysfunction of the bladder–brain connection arising from the pontine tegmentum (micturition center), may explain why the child does not awaken.

The fact that parents and siblings of children with nocturnal enuresis also report a history of the problem has led to the establishment of a genetic contribution. Enuresis is associated with such conditions as constipation, fecal incontinence, attention-deficit hyperactivity disorder, and sleep disturbances.

Diagnosis. Clinical workup for enuresis includes a thorough history of elimination patterns via a bladder diary and a physical examination to identify gross anatomic abnormalities. Children should also be tested for comorbidities such as diabetes mellitus, constipation, developmental difficulties, and behavioral problems. Additional diagnostic tests, including urinalysis, flow cystometry, and urinary tract imaging, are warranted for children who present with nonmonosymptomatic nocturnal enuresis. Invasive diagnostic procedures are typically reserved for children suspected of having organic or neurologic abnormalities.

Treatment. Treatment for enuresis, not recommended before age 5, begins with education about appropriate daytime and bedtime voiding patterns as well as appropriate fluid intake and toileting patterns. This is sometimes termed *urotherapy.* If constipation is present, it is managed with fiber intake, physical activity, and stool softeners. Enuresis alarms, designed with a moisture sensor in the bed linen or nightclothes, arouse and/or condition children to contract pelvic muscles; alarm therapy may be appropriate if enuresis is infrequent.

Pharmacotherapy may be added if nonpharmacologic treatments are ineffective. Desmopressin, a vasopressin (ADH) analog, effects a full response in 30% of children and a partial response in 40%. Anticholinergics such as oxybutynin and tolterodine are sometimes used, often in combination with desmopressin, for children who have not responded to desmopressin alone. Urinary retention and constipation are side effects of anticholinergics that are worrisome in the setting of enuresis. Imipramine, a tricyclic antidepressant, may be effective in children refractory to urotherapy and other pharmacologic treatments. However, its risk of cardiotoxicity, increased suicidality, and overdose mandates careful use in selected children. In children who remain enuretic after many months of standard treatment, additional diagnostic testing is recommended. Efforts to manage enuresis without treating an underlying pathology are likely to be unsuccessful.

Overactive Bladder Syndrome

Overactive bladder syndrome (OAB) is a symptom syndrome characterized by urinary urgency, often associated with increased daytime frequency and nocturia, though not necessarily with incontinence. Not all people with OAB experience urgency incontinence, but by definition all people with urgency incontinence have OAB. Among 20,000 U.S. residents who completed an Internet survey about their LUTS, over 35% reported OAB symptoms at least sometimes. Prevalence increased with advancing age and was higher in women compared with men. Another study of OAB prevalence among 10,000 ethnically diverse adults in the United States found an overall prevalence of OAB symptoms at least sometimes of 23.2%, with a higher prevalence in women and in African American men and women compared with Hispanic and white men and women. In children, the incidence of OAB peaks between ages 5 and 7.

Pathogenesis. Multiple factors are associated with OAB. Urgency is thought to be a result of involuntary detrusor muscle contractions, hypersensitivity of afferent nerves due to dysfunction of urothelial or interstitial cells in the mucosal layer of the bladder, or abnormal processing in the CNS of afferent signals from the bladder. The anatomic and physiologic changes of aging, including decreased compliance and capacity of the bladder, lead to increased risk. In addition to increasing age and female gender, diabetes is a risk factor for OAB. Conditions associated with OAB include chronic constipation, fecal incontinence, and poorer sexual health, especially when OAB includes incontinence.

Diagnosis and clinical manifestations. A diagnosis of OAB is established for most people by history, physical examination, and a urinalysis. Clinical manifestations typically include urgency, urinary frequency, urge incontinence, and nocturia. Other conditions that might cause these symptoms, such as obstruction of the bladder outlet or UTI, must be ruled out.

Treatment. First-line treatment for OAB includes behavioral therapies such as adopting a voiding schedule, pelvic floor muscle strengthening, limiting nighttime fluid intake, and reducing caffeine. Second-line pharmacologic treatments include antimuscarinic anticholinergics (e.g., tolterodine, solifenacin) and beta-3 agonists (e.g., mirabegron). Transdermal oxybutynin may also be useful. Of note, the cognitive (e.g., delirium) and gastrointestinal (e.g., constipation) side effects of antimuscarinic agents can be especially troublesome in the older adult population. Third-line therapies, options when first- and second-line treatments have been ineffective, include onabotulinumtoxinA injections into the detrusor muscle and neural stimulation of the detrusor via peripheral nerve pathways using an implanted stimulator.

Bladder Pain Syndrome/Interstitial Cystitis

Bladder pain syndrome (BPS), also called *painful bladder syndrome,* is a genitourinary pain syndrome in which suprapubic pain is experienced with bladder filling, along with other LUTS such as frequency, unaccompanied by proven UTI or other painful condition. The syndrome is also called *interstitial cystitis* (IC). Other definitions of the syndrome specify that the pelvic pain must be of at least 6 months' duration and be accompanied by at least one other LUTS. Urgency, dysuria, and dyspareunia may accompany the pain, which may worsen with consumption of certain foods and drinks such as alcohol- and caffeine-containing beverages, citrus juices and fruits, and hot peppers. Pain is often relieved with voiding. "Classic" IC has been described as a subtype of BPS, characterized by visible inflammatory lesions of the bladder mucosa

and submucosa called *Hunner ulcers*. However, only an estimated 10% of people with BPS exhibit these ulcers.

BPS/IC is more prevalent in women, with a 5:1 ratio of females to males. Estimates of overall prevalence are 500 cases per 100,000 people worldwide. Less is known about the prevalence of BPS/IC in children and adolescents. This condition can have a serious impact on quality of life, interfering with work, daily activities, sleep, family life, and sexual activities.

Pathogenesis. The specific etiology of BPS/IC is unknown, and the pathophysiology is unclear, but a fundamental mechanism appears to be urothelial dysfunction. One theory is that an abnormal bladder glycosaminoglycan layer reduces protective mucous production, leading to increased permeability of the bladder wall and absorption of damaging urine components such as urea and potassium ions. An inflammatory response follows, with mast cell activation and histamine and other neurotransmitter release. Sensory afferents in the bladder are up-regulated, and previously silent C-fibers may become active, causing pain. Research is ongoing into the role of cytokines and chemokines in bladder inflammation, dysfunction, and pain. People with BPS are more likely to be diagnosed with other pain syndromes such as fibromyalgia and with conditions such as inflammatory bowel disease and rheumatoid arthritis, suggestive of neurogenic and autoimmune mechanisms, respectively.

Diagnosis and clinical manifestations. Diagnosis of BPS/IC is by exclusion and is challenging due to the nonspecific nature of the syndrome and absence of specific diagnostic tests. The workup should include a history, physical examination, bladder diary, pain assessment, urinalysis, and a urine culture and cytology to rule out subclinical infection and bladder cancer. Cystoscopy and urodynamic studies are recommended only in complex cases; BPS/IC is diagnosed as a symptom-based syndrome after eliminating other pathologies.

Treatment. BPS/IC treatment strategy is typically multimodal, with no one therapy proving effective over time for most patients. Pain management is a key component. First-line treatments include patient education about the syndrome and self-management strategies, including adjusting fluid intake to manage the concentration and volume of urine, using local heat or cold application, avoiding foods and fluids that trigger pain, and relaxation of the pelvic muscle floor. Second-line treatments include manual physical therapy and pharmacologic management of pain with options including pentosan polysulfate, amitriptyline, cimetidine, and hydroxyzine. For people for whom first- and second-line therapies are ineffective, subsequent treatments may be selected from cystoscopy with hydrodistention, fulguration of Hunner lesions, intradetrusor injection of onabotulinumtoxinA, implantation of neurostimulation devices (e.g., sacral, pudendal), oral cyclosporine A, and, finally, major surgery such as cystectomy with urinary diversion as a last resort.

KEY POINTS

- Lower urinary tract symptoms (LUTS) may be secondary to disorders of the lower urinary tract (LUT), attributable to pathologies affecting the central, autonomic, and peripheral nervous systems, or associated with a wide variety of factors affecting control of micturition, including medications and access to toileting facilities.
- Urgency incontinence may be attributable to detrusor muscle overactivity, bladder infections that irritate the bladder lining, radiation therapy, tumors or stones in the urinary tract, or central nervous system (CNS) damage to inhibitory pathways (as would occur with stroke, dementia, Parkinson disease, and multiple sclerosis).
- Decreased pelvic muscle support of the bladder and urethra or intrinsic urethral sphincter deficiency results in stress incontinence.
- Mixed incontinence is a combination of both stress and urgency incontinence.
- Bladder outlet obstruction may trigger overflow incontinence.
- Functional incontinence is related to physical or environmental limitations in reaching a toilet in time to void.
- Treatment options for voiding dysfunction include behavioral, pharmaceutical, and surgical interventions.
- Enuresis is inappropriate wetting of clothing or bedding, with the term usually reserved for nocturnal incontinence in children. Treatment for enuresis is typically behavioral modification with or without pharmacologic intervention.
- Overactive bladder syndrome (OAB) is characterized by urgency with or without incontinence. Prevalence is higher in women and older adults. The most common treatments are behavioral and pharmaceutical.
- Bladder pain syndrome/interstitial cystitis (BPS/IC) is a chronic condition consisting of bladder pain and often urgency, frequency, and dyspareunia when no other etiology can be identified. It is diagnosed based on this information. Treatment is multimodal and tailored to manage pain and maximize quality of life.

NEUROGENIC BLADDER

Neurogenic bladder is a broad classification of voiding dysfunction in which the specific cause is pathology that disrupts the nervous communication governing micturition. Patients with neurogenic bladder include those with CNS disorders (e.g., multiple sclerosis, Parkinson disease, stroke), disorders affecting the autonomic innervation of the bladder (e.g., spinal cord injury), and neuropathy of diabetes mellitus. The pathogenesis of the voiding dysfunction depends on the etiologic condition. Demyelinization of neurons and axonal degradation in multiple sclerosis are thought to affect detrusor function, resulting in detrusor overactivity with storage symptoms (e.g., UUI and frequency) or detrusor areflexia with voiding symptoms (e.g., hesitancy, slow stream, retention). The central neurodegeneration of Parkinson disease results in both storage symptoms and voiding symptoms. A spinal cord injury proximal to the sacral cord may cause detrusor overactivity (spasticity), whereas injury of the sacral cord or cauda equina may cause detrusor arreflexia. Oxidative stress due to hyperglycemia of diabetes mellitus is thought to alter detrusor structure and function, leading to voiding dysfunction.

Along with a history and physical examination, urodynamic tests and measurement of postvoid residual urine volume are useful to diagnose the voiding dysfunction of neurogenic bladder. Management of storage symptoms may include antimuscarinic medications and desmopressin. Injections of the detrusor with onabotulinumtoxinA, as well as sacral neuromodulation (stimulation), and, if more conservative measures fail, surgical procedures such as urinary diversion are also treatment options. If spontaneous voiding cannot be established, clean intermittent catheterization is preferred over the use of indwelling catheters to reduce the risk of UTIs.

KEY POINTS

- *Neurogenic bladder* is a broad classification of voiding dysfunction in which the specific cause is a pathology that produces a disruption of neurologic communication governing micturition.
- Both storage (e.g., UUI, frequency) and voiding (e.g., hesitancy, slow stream, retention) symptoms may characterize neurogenic bladder.
- Pharmacologic, neuromodulation, and surgical options exist for managing storage symptoms, whereas clean intermittent catheterization is preferred for managing inadequate spontaneous voiding.

CONGENITAL DISORDERS

Primary Vesicoureteral Reflux

Reflux of urine from the bladder to the ureter and renal pelvis, known as vesicoureteral reflux (VUR), is usually due to incompetence of the valvular mechanism at the ureter–bladder junction. Primary VUR is due to a congenital anatomic anomaly.

The incidence of vesicoureteral reflux in the general population of children is low, estimated at 0.4% to 1.8%. However, the incidence in infants and children with a history of febrile UTIs is much greater, estimated at 30%. More common in females, a genetic component is also present; as many as 34% of people with VUR have siblings with the condition. The specific mode of inheritance has not yet been identified; dominant inheritance with variable penetration is hypothesized.

Pathogenesis. VUR is usually due to incompetence of the valvular mechanism at the ureterovesical junction. Normally, the ureters enter the bladder at an oblique angle and then continue for 1 to 2 cm under the bladder mucosa before exiting inside the bladder cavity. As the bladder fills, pressure within it increases against the muscle wall and closes the ureteral passageway. In the case of VUR, there is a shortened ureteral tunnel through the bladder wall, which results in lateral displacement of the valvular mechanism, rendering it incompetent. Urine is permitted to flow backward into the ureters, sometimes reaching the kidney. Impaired bladder dynamics may be a contributing factor to VUR; the prevalence reports of bladder dysfunction in children with VUR have ranged from 38% to 75%. Reflux may be bilateral or unilateral; the extent of reflux is graded from I to V (Fig. 29.2). Spontaneous resolution of reflux occurs commonly, though it is more likely to occur with lower grades of reflux.

Reflux of urine may cause increased renal pelvis pressure. Additionally, migration of bacteria from nonsterile urine to the kidneys may result in pyelonephritis or renal scarring that can ultimately cause reflux nephropathy leading to hypertension or renal insufficiency.

Diagnosis and clinical manifestations. VUR may be discovered prenatally, during ultrasonography, when hydronephrosis is identified. It may also be identified during evaluation of infants and children with

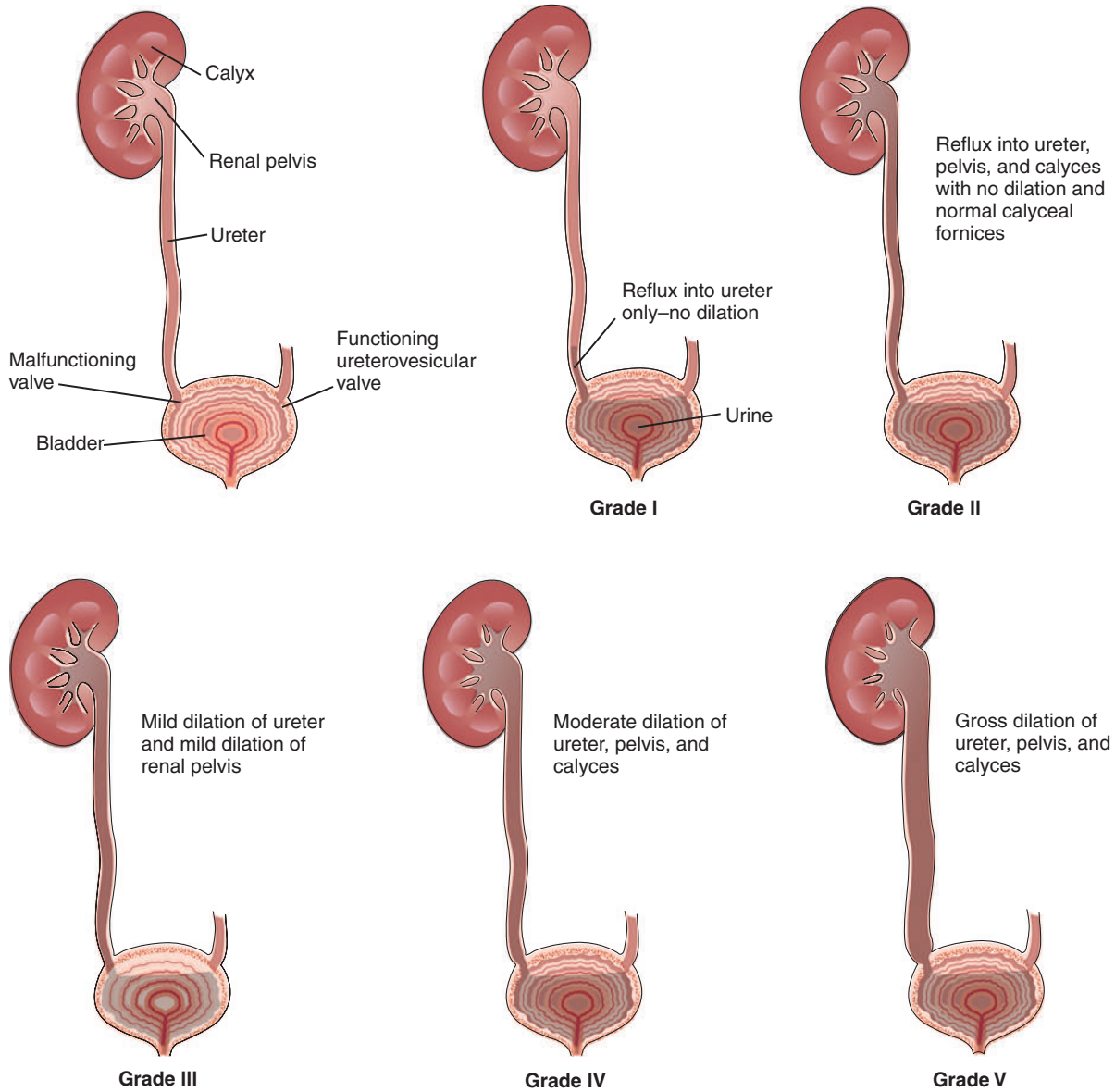


FIG 29.2 International classification of vesicoureteral reflux. (From James SR et al: *Nursing care of children: principles and practice,* ed 4, Philadelphia, 2013, Saunders, p 576.)

recurrent febrile UTI. Finally, siblings of children with VUR may be screened for the condition. In addition to a history, physical examination, and urinalysis, renal ultrasound to evaluate the upper urinary tract and dimercaptosuccinic acid (technetium-99m-labeled dimercaptosuccinic acid) scan to assess for renal scarring are recommended. A VCUG is recommended for the neonate or infant with significant hydronephrosis on prenatal or early postnatal ultrasound.

Treatment. In nearly 80% of cases, reflux resolves spontaneously as the child grows. The factors predictive of spontaneous resolution are younger age at diagnosis, classification as grade I or II reflux, higher bladder capacity at the onset of reflux manifestations, and a history of prenatal hydronephrosis. This information is helpful in determining when and how to treat patients, given questions about safety and efficacy of continuous antibiotic prophylaxis and surgical interventions.

A conservative approach to VUR may consist of expectant management with antimicrobial therapy for acute UTIs; this strategy is appropriate for patients with low-grade VUR. Continuous antibiotic prophylaxis is an appropriate option for children with a history of UTI. Children who experience breakthrough UTIs despite continuous antibiotic therapy are candidates for invasive procedures. Endoscopic correction of reflux may be performed with subureteric injection of dextranomer/hyaluronic acid copolymer, a bulking agent that bolsters tissue around the valvular mechanism. Finally, laparoscopic and open surgical ureteral reimplantation procedures typically offer a high rate of reflux resolution.

Obstruction of the Ureteropelvic Junction

Ureteropelvic junction obstruction (UPJO) is defined as a blockage (partial or complete) in urinary flow from the renal pelvis at the entry point of one or both ureters. It is diagnosed more often in males and more frequently presents unilaterally at the left ureter. Congenital UPJO is the most common diagnosis associated with antenatal hydronephrosis. The condition has an incidence of 10% to 30% among infants found to have hydronephrosis. An autosomal-dominant model of inheritance has been proposed; research is ongoing to identify the genetic abnormalities that give rise to congenital anomalies of the kidney and urinary tract, including UPJO. An increased incidence in premature infants and twins has been noted, as well as an increased incidence of other urologic abnormalities such as horseshoe kidney in conjunction with UPJO. Acquired UPJO may be due to urinary tract stones, postoperative or inflammatory strictures, or neoplasms of the ureters.

Pathogenesis. The exact etiology of UPJO remains unknown, but in the majority of cases, stenosis (narrowing) of the junction by either intrinsic or extrinsic factors is the cause of the obstruction. Excessive collagen in the muscle cells of the junction is a usual intrinsic source of UPJO, whereas compression from renal veins or arteries "crossing vessels" is a possible external cause. The result of intrinsic and extrinsic UPJO is inefficient ureter drainage leading to progressive dilation of the renal pelvis and hydronephrosis with risk of structural damage to the renal parenchyma.

Diagnosis and clinical manifestations. Increasingly, a UPJO diagnosis is made prenatally during maternal ultrasonography. Neonates with hydronephrosis attributable to UPJO frequently have elevated serum creatinine levels and reduced urinary output, but it is not common for renal failure to be present in neonates with either unilateral or bilateral UPJO. Other early signs and symptoms include a palpable flank mass in a newborn infant; abdominal, flank, or back pain; a UTI with fever; or hematuria without significant trauma. UPJO may also be asymptomatic and discovered incidentally on renal ultrasonography. A VCUG may be performed to rule out other conditions such as vesicoureteral reflux.

Treatment. Surgical intervention for UPJO has decreased in recent decades in favor of more conservative approaches. In the case of prenatally diagnosed UPJO, observation is considered most appropriate. Expectant management includes serial diagnostic tests to monitor renal function and degree of hydronephrosis. The intent of surgery is to relieve symptoms and/or maintain renal function. The timing of surgical intervention, when found to be necessary to correct UPJO, is controversial. Early surgical repair is usually warranted if function of the affected kidney decreases, in cases of bilateral obstruction, and in cases of congenital single kidney with obstruction. Ultimately, it is a clinical decision based on the degree of obstruction, a careful analysis of kidney function, and the overall health of the infant or child.

Pyeloplasty is the surgical reconstruction of the ureteropelvic junction. It usually involves removal of the stenosed area of the junction and anastomosis of the ureter and renal pelvis. Minimally invasive laparoscopic techniques, including robotic-assisted approaches, are increasingly used and have yielded similar complication rates and lengths of hospital stay compared with open pyeloplasty.

Ureteral Ectopy

An ectopic ureter is a ureter implanted caudal to the normal location of ureteral implantation on the trigone. Alternative or duplicate sites of ureter implantation predispose the patient to UTI and a potential reduction in renal function.

Pathogenesis. Ureters may implant anywhere along the route of migration of the mesonephric duct during fetal development; ectopic implantation results from delayed or failed separation of the ureteral bud from the duct. Ectopic ureters are significantly more common in females than in males. In males, the ectopy is most frequently found implanted in the bladder neck, seminal vesicle, and vas deferens. In females, ectopic ureters may be implanted in the urethra, uterus, vagina, or cervix.

Diagnosis and clinical manifestations. An ectopic ureter is often found in conjunction with other genitourinary pathologies, including duplicate ureters and duplex or nonfunctioning kidney. Ureteral ectopy may be suspected, as with other congenital urologic anomalies, with antenatal detection of hydronephrosis on maternal ultrasonography. Postnatally, the condition may be challenging to diagnose unless the ectopic ureter arises from a functioning kidney or part of a kidney (moiety). Renal ultrasonography, VCUG, and magnetic resonance urography may be useful for diagnosis.

The clinical manifestations of ureteral ectopy vary depending on the site of implantation. Infant females are likely to present with UTI, whereas older girls more commonly exhibit incontinence. Epididymitis, orchitis, or UTI may be the initial problem in males, who are less likely to experience incontinence because the ectopic implantation is typically located above the external sphincter.

Treatment. Accompanying UTIs are treated appropriately. Surgical alternatives vary according to the site of ureteral ectopy, the function of the affected kidney(s), and the presence of any other pathology. In the case of a single ectopic ureter, when the opposing kidney is normal, nephroureterectomy is the recommended course of treatment. If the involved kidney has adequate function, the ureter may be reimplanted in a more physiologically acceptable site. Heminephrectomy and ureteropyelostomy are not uncommon. Laparoscopic techniques are increasingly used because of the reduced mortality and better visualization of the surgical field with minimally invasive procedures.

Ureterocele

A ureterocele is a congenital cystic dilation (outpouching) of the distal end of the ureter. These cystic dilations are called *intravesical* or *orthotopic* ureteroceles when they are entirely within the bladder itself and *extravesical* or *ectopic* when they extend into the neck of the bladder or the urethra. Ectopic ureteroceles are the more common form in the pediatric

population. A ureterocele with a kidney that has just one ureter is called a *single system* ureterocele; *duplex system* ureteroceles are found with a kidney possessing two ureters. Duplex system ureteroceles are more common. Ureteroceles occur more often in females than in males (6:1 ratio) and almost exclusively in Caucasians.

Etiology and pathogenesis. It is most likely that the etiology of ureteroceles is complex rather than unified. There are multiple points in embryogenesis when ureteral development could go awry and ureterocele formation occur, often concomitantly with other anomalies of the kidney(s). Ureteroceles may be classified as simple structures when they are not associated with duplicate collecting systems, but the majority present as duplicate systems with ectopic implantation. They are infrequently bilateral. The small orifice of the ureter poses an obstruction in the collecting system and results in ureteral and renal calyx dilatation, facilitating reflux and infection. If the ureterocele is large, obstruction of the bladder outlet may occur.

Diagnosis and clinical manifestations. Ureteroceles are increasingly identified serendipitously during prenatal ultrasonography by the appearance of hydronephrosis and evidence of cystic dilatation. After birth, confirmation with further assessment is necessary. UTIs are the most common manifestation in infants. In addition to UTI, patients with ureteroceles may present with either urinary retention or UI if the bladder outlet is obstructed, ureteral calculus, hematuria, urosepsis, or a general failure to thrive.

Anatomically, ureteroceles present with significant variability, requiring a thorough history, physical examination, and diagnostic testing. Ultrasonography is recommended for detection of the ureteral dilatation. The structure and function of the urinary tract, including the presence of associated conditions such as VUR, are also evaluated with intravenous pyelograms, VCUGs, magnetic resonance urograms, and nuclear renal scans. In addition to determining the precise characteristics of the ureterocele, diagnostic test results assist in determining the most appropriate intervention.

Treatment. Treatment goals for ureterocele include controlling infection, preserving urinary tract function, maintaining urinary continence, and removing obstruction. Expectant management may be appropriate in select patients, particularly children without obstruction. Continuous antibiotic prophylaxis may be prescribed. Endoscopic ureterocele decompression by incision or puncture is recommended in the setting of infection or severe obstruction. Surgical procedures may be selected depending on multiple factors, including patient age and clinical condition, kidney function, and presence of obstruction or reflux. Depending on the anomalies found during diagnostic evaluation, the intervention may include uretero-ureterostomy, excision with reimplantation of the ureter, or partial nephrectomy and ureterectomy. In an acutely septic patient, a percutaneous nephrostomy to drain the upper collecting system may be needed.

KEY POINTS

- Congenital abnormalities of the bladder include misimplantation of ureters, strictures, duplicate ureter, and ureterocele.
- These disorders cause problems by obstructing normal urine flow and predisposing to retrograde urine flow, urinary stasis, and secondary infection; treatments range from expectant management to surgical interventions.

NEOPLASMS

Primary cancers of the ureters or urethra are rare. Bladder cancer, however, is quite common.

Bladder Cancer

Bladder cancer is the fifth most common cancer diagnosed in the United States (after breast, lung, prostate, and colorectal cancers). It accounts for 4.5% of all new cancer cases, with an estimated 74,000 new cases in 2015. An estimated 430,000 patients worldwide receive the diagnosis every year, with the highest incidence in the Middle East, North Africa, and Europe. Bladder cancer is more prevalent in males, older adults, and whites. The overall median age at diagnosis in the United States is 73 years, with 9 of 10 cases occurring in people older than 55 years of age. The 5-year survival rate for people with bladder cancer is 77.4%; the median age at death is 79 years. However, African Americans have a significantly lower 5-year survival rate. At diagnosis, about half of bladder cancer patients have in situ cancers, and about 35% have localized bladder cancer.

More than 90% of all cases of bladder cancer are *urothelial (transitional cell)* carcinomas originating in the transitional epithelium, or urothelium. There are two subtypes of urothelial carcinomas; the majority exhibit a papillary appearance (slender projections into the bladder), with the remainder having a flat appearance. Bladder cancer is said to be noninvasive if it remains in the transitional (urothelial) cells; invasive cancer has spread into the lamina propria layer or even deeper into the bladder muscle. The flat subtype of urothelial lesions tends to be more muscle invasive and thus has a poorer prognosis.

Other types of bladder cancer are much less common. *Squamous cell* carcinoma of the bladder represents only approximately 1% to 2% of the overall cases in the United States. This type of bladder cancer is microscopically similar to skin cancer and is quite invasive. About 1% of bladder cancers are *adenocarcinomas.* These tumors, arising from glandular tissue, are also invasive in nature. *Small cell* bladder cancers account for less than 1% of all bladder cancers; these cancers arise from neuroendocrine cells. A very rare cancer called *sarcoma* originates in the muscle of the bladder.

Risk factors. Bladder cancer risk usually is divided into environmental and inherited classifications. Environmental risk factors for bladder cancer are shown in Box 29.2. Smoking is the greatest risk factor, increasing risk by 3.89 in men and 4.65 in women. Carcinogenic chemicals in cigarette smoke are inhaled, enter the bloodstream, and then are filtered and concentrated in the urinary tract. About half of the bladder tumors in men and women are attributable to smoking. Concurrent occupational exposure to carcinogens and smoking may increase bladder cancer risk synergistically. People with occupational exposure to aniline (an aromatic amine) textile dyes, hair dyes (especially for longer than 10 years), paint and leather, carpet, rubber, and cement are at increased risk for bladder cancer. Exposure to arsenic in water is a significant risk factor in some parts of the world. Chronic inflammation of the bladder by UTIs is associated with an increased risk of bladder cancer, especially

BOX 29.2 Environmental Risk Factors for Bladder Cancer

Tobacco smoking (aromatic amines)
Certain types of dyes, including hair, medical, and industrial dyes (aromatic amines)
Certain chemicals used in rubber tire production
Certain pesticides that contain aromatic amines
Certain chemotherapeutic agents (e.g., cyclophosphamide)
Diesel exhaust (polycyclic aromatic hydrocarbons)
Arsenic in drinking water
Low fluid intake
Pelvic radiation therapy

the invasive squamous cell type. Other causes of inflammation, such as stones, have also been associated with the disease. Certain chemotherapy agents (e.g., cyclophosphamide, ifosfamide) used in the treatment of other malignancies, as well as radiation therapy to the pelvis, increase the risk of bladder cancer. Low fluid intake is a risk factor for bladder cancer. A family history of bladder cancer increases a person's risk, likely through inherited gene syndromes and/or shared exposure to environmental carcinogens. One heritable risk factor is a mutation of the retinoblastoma gene responsible for an infantile cancer of the eye that is associated with increased bladder cancer risk.

In children, the congenital anomaly of exstrophy of the bladder (bladder outside the abdominal cavity) may predispose to development of bladder tumors, particularly adenocarcinomas. Parasitic infections from schistosomiasis, prevalent in the Middle East and Africa (especially Egypt), have been associated with squamous cell carcinoma as a result of urine-borne carcinogens formed during the infectious process and irritation by the parasitic ova.

Pathogenesis. The pathways by which urothelial carcinomas develop differ by tumor type. Papillary non–muscle-invasive tumors are associated with activation of the cellular growth pathway involving Ras, a G-protein, and receptor tyrosine kinase. Other mechanisms, including alterations of the pathways involving retinoblastoma 1 (RB1) and tumor suppressor protein p53 (TP53), disrupt normal cell replication and produce muscle-invasive tumors.

Bladder cancer metastasis occurs directly through the bladder wall to adjacent organs (e.g., prostate, bowel, vagina, uterus) or via lymph nodes in the pelvis and abdomen. Once treated, tumors can recur at the original site, or an entirely new tumor may develop at another site. The sites of metastasis most commonly include the lymph nodes, liver, lungs, and bone. Specific details of the pathogenesis of cancer are found in Chapter 7.

Clinical manifestations. Painless hematuria is usually the initial sign of bladder cancer. However, in the early stages, both gross and microscopic hematuria are often intermittent, and hematuria is associated with numerous other urinary tract pathologies. Other manifestations related to bladder cancer include urinary frequency, urgency, and dysuria. But again, all these symptoms are seen with other conditions involving the urinary tract, including UTI. Because the signs and symptoms are so indefinite, early diagnosis of bladder cancer requires a thorough workup by the health care provider, especially in persons more than 40 years of age.

Diagnosis. There is insufficient evidence regarding routine screening for bladder cancer in asymptomatic adults to determine the balance of harm with benefit. In most cases, diagnostic evaluation for bladder cancer is initiated because of the development of one or more of the previously listed clinical manifestations. A thorough history may reveal risk factors for bladder cancer, or physical examination and diagnostic testing may indicate another etiology.

The recommended diagnostic test for suspected bladder cancer is cystoscopy, with biopsy of any questionable tissue and washings of free cells for cytologic examination. White light cystoscopy allows direct visualization of tumors. Transurethral resection of bladder tumors (TURBT) is recommended during cystoscopy to obtain samples of tumor and detrusor muscle as well as to resect visible tumors. Urine specimens may be tested for the presence of tumor markers. However, the sensitivity and specificity of available tumor markers vary widely; at this time, no tumor marker is reliable enough to replace cystoscopy. Additional imaging tests may include CT or MRI scans and fluorodeoxyglucose positron emission tomography CT to assist with staging, especially for muscle-invasive cancer. The stage of the malignancy is an important contributor to treatment decisions and aids in the determination of prognosis. The TNM (tumor, node, metastasis) staging system is commonly used for bladder cancer. Examination of biopsied tissue, cells obtained during cystoscopy, and cells found in the urine is essential to determining the specific type and grade of the tumor. (See Chapter 7 for a discussion of staging and grading.) All of this information is used to guide the selection of the specific treatment approach (Fig. 29.3).

Treatment. Treatment protocols are based on the tumor's features: the type of bladder cancer and its grade and stage. More aggressive treatment is needed for muscle-invasive cancers. Non–muscle-invasive tumors have a lower progression rate but tend to recur; they require lifelong surveillance and may require retreatment. The primary options are surgery, radiation therapy, chemotherapy, and immunotherapy. Treatment selection is evidence based, with available information continuously being updated. Surgery as the single treatment, or with other adjuvants, is the intervention for the majority of bladder cancers, especially early stage. The specific approach to surgery varies with the stage of the tumor. For non–muscle-invasive carcinoma, endoscopic TURBT with appropriate cystoscopic and cytologic follow-up is recommended. Adjuvant intravesical instillation of mitomycin C has been shown to reduce recurrence. Intravesical immunotherapy with Bacillus Calmette-Guérin (BCG) may be used for high-risk non–muscle-invasive bladder cancer.

If the tumor is at high risk for progression, unresponsive to BCG, or muscle invasive, more extensive surgical procedures are employed. Neoadjuvant chemotherapy may be given preoperatively to reduce tumor size and improve survival. Radical radiotherapy may also be offered. Surgical removal of the bladder (cystectomy), either partial or total, may be performed; a total cystectomy requires urinary diversion to provide for storage and elimination of urine. Urinary diversions typically involve the creation of a reservoir, using a portion of the intestine, into which the ureters are implanted. Urine is drained from the reservoir, often through a stoma created through the abdominal wall. However, substitute bladder reservoir (neobladder) procedures involve connecting the urethra to the reservoir, allowing the patient to void normally. In cases where the tumor is large or there are multiple bladder tumors, a radical cystectomy is performed, in which the bladder and surrounding nodes are removed; in men the prostate gland is also removed, and in women the uterus, ovaries, fallopian tubes, and part of the vagina are also often removed. Systemic chemotherapy may be an option for people with metastatic bladder cancer.

The prognosis for stage 0 (noninvasive papillary carcinoma or carcinoma in situ) is excellent, with a relative 5-year survival rate of 98%. As the tumor invades deeper into the bladder wall or metastasizes beyond the bladder, the survival rate declines. The 5-year relative survival rate for muscle-invasive (stage II) bladder cancer is 63% and drops to 15% for metastatic (stage IV) cancer. Risk of recurrence of bladder cancer is high, even for superficial malignancies; ongoing follow-up with periodic diagnostic tests is necessary.

KEY POINTS

- Bladder cancer is the fifth most common cancer diagnosed in the United States.
- The risk of developing bladder cancer increases with age; smoking and occupational exposure to carcinogenic chemicals are thought to be the main predisposing factors.
- The types of bladder cancer are differentiated by their tissue of origin and histologic appearance. About 95% of bladder tumors originate from the transitional epithelium (urothelium) lining the urinary tract.
- Bladder cancer is primarily manifested as hematuria. Frequency, urgency, and dysuria may also be present.

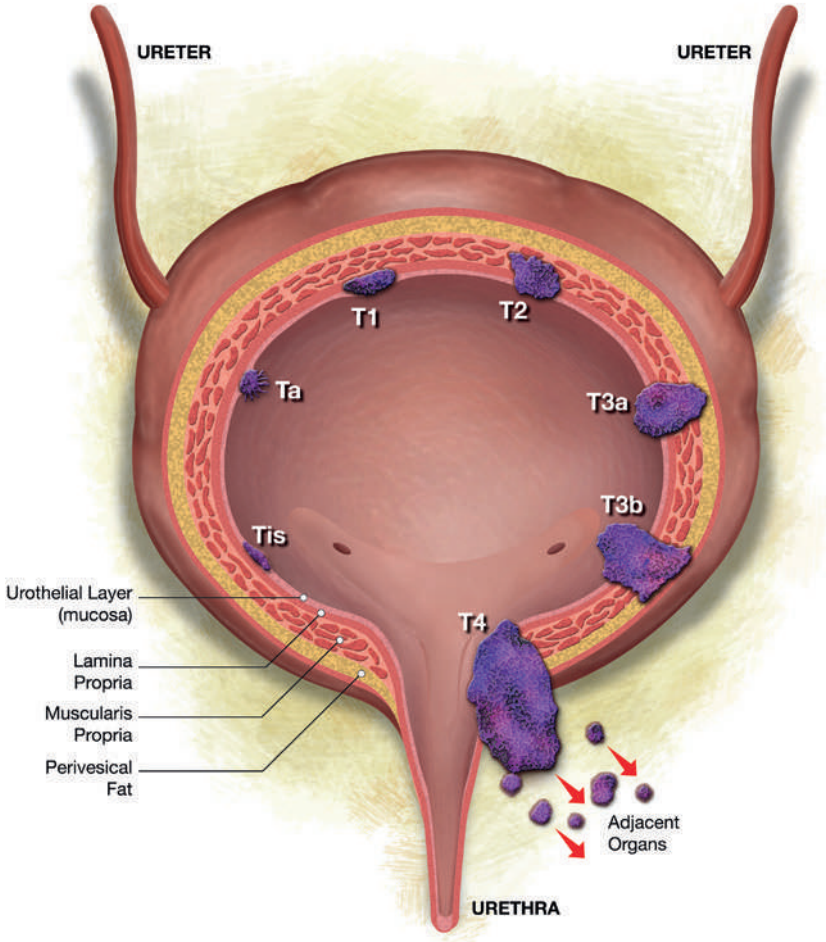


FIG 29.3 TNM staging of bladder tumors. (From Harshman LC et al: Diagnosis of bladder cancer. *Surg Pathol Clin* 2015;8(4):677-685, Elsevier.)

- The most frequently used invasive test in the diagnosis of bladder cancer is cystoscopy, with biopsy of any questionable tissue and washings of free cells for cytologic examination.
- Treatment protocols are based on the tumor's features: the type of bladder cancer and its grade and stage. The primary options are chemotherapy, immunotherapy, surgery, and radiation therapy.

INFLAMMATION AND INFECTION

Most inflammations of the LUT are due to infection. The normal defense mechanisms of the urinary tract are presented in detail in Chapter 27 with the discussion of pyelonephritis (infection of the kidney). UTIs are typically ascending in nature; they begin in the LUT and may progress to the kidney (upper urinary tract). By the time infection reaches the kidneys, the bladder and urethra are already infected. Involvement of the ureters does not seem to present with clinical manifestations.

Urethritis

Urethritis is an inflammation of the urethra. It can be caused by infection, external irritants, and, in women, insufficient estrogen levels. Infection of the urethra may be due to a wide variety of organisms. The most common sexually transmitted causative organisms are *Neisseria gonorrheae* and *Chlamydia trachomatis.* Urethritis is the most commonly occurring sexually transmitted infection (STI) in men. Other causative organisms of nonchlamydial nongonococcal urethritis include *Mycoplasma genitalum* and *Trichomonas vaginalis.* Postmenopausal women are at increased risk for irritation and inflammation of the urethra due to reduced estrogen levels; the urethra is an estrogen-dependent structure.

Inflammation of the urethra may lead to pain, dysuria, urethral discharge, and abscess, and if it is attributable to an infectious organism, it may progress to infective cystitis. Men may also experience epididymitis and prostatitis. Some people are asymptomatic; if the cause is an STI, they may not know they are infected until notified that their partner has an infection. If urethritis is due to infection, targeted pharmacologic therapy is indicated. Application of topical estrogen to the urethral opening helps maintain postmenopausal mucosal health.

Cystitis

Cystitis, or inflammation of the bladder lining, may result from bacterial, fungal, or parasitic infections; chemical irritants; foreign bodies (e.g., stones); or trauma. By far the most common cause of cystitis—and the focus of this discussion—is bacterial infection.

UTIs are highly prevalent; they represent the most common bacterial infections seen in outpatient settings, with annual treatment costs at $2 to $3 billion in the United States alone. Only respiratory conditions prompt more pediatric office visits than cystitis. *Uncomplicated UTIs* are infections not associated with abnormalities in LUT anatomy or function, urinary catheterization, pregnancy, or diseases such as diabetes.

Complicated UTIs are those in which the host is immunocompromised or has genitourinary abnormalities or that are caused by multidrug-resistant bacteria. Incidence of UTI is much higher in women and girls, with a 4:1 ratio of occurrence in girls compared with boys. An estimated 11% of women report having had a UTI each year. A shorter urethra, as well as a colonization route from both the rectum and the vagina to the urethra, is thought to explain the increased rate of infections in women. Prostatic secretions, which are antibacterial, inhibit cystitis in men younger than 50 years. The incidence of complicated UTIs is relatively higher in men, however, especially in older adult males. UTIs are common occurrences during pregnancy and have been associated with an increased risk of premature delivery. *Recurrent UTIs* are repeated infections within a short period after verified resolution of the earlier infection. Approximately one-quarter of young women with a bladder infection experience a recurrent UTI within 6 months of their initial UTI. The shorter the time frame between the initial and recurrent UTI, the greater the likelihood of a same-strain recurrent infection. Rates of bacteriuria significantly increase with age.

Risk factors are listed in detail in Box 29.3. They include factors that reduce the flow of urine such as decreased fluid intake and bladder outlet obstruction or other conditions that impair bladder emptying or cause urinary retention (e.g., neurogenic bladder). Additional risk factors include urinary catheterization and estrogen depletion. Use of spermicidal agents with diaphragms, recent sexual intercourse, and recurrent UTIs also increase risk. UTIs in older adults are associated with increased risk of falls and sepsis. Investigation continues into the association between low-bacterial-count cystitis and detrusor overactivity, as well as between recurrent UTIs and bladder cancer.

BOX 29.3 Risk Factors for Cystitis

Factors Increasing Urinary Stasis
Intrinsic obstruction (stone, tumor of urinary tract, urethral stricture, BPH)
Extrinsic obstruction (tumor, fibrosis compressing urinary tract)
Urinary retention (e.g., neurogenic bladder)
Renal impairment

Foreign Bodies
Urinary tract calculi
Catheters (indwelling, external condom catheter, urethral stent, nephrostomy tube, intermittent catheterization)
Urinary tract instrumentation (cystoscopy)

Anatomic Factors
Congenital defects leading to obstruction or urinary stasis
Fistula (abnormal opening) exposing urinary stream to skin, vagina, or fecal stream
Shorter female urethra and colonization from normal vaginal flora
Obesity

Factors Compromising Immune Response
Aging
Human immunodeficiency virus infection
Diabetes mellitus

Functional Disorders
Constipation
Voiding dysfunction with detrusor sphincter dyssynergia

Other Factors
Pregnancy
Multiple sex partners (women)
Use of spermicidal agents, contraceptive diaphragm (women), bubble baths, feminine sprays
Poor personal hygiene
Habitual delay of urination ("nurse's bladder," "teacher's bladder")

From Lewis S et al: *Medical-surgical nursing: assessment and management of clinical problems*, ed 10, St Louis, MO, 2017, Elsevier.
BPH, Benign prostatic hyperplasia.

Etiology and pathogenesis. Normally, bacteria are cleared from the bladder by the flushing and dilutional effects of voiding. The high urea concentration with the high osmolarity and low pH of urine act as natural barriers to invading pathogens. Additionally, the mucous layer lining the bladder, as well as the ability of urothelial cells to initiate an immune response, serve as host defenses.

Escherichia coli is responsible for 85% of community cases of bacterial cystitis. The microbial characteristics of *E. coli* promote its adherence to bladder epithelium, colonization, and ability to evade the host's immune response while invading host cells and replicating within them, creating intracellular bacterial communities that are released during epithelial cell apoptosis and invade more cells.

Clinical manifestations. The majority of patients with cystitis experience an acute onset of frequency, urgency, and dysuria; pain may be present in the suprapubic area. The urine may appear pink because of hematuria or cloudy as a result of the infectious organism. Infants and small children may exhibit fever, irritability, poor feeding, vomiting, diarrhea, and lethargy. Additional signs of UTI in older adults include delirium and new-onset incontinence. Untreated cystitis may lead to renal damage secondary to ascension of the infection to the upper urinary tract. Therefore prompt intervention is essential.

Diagnosis and treatment. Screening for suspected UTI is often performed using a clean-catch urine specimen and a simple nitrite and leukocyte esterase dipstick test. Nitrites are present in UTI because many urinary pathogens produce enzymes that reduce nitrates to nitrites; the presence of leukocyte esterase is indicative of pyuria (white blood cells in the urine). The dipstick test is quite sensitive in detecting UTIs. Febrile infants and small children may need to be catheterized to obtain a reliable urine specimen. A urine culture is not always necessary for the diagnosis of UTI in adults; uncomplicated infections, caused by common organisms, may be diagnosed on the basis of symptoms and a positive dipstick test. Urine culture is recommended in situations in which the patient has manifestations suggesting pyelonephritis, fails to respond to empirical pharmacologic therapy, or presents with atypical symptoms. *Unresolved infections* are those in which bacteriuria remains after the initial treatment. Under these circumstances, cultures precisely identify organisms and permit targeted pharmacologic therapy.

Men, children, and those women with recurrent infections, unresolved infections, atypical manifestations, or other problems previously described have complicated UTIs and require physical examination and diagnostic testing. Recurrent infections in men should trigger a urologic examination that may include endoscopy and a CT scan.

Treatment algorithms for acute cystitis and recurrent UTIs are shown in Figs. 29.4 and 29.5, respectively. Acute, uncomplicated UTIs in women with no anatomic anomalies of the urinary tract, with no recent history of cystitis, and with no urologic comorbidities may be treated empirically with a 3-day course of trimethoprim-sulfamethoxazole or a 5-day course of nitrofurantoin. Complicated cystitis is managed with a longer course of antibiotics, 7 to 10 days or longer depending on causative organism, comorbidities, and whether it is an unresolved or recurrent infection. Antibiotic therapy in pregnant women must be carefully chosen to avoid fetal harm. Routine follow-up urine culture and sensitivity are not typically necessary for uncomplicated UTI but may be warranted with complicated UTIs to ensure antibiotic effectiveness. Women with recurrent cystitis may be candidates for several months of antibiotic prophylaxis; women with postcoital recurrence may benefit from self-administering a single dose of antibiotic after sexual intercourse.

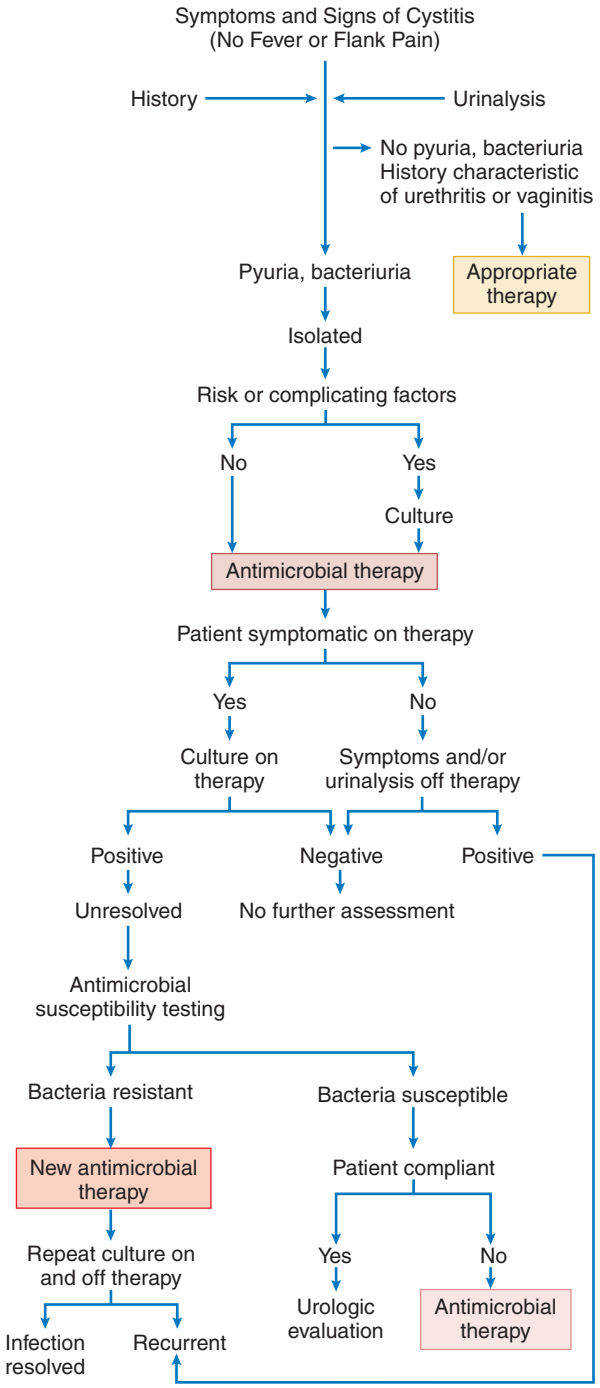


FIG 29.4 Management of acute cystitis. (From Wein AJ et al, editors: *Campbell-Walsh urology,* ed 11, St Louis, MO, 2016, Elsevier.)

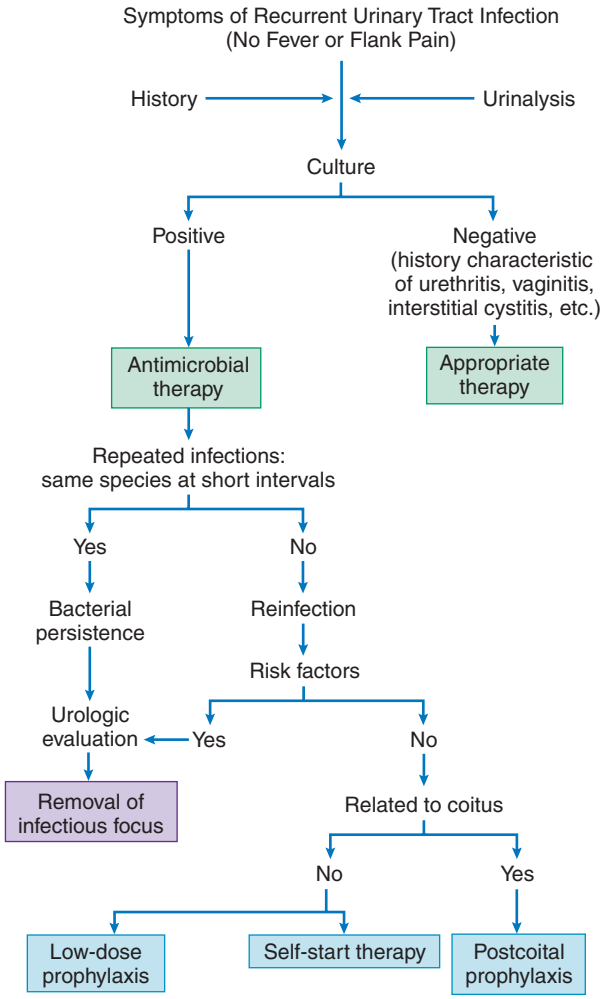


FIG 29.5 Management of recurrent UTI. (From Wein AJ et al, editors: *Campbell-Walsh urology,* ed 11, St Louis, 2016, Elsevier.)

In children, administration of antibiotics for 7 to 14 days is the recommended treatment, with the selection of drug made according to sensitivity of the infecting organism. UTIs in children may be indicative of an underlying pathologic process (e.g., VUR, UPJO) and warrant additional urologic evaluation. There is a higher incidence of asymptomatic bacteriuria in older adult women. Bacteriuria without clinical signs of infection does not warrant antibiotic treatment. Antibiotic therapy for UTI in older adults must be prescribed with attention to underlying renal function and risk for interaction with other medications being taken. Finally, all patients will benefit from education on UTI prevention, including maintaining an adequate fluid intake.

KEY POINTS

- Urethritis is inflammation of the urethra. It is most often due to infection; common causative organisms are transmitted sexually. Urethritis may also be due to external factors such as frequent catheterizations or poor personal hygiene. Treatment depends on the cause.
- Cystitis is an inflammation of the bladder lining that may be due to infection, chemical irritants, stones, or trauma. Most cases have an infectious etiology.
- Factors predisposing to cystitis include female gender, increased age, catheterization, diabetes mellitus, bladder dysfunction, and any disorder causing urinary stasis. Manifestations include frequency, urgency, dysuria, possible suprapubic pain, and cloudy urine, although the infection may be asymptomatic. Most women with uncomplicated cystitis are treated empirically based on symptoms and positive dipstick test. More complicated situations, as well as cystitis in men and children, may require urine culture and/or further assessment.
- Symptoms of cystitis in older adults may be atypical and include confusion and new-onset incontinence. Antibiotic treatment for symptomatic cystitis in the elderly should be managed with close drug monitoring to avoid toxicity. Asymptomatic bacteriuria in the elderly should not be treated with antibiotics.

OBSTRUCTION

Obstruction of the LUT may develop when a stone obstructs urine transport. Stones in the LUT produce some manifestations that are similar to stones formed and lodged in the kidney.

Lower Urinary Tract Calculi

Stones, or calculi, usually form in the kidneys *(renal calculi* or *nephrolithiasis)*. Renal calculi are discussed in Chapter 27. The term *urolithiasis* is used for stones forming anywhere in the urinary tract, primarily in the kidney or ureters, as well as those that form or travel into the ureters, bladder, or urethra. The manifestations of stones in the ureters and bladder vary somewhat from those associated with kidney stones, but the risk factors, stone composition, diagnostic tests, and treatment aspects are essentially the same. Refer to Chapter 27 for a detailed discussion of these topics. If diagnosis and treatment are not completed in a timely manner, the patient is at risk for the development of postrenal acute kidney injury (Chapter 28).

Ureteral Calculi

Calculi in the ureters are usually those that were able to pass from the kidney through the ureteropelvic junction, as opposed to having formed in the ureter. In many cases, ureteral stones are small, pass easily through the ureters, and cause no clinical manifestations. Calculi with a diameter of ≤5 mm have an estimated 68% probability of being spontaneously expelled. The ureteropelvic and ureterovesical junctions are locations at which larger calculi may become lodged. When a larger stone becomes wedged in the ureter, manifestations are those characteristic of renal calculi. Ureteral stretching and spasm contribute to pain that is termed *ureteral colic,* a sharp, spasmodic pain experienced in the flank area, possibly radiating into the umbilical region. When the stone approaches the distal portion of the ureters, the pain may continue to be sharp in quality but intermittent in nature. Alternatively, the pain may increase, with radiation into the groin region, testicles, or labia. Irritation of the ureters may result in hematuria. Manifestations associated with ureteral colic include tachycardia, tachypnea, diaphoresis, nausea, and vomiting.

The recommended initial diagnostic procedure is CT of the abdomen and pelvis without contrast; ultrasound color Doppler may be used for pregnant women. Ureteral stones less than 1 mm in diameter with no evidence of accompanying sepsis warrant observation with pain management while waiting for spontaneous passage. However, failure of the stone to move or unrelenting pain are indications for stone removal. Medical expulsion therapy using α1-adrenergic antagonists such as tamsulosin may facilitate passage of stones. These drugs decrease ureteral muscle tone and decrease the frequency of peristalsis, relaxing the ureteral wall. Larger stones (>10 mm) may require stone removal; shockwave lithotripsy and ureteroscopy are effective treatments for both adults and children. Percutaneous ureteroscopy may be used for larger impacted stones. Either laparoscopic or open surgical removal of stones may be used in cases of first-line treatment failure. Additional therapeutic and preventive interventions for urinary tract calculi are presented in Chapter 27.

Bladder (Vesical) Calculi

Bladder calculi represent only about 5% of urinary tract stones in Western countries; the incidence has decreased in recent decades. Bladder stones have been classified into three categories. *Migrant* vesical calculi form in the upper urinary tract and migrate to the bladder. *Primary* bladder calculus formation occurs in the bladder absent of underlying pathology. More common in children, the etiology appears to be a combination of metabolic (low urinary phosphate, increased calcium oxalate, uric acid, and ammonia excretion) and nutritional (vitamin and mineral deficiencies, protein-poor diet) factors. Primary bladder stones tend to be composed of uric acid or calcium oxalate. Finally, *secondary* bladder stones arise in association with a range of conditions, including bladder outlet obstruction with urinary stasis (e.g., benign prostatic hyperplasia), neurogenic bladder due to spinal cord injury, chronic bladder infection, and foreign bodies in the bladder. There is evidence that bladder outlet obstruction is a causative factor in the majority of secondary bladder calculi.

Some patients with bladder stones are asymptomatic; those with symptoms may experience dysuria, suprapubic pain, and urinary hesitancy. A stone in the bladder may irritate the urothelium and result in hematuria; once the bladder neck or urethral orifice is reached, manifestations of obstruction may appear. Diagnostic tests are similar to those for calculi in other segments of the urinary tract. Treatment includes antimicrobial therapy for infection. Stone expulsion is often spontaneous; stones causing obstruction pose a risk for postrenal acute kidney injury and require prompt intervention. Stone removal procedures include shockwave lithotripsy, transurethral lithotripsy using ultrasonic or laser techniques, percutaneous cystolithotomy, and open cystolithotomy for stones ≥4 cm. The choice of procedure depends on patient and calculus characteristics.

KEY POINTS

- Ureteral calculus is most often due to stones traveling from the kidney.
- Manifestations of ureteral calculus include colicky pain; associated signs and symptoms may include hematuria, tachycardia, tachypnea, diaphoresis, nausea, and vomiting.
- Bladder calculi may be migrant or form in the bladder as primary or secondary stones.
- Manifestations of bladder stones are often limited to dysuria and pain; hematuria is possible. Obstructive signs and symptoms will present if stones obstruct the bladder neck or urethral orifice.
- Ureteral and bladder calculi are similar to nephrolithiasis in terms of risk factors and stone characteristics. If infection is present, it is treated with appropriate antimicrobials. Several interventions are available for stones that do not pass spontaneously.

SUMMARY

Disorders involving the LUT are quite common. In adults, the most prevalent pathologic conditions are infections, incontinence and overactive bladder syndrome, malignancies, and urinary stones. Incontinence as a result of bladder dysfunction affects millions of adults. The prevalence of incontinence increases with age and is more common in women than in men. However, incontinence is not a result of the normal aging process, and patients should be urged to seek evaluation and treatment. Behavioral, pharmacologic, and surgical approaches may be used to manage voiding dysfunction. IC/PBS is a chronic condition consisting of bladder pain and a variety of other manifestations typically associated with other bladder pathologies but with no other identifiable pathology. Treatment is individualized and multimodal. In the United States bladder cancer is the fifth most common type of cancer in adults. The primary clinical manifestation of bladder cancer is hematuria. Treatment is based on stage, grade, and type of cancer. Interventions include surgery, radiation, and chemotherapy.

LUT infections are common; they require appropriate treatment to prevent ascending infection and kidney involvement. Both childhood UTIs and hydronephrosis should trigger more detailed urologic evaluation to rule out congenital disorders such as UPJO, ectopic ureters, ureterocele, and VUR. VUR is the most prevalent congenital LUT disorder. It predisposes children to UTIs, which can result in kidney scarring and permanent renal impairment. Most cases of VUR resolve spontaneously as the child ages. Before resolution, close medical management is necessary to prevent upper UTI and kidney damage.

Stones in the LUT usually arrive there after being formed in the kidney. Risk factors and treatment are much the same as those for nephrolithiasis; however, manifestations of calculi in the ureters or bladder differ.

RESOURCES

Urinary Physiology and Diagnostic Testing

Abrams P, et al: The standardization of terminology of lower urinary tract function, *Neurourol Urodyn* 21:167–178, 2002.

Brown ET, Krlin RM, Winters JC: Urodynamics: examining the current role of UDS testing. What is the role of urodynamic testing in light of recent AUA Urodynamics and Overactive Bladder Guidelines and the VALUE study? *Curr Urol Rep* 14:403-408, 2013.

Casey G: Incontinence and retention—how the bladder misfunctions, *New Zealand Nurs J* 17:26–31, 2011.

Gray M: Traces: making sense of urodynamics testing – part 10: evaluation of micturition via the voiding pressure-flow study, *Urol Nurs* 32(2):71-78, 2012.

Hall JE The urinary system. In: JE Hall, ed. *Guyton and Hall textbook of medical physiology*. ed 13. St. Louis, MO: Elsevier; 2016:26.

Mangera A, Osman NI, Chapple CR: Anatomy of the lower urinary tract, *Surgery (Oxford)* 31(7), 2013.

National Guideline Clearinghouse (NGC). *Guideline summary: bladder cancer: diagnosis and management*. Agency for Healthcare Research and Quality (AHRQ). Available at http://www.guideline.gov/content.aspx?id=49119&search=bladder+cancer+diagnosis+and+management. Summary completed 8/28/15. Accessed 31 October 2015.

O'Neill WC, et al: Renal relevant radiology: use of ultrasound in kidney disease and nephrology procedures, *Clin J Am Soc Nephrol* 9:373-381, 2014

Pagana KD, Pagana TJ: *MOSBY's manual of diagnostic and laboratory tests*. ed 5. St Louis, MO: Mosby; 2014.

Urinary Incontinence and Overactive Bladder

Abrams P, et al: Reviewing the ICS 2002 Terminology Report: the ongoing debate, *Neurourol Urodyn* 28:287, 2009.

Austin PF, et al: The standardization of terminology of lower urinary tract function in children and adolescents: update report from the standardization committee of the International Children's Continence Society, *Neurourol Urodyn* 9999:1-11, 2015.

Coyne KS, Sexton CC, Bell JA, et al: The prevalence of lower urinary tract symptoms (LUTS) and overactive bladder (OAB) by racial/ethnic group and age: results from OAB-POLL, *Neurourol Urodyn* 32:230-237, 2013.

Coyne KS, Sexton CC, Vats V, et al: National community prevalence of overactive bladder in the United States stratified by sex and age, *Urology* 77:1081-1087, 2011.

DiBianco JM, Morley C, Al-Omar O: Nocturnal enuresis: a topic review and institution experience, *Avicenna J Med* 4(4):77-87, 2014.

DuBeau CE, Kuchel GA, Johnson T II, et al: Incontinence in the frail elderly: report from the 4th International Consultation on Incontinence, *Neurourol Urodyn* 19:165-178, 2010.

Franco I: Functional bladder problems in children, pathophysiology, diagnosis, and treatment, *Pediatr Clin N Am* 59:783-817, 2012.

Gibson W, Wagg A: New horizons: urinary incontinence in older people, *Age Ageing* 43:157-163, 2014.

Gorina Y, Schappert S, Bercovitz A, et al: Prevalence of incontinence among older Americans. National Center for Health Statistics. *Vital Health Stat* 3(36), 2014.

Griebling TL: Overactive bladder in elderly men: epidemiology, evaluation, clinical effects, and management, *Curr Urol Rep* 14:418-425, 2013.

Hanna-Mitchell AT, Kashyap M, Chan WV, et al: Pathophysiology of idiopathic overactive bladder and the success of treatment: a systematic review from ICI-RS 2013, *Neurourol Urodyn* 33:611-617, 2014.

Hillard T: The postmenopausal bladder, *Menopause Int* 16:74–80, 2010.

Khandalwal C, Kistler C: Diagnosis of urinary incontinence, *Am Fam Physician* 87(8):543-550, 2013.

Manella P, Palla G, Bellini M, et al: The female pelvic floor through midlife and aging, *Maturitas*, 76:230-234, 2013.

Maternik M, Krzeminska K, Zurowska A: The management of childhood urinary incontinence, *Pediatr Nephrol* 3:41-50, 2015.

Natalin R, Lorenzetti F, Dambros M: Management of OAB in those over age 65, *Curr Urol Rep* 14:379-385, 2013.

National Guideline Clearinghouse (NGC). *Guideline summary: nonsurgical management of urinary incontinence in women: a clinical practice guideline from the American College of Physicians*. Agency for Healthcare Research and Quality (AHRQ). Available at http://www.guideline.gov/content.aspx?id = 48543&search = nonsurgical + management + of + urinary + incontinence + in + women. Summary completed 10.22.14. Accessed 31 October 2015.

National Guideline Clearinghouse (NGC). *Guideline summary: lower urinary tract symptoms in men: assessment and management*. Agency for Healthcare Research and Quality (AHRQ). Available at http://www.guideline.gov/content.aspx?id = 49273&search = lower + urinary + tract + symptoms. Summary completed 1/5/11. Updated 8/6/15. Accessed 8 November 2015.

National Guideline Clearinghouse (NGC). *Guideline summary: urinary incontinence: the management of urinary incontinence in women*. Agency for Healthcare Research and Quality (AHRQ). Available at http://www.guideline.gov/content.aspx?id = 47099&search = incontinence + women. Summary completed 3/30/2009. Updated 11/27/13. Accessed 8 November 2015.

National Guideline Clearinghouse (NGC). *Guideline summary: nocturnal enuresis: the management of bedwetting in children and young people*. Agency for Healthcare Research and Quality (AHRQ). Available at http://www.guideline.gov/content.aspx?id = 25680&search = nocturnal + enuresis. Summary completed 6/24/11. Accessed 8 November 2015.

National Guideline Clearinghouse (NGC). *Guideline summary: diagnosis and treatment of overactive bladder (non-neurogenic) in adults: AUA/SUFU guideline*. Agency for Healthcare Research and Quality. Available at http://www.guideline.gov/content.aspx?id=48226&search=diagnosis+treatment+overactive+bladder. Summary created 6/4/12. Summary updated 9/8/14. Accessed 8 November 2015.

Rahn DD, Roshanravan SM: Pathophysiology of urinary incontinence, voiding dysfunction, and overactive bladder, *Obstet Gynecol Clin North Am* 36:463–474, 2009.

Testa A: Understanding urinary incontinence in adults, *Urol Nurs* 35(2):82-86, 2015.

Bladder Pain, Dysfunction and Congenital Anomaly

Altobelli E, Nappo SG, Guidotti M, et al: Vesicoureteral reflux in pediatric age: where are we today?, *Urologia* 81(2):76-87, 2014.

American Urological Association (AUA). *Diagnosis and treatment of interstitial cystitis/bladder pain syndrome*, 2014. Available at https://www.auanet.org/education/guidelines/ic-bladder-pain-syndrome.cfm. Accessed 10 January 2016.

American Urological Association (AUA): *Management and screening of primary vesicoureteral reflux in children: AUA guideline*, 2010. Available at

https://www.auanet.org/common/pdf/education/clinical-guidance/Vesicoureteral-Reflux-a.pdf. Accessed 10 January 2016.
American Urological Association (AUA). *Pediatric urinary tract infections*. Available at https://www.auanet.org/education/pediatric-urinary-tract-infections.cfm. Updated June 2012. Accessed 11 January 2016.
Arlen AM, Cooper CS: Controversies in the management of vesicoureteral reflux, *Curr Urol Rep* 2014;16:64. doi:10.1007/s11934-015-0538-2.
Chowdhary SK, Kandpal DK, Sibal A, et al: Management of complicated ureteroceles: different modalities of treatment and long-term outcome, *J Indian Assoc Pediatr Surg* 19(3):156-161, 2014.
Chrysanthopoulou EL, Doumouchtsis SK: Challenges and current evidence on the management of bladder pain syndrome, *Neurourol Urodyn* 33:1193-1201, 2014.
Dinis S, de Oliveira JT, Pinto R, et al: From bladder to systemic syndrome: concept and treatment evolution of interstitial cystitis, *Int J Womens Health* 7:735-744, 2015.
DiRenzo D, Ellsworth PI, Caldamone AA, et al: Transurethral puncture for ureterocele – which factors dictate outcomes?, *J Urol* 184:1620-1624, 2010.
Figuera VH, Chavhan GB, Oudjhane K, et al: Utility of MR urography in children suspected of having ectopic ureter, *Pediatr Radiol* 44:956-962, 2014.
Ginsberg J: The epidemiology and pathophysiology of neurogenic bladder, *Am J Manag Care* 19:S191-S196, 2013.
Gonzales EJ, Arms L, Vizzard MA: The role(s) of cytokines/chemokines in urinary bladder inflammation and dysfunction, *Biomed Res Int* 2014; doi:10.1155/2014/120525.
Heinlen JE, Manatt CS, Bright BC, et al: Operative versus nonoperative management of ureteropelvic junction obstruction in children, *Urology* 73:521-525, 2009.
Jaiman S, Ulhoj BP: Bilateral intravesical ureterocele associated with unilateral partial duplication of the ureter and other anomalies, *APMIS* 118:809–814, 2010.
Karnak I, et al: Prenatally detected ureteropelvic junction obstruction: clinical features and associated urologic abnormalities, *Pediatr Surg Int* 24:395–402, 2008.
Kirschner-Hermanns R, Daneshgari F, Vahabi B, et al: Does diabetes mellitus-induced bladder remodeling affect lower urinary tract function?:ICI-RS 2011, *Neurourol Urodyn* 31:359-364, 2012.
Knudson MJ, Austin JC, McMillan ZM, et al: Predictive factors of early spontaneous resolution in children with primary vesicoureteral reflux, *J Urol* 178:1684-1688, 2007.
Lam JS, Breda A, Schulam PG: Ureteropelvic junction obstruction, *J Urol* 177:1652–1658, 2007.
Lee H, Lee YS, Im YJ, et al: Vesicoureteral reflux and bladder dysfunction, *Transl Androl Urol* 1(3):153-159, 2014.
Liu DB, Armstrong WR, Maizels M: Hydronephrosis: prenatal and postnatal evaluation and management, *Clin Perinatol* 41:661-678, 2014.
Marguerian PA, Taenzer A, Knoerlein K, et al: Variation in management of duplex system intravesical ureteroceles: a survey of pediatric urologists, *J Urol* 184:1625-1630, 2010.
Martin E, Sheaves C, Childers K: Underlying mechanisms and optimal treatment for interstitial cystitis: a brief overview, *Urol Nurs* 35(3):111-116, 2015.
Miranda DM, dos Santos Jr AC, Sarubi HC, et al: Association of angiotensin type 2 receptor gene polymorphisms with ureteropelvic junction obstruction in Brazilian patients, *Nephrology (Carlton)* 19:714-720, 2014.
Panicker J, de Seze M, Fowler CJ: Neurogenic lower urinary tract dysfunction and its management, *Clin Rehabil* 24:579-589, 2010.
Paye-Jaouen A, Pistolesi F, Botto N, et al: Long-term bladder function after ureterocele decompression in children, *J Urol* 193:1754-1759, 2015.
Pohl HG: Recent advances in the management of ureteroceles in infants and children: why less may be more, *Curr Opin Urol* 21:322–327, 2011.
Timberlake MD, Corbett ST: Minimally invasive techniques for management of the ureterocele and ectopic ureter, upper tract versus lower tract approach, *Urol Clin N Am* 42:61-76, 2015.
Tubre RW, Gatti JM: Surgical approaches to pediatric ureteropelvic junction obstruction, *Curr Urol Rep* 2015:16:72. doi:10.1007/s11934-015-0539-1.
Vivante A, Kleppa M-J, Schulz J, et al: Mutations in *TBX18* cause dominant urinary tract malformations via transcriptional dysregulation of ureter development, *Am J Hum Gen* 97:291-301, 2015.
Yoost JL, Hertweck SP, Loveless M: Diagnosis and treatment of interstitial cystitis in adolescents, *J Pediatr Adolesc Gynecol* 25:162-171, 2012.

Bladder Cancer

American Cancer Society (ACS). *Bladder cancer*, 2014. Available at www.cancer.org/acs/groups/cid/documents/webcontent/003085-pdf. Accessed 8 November 2015.
Barrow TM, Michels KB: Epigenetic epidemiology of cancer, *Biochem Biophys Res Commun* 455:70-93, 2014.
National Cancer Institute. *Surveillance Epidemiology and End Results (SEER) stat fact sheets: bladder*. Available at http://seer.cancer.gov/statfacts/html/urinb.html. Accessed 5 November 2015.
National Guideline Clearinghouse (NGC). *Guideline summary: screening for bladder cancer: U.S. Preventive Services Task Force recommendations statement*. Agency for Healthcare Research and Quality (AHRQ). Available at http://www.guideline.gov/content.aspx?id=34289&search=bladder+cancer+screening. Summary completed 6/30/98. Updated 10/12/11. Accessed 11 January 2016.
National Guideline Clearinghouse (NGC). *Guideline summary: bladder cancer: diagnosis and management*. Agency for Healthcare Research and Quality (AHRQ). Available at http://www.guideline.gov/content.aspx?id=49119&search=bladder+cancer+diagnosis+and+management. Summary completed 8/28/15. Accessed 31 October 2015.
Tanaka MF, Sonpavde G: Diagnosis and management of urothelial carcinoma of the bladder, *Postgrad Med* 123:43–55, 2011.
Ye F, Wang L, Castillo-Martin M, et al: Biomarkers for bladder cancer management: present and future, *Am J Clin Exp Urol* 2(1):1-14, 2014.

Urinary Tract Infections

American Urological Association (AUA). *Adult UTI*. Available at https://www.auanet.org/education/adult-uti.cfm. Updated June 2012. Accessed 12 January 2016.
Armstrong K: Diagnosing and treating urinary tract infections in older people, *British J Community Nurs* 20(5):226-230, 2015.
Brill JR: Diagnosis and treatment of urethritis in men, *Am Fam Physician* 81:873–878, 2010.
Colgan R, Williams M: Diagnosis and treatment of acute uncomplicated cystitis, *Am Fam Physician* 84(7):771-776, 2011.
Dielubanza EJ, Mazur DJ, Schaeffer AJ: Management of non-catheter-associated complicated urinary tract infection, *Infect Dis Clin N Am* 28:121-134, 2014.
Gupta K, Hooton TM, Naber KG, et al: International clinical practice guidelines for the treatment of acute uncomplicated cystitis and pyelonephritis in women: a 2010 update by the Infectious Diseases Society of America and the European Society for Microbiology and Infectious Diseases, *Clin Infect Dis* 52(5):e103-e120, 2011.
Khatib N, Bradbury C, Chalker V, et al: Prevalence of *Trichomonas vaginalis, Mycoplasma genitalium* and *Ureaplasma urealyticum* in men with urethritis attending an urban sexual health clinic, *Int J STD AIDS* 26(6):388-392, 2015.
Moore KH, Malykhina AP: What is the role of covert infection in detrusor overactivity, and other LUTD? ICI-RS 2012, *Neurourol Urodyn* 33:606-610, 2014.
Nazarko L: Recurrent urinary tract infection in older women, *Nurse Prescribing* 12(12):608-613, 2014.
Nelson JM, Good E: Urinary tract infections and asymptomatic bacteriuria in older adults, *Nurse Pract* 40(8):43-48, 2015.
Roberts KB: Revised AAP guideline on UTI in febrile infants and young children, *Am Fam Physician* 86(10):940-946, 2012.
Saunders JM, Mercer CH, Sutcliffe LJ, et al: Factors associated with asymptomatic non-chlamydial non-gonococcal urethritis in heterosexual men: findings from a case-control study, *Int J STD AIDS* 24:627-631, 2013.

Silverman JA, Schreiber HL, Hooton TM, et al: From physiology to pharmacy: developments in the pathogenesis and treatment of recurrent urinary tract infections, *Curr Urol Rep* 14:448-456, 2013.

Vermeulen SH, Hanum N, Grotenhuis AJ, et al: Recurrent urinary tract infection and risk of bladder cancer in the Nijmegen bladder cancer study, *Br J Cancer* 112:594-600, 2015.

Workowski KA, Bolan GA: Sexually transmitted diseases treatment guidelines, 2015. *MMWR Recomm Rep* 64(3):1-137, 2015.

Urinary Stones

American College of Radiology (ACR). *ACR Appropriateness Criteria® Acute onset flank pain – suspicion of stone disease (urolithiasis)*. Available at https://acsearch.acr.org/docs/69362/Narrative/. Published 1995. Updated 2015. Accessed 12 January 2016.

American Urological Association (AUA) and European Association of Urology (EAU). *2007 guideline for the management of ureteral calculi*, 2007. Available at http://www.auanet.org/common/pdf/education/clinical-guidance/Ureteral-Calculi.pdf. American Urological Association. Accessed 12 January 2016.

Bender LC, Dubinsky TJ: The impacted ureteral stone, *Ultrasound Q* 28:235-237, 2012.

Philippou P, Moraitis K, Masood J, et al: The management of bladder lithiasis in the modern era of endourology, *Urology* 79:980-986, 2012.

Torricelli FCM, Mazzucchi E, Danilovic A, et al: Surgical management of bladder stones: literature review, *Rev Col Bras Cir* 40(3):227-233, 2012.

Zehri AA, Ather MH, Abbas F, et al: Preliminary study of efficacy of doxazosin as a medical expulsive therapy of distal ureteric stones in a randomized clinical trial, *Urology* 75:1285-1288, 2010.

30

Male Genital and Reproductive Function

Marvin Van Every

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- What is the role of Sertoli cells in spermatogenesis?
- What is the function of Leydig cells?
- Which branch of the autonomic nervous system is responsible for penile erection? Ejaculation?
- Which genitourinary structures develop embryologically from the wolffian ductal system in males?
- How do the hypothalamic-pituitary gonadotropic hormones influence male reproductive function?
- How do the processes of capacitation and acrosome reaction affect the fertilization process?

CHAPTER OUTLINE

Anatomy, 626
Upper Genitourinary Tract, 626
Lower Genitourinary Tract, 627
Bladder, 627
Urethra, 628
Auxiliary Genital Glands, 628
Prostate, 628
Seminal Vesicles, 629
Bulbourethral Glands, 629
External Genitalia, 629
Scrotum, 629
Testes, 629
Epididymis and Ductus Deferens, 630
Penis, 631
Embryology, 632
Nephric System, 632
Vesicourethral Unit, 633
Gonads, 633
Genital Duct System, 633
External Genitalia, 633
Male Reproductive Physiology, 633
Hypothalamic-Pituitary-Testicular Axis, 633
Spermatogenesis, 637
Anatomy of Spermatozoa, 637
Transport of Spermatozoa, 637
Erection, Emission, and Ejaculation, 639
Capacitation, 639
Acrosome Reaction, 639

This chapter provides a foundation for comprehending male genital and reproductive disorders, which are presented in Chapter 31. The anatomy and embryology of the male genitourinary tract—those organs involved in the processes of sexual reproduction and elimination of nitrogenous wastes—will be presented first. Because these organs are derived from common embryologic structures, the anatomy and embryology of the male genitalia and urinary system will be emphasized, and the differences in embryologic development between males and females will be considered when pertinent. The remainder of this chapter will deal with the physiologic processes of male reproduction.

ANATOMY

Upper Genitourinary Tract

The upper genitourinary tract consists of the kidneys and ureters. The kidneys receive their blood from the renal arteries, which arise directly

from the aorta. They are usually solitary but will at times be duplicated. The ureteral blood supply is derived from multiple sources. The renal pelvis and upper part of the ureter receive blood from branches of the renal artery. The arterial blood supply of the middle ureter segment comes from the internal spermatic artery (gonadal artery), and the lowermost ureter sections receive blood from the branches of the common iliac, internal iliac, and vesical arteries (Fig. 30.1). The veins of the renal pelvis and ureter are usually paired with the arteries.

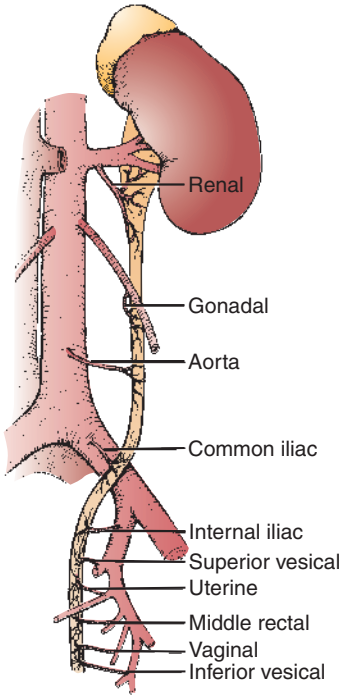


FIG 30.1 Sources of ureteral blood supply. (From Wein AJ et al, editors: *Campbell-Walsh urology,* ed 10, Philadelphia, 2012, Saunders.)

Lower Genitourinary Tract

Bladder

The bladder is a hollow muscular organ that serves as a reservoir for urine. The adult bladder normally has a capacity of 450 to 500 mL. When empty, the bladder lies behind the pubic symphysis and is mainly a pelvic organ. With overdistention or chronic urine retention, the abdomen may bulge, allowing easy palpation of the bladder in the suprapubic region.

The ureters enter the bladder posteroinferiorly. The ureteral orifices are situated on a crescent-shaped ridge and are approximately 2.5 cm apart. The triangular area demarcated by this interureteric ridge and bladder neck is called the *trigone* (Fig. 30.2). As will be discussed later in the chapter, the trigone has a different embryologic origin from the rest of the bladder body, or fundus. The trigone is composed of mesoderm, and the fundus is composed of endoderm.

In males, the bladder lies anterior to the seminal vesicles, vasa deferentia, ureters, and rectum. The dome and part of the posterior bladder surfaces are covered by peritoneum and are thus in close proximity to the small bowel and the sigmoid colon. The neck of the bladder, which is the most inferior part, leads to the urethra. In males, the prostate lies between the bladder and the muscle layers of the pelvic floor that composes the urogenital diaphragm.

The arterial blood supply of the bladder comes from the superior, middle, and inferior vesical arteries, which originate from the anterior division of the hypogastric artery. Venous drainage occurs by a rich plexus of veins that surround the bladder and ultimately drain into the hypogastric veins.

The bladder and urethra receive their nerve supply from both the sympathetic and the parasympathetic divisions of the autonomic nervous system. The sympathetic fibers, originating mainly from the lower thoracic and upper lumbar segments (T11–T12 and L1–L2), innervate the bladder and urethra as the hypogastric nerves. These sympathetic fibers are distributed more densely in the bladder base and proximal end of the urethra than in the bladder dome. The sympathetic nerves facilitate storage of urine. Studies have revealed differences in the bladder muscle receptors, with cholinergic receptors concentrated in the fundus

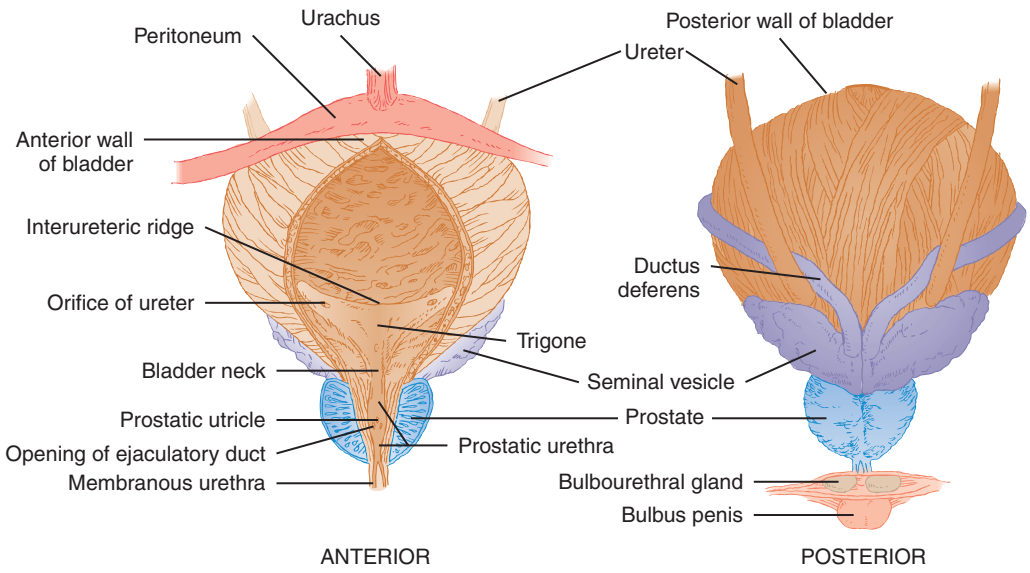


FIG 30.2 Anterior and posterior views of the prostate gland and related structures. The triangular area demarcated by the interureteric ridge and the bladder neck is the trigone. (From Black JM et al: *Medical-surgical nursing: clinical management for positive outcomes,* ed 6, Philadelphia, 2001, Saunders, p 940.)

and adrenergic receptors present in the trigone and proximal end of the urethra (Fig. 30.3).

The parasympathetic nerve supply originates from the sacral segments (S2–S4), which proceed to form a plexus surrounding the bladder. In the male, a separate segment will reach the prostate and form the prostatic plexus. From this plexus, nerves emerge to innervate the erectile tissue of the male penis and the clitoris of the female (see Fig. 30.3).

Branches of the bladder plexus penetrate the muscular coat of the bladder and become distributed throughout the detrusor. Parasympathetic muscle receptors are cholinergic in nature, and parasympathetic stimulation induces a detrusor contraction that causes bladder emptying.

Urethra

The male urethra, which extends from the bladder to the external opening (urethral meatus) at the tip of the penis, functions as a conduit for both urinary and genital systems. It is commonly divided into three segments: the *prostatic,* the *membranous,* and the *penile* or *spongy urethra* (Fig. 30.4).

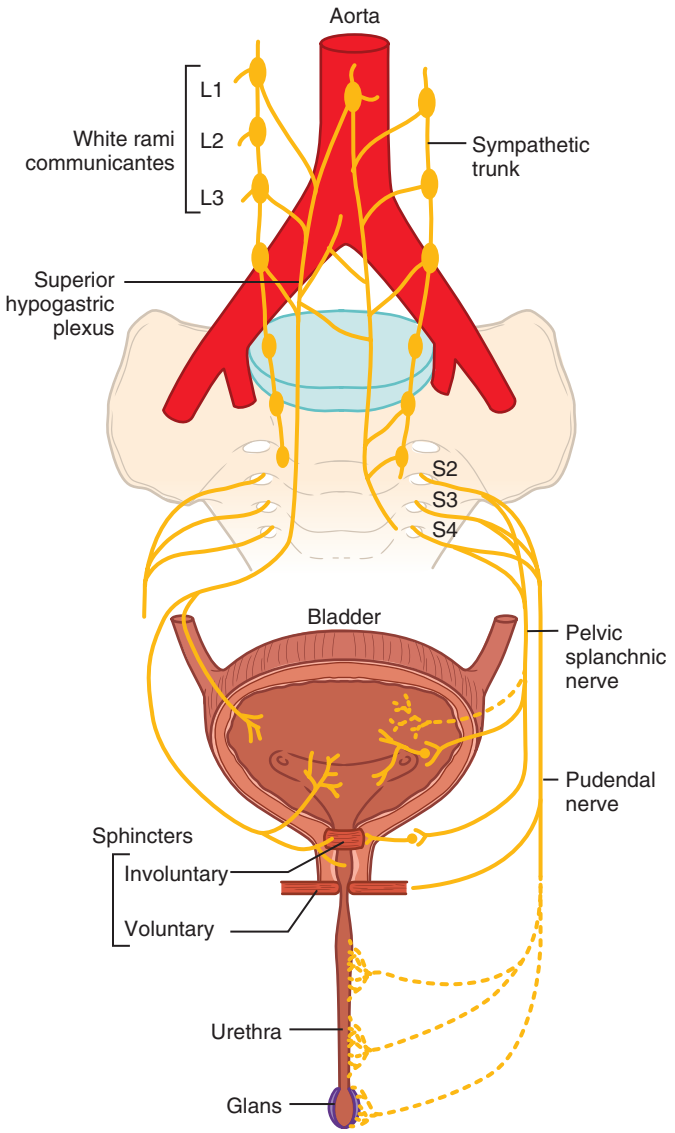


FIG 30.3 Diagram of nerve supply to bladder and urethra.

Auxiliary Genital Glands

The auxiliary genital glands of the male consist of the *prostate,* the *seminal vesicles,* and the *bulbourethral glands.* These glands secrete products that contribute to the seminal fluid.

Prostate

The prostate lies below the bladder and has both a muscular and a glandular component. The normal prostate weighs about 20 g and measures about 3.5 cm transversely and about 2.5 cm in its vertical and anteroposterior dimensions. The prostate is conical and is anterior to the rectum. Its base is continuous with the bladder neck, and the inferior aspect of the prostate gland, or apex, lies adjacent to the urogenital diaphragm (Fig. 30.5).

The prostate consists of a thin fibrous capsule with internally circular smooth muscle fibers and collagenous tissue that surround the urethra. Deep in this layer of connective and elastic tissue lies the prostatic stroma, which contains the prostatic epithelial glands. These glands drain into excretory ducts, which open chiefly on the floor of the urethra between the verumontanum and the vesical neck. The prostate is primarily a reproductive organ. In conjunction with the seminal vesicles, the prostate produces the fluid that supports the sperm. In fact, the sperm constitute a small amount of the semen, with the vast majority of the seminal fluid coming from the prostate and seminal vesicles. A further function of the prostate gland is to act as a valve for the bladder.

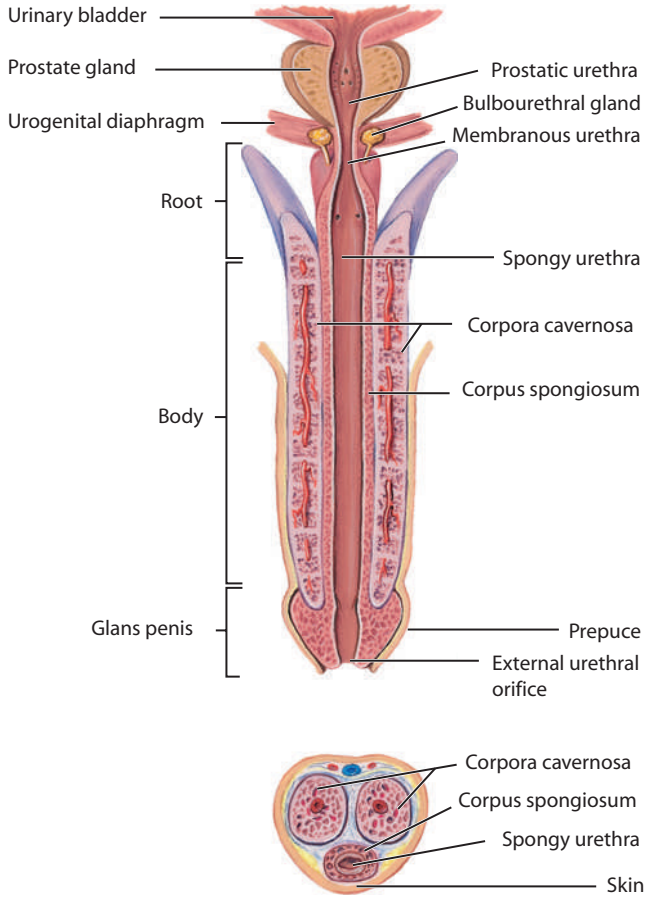


FIG 30.4 Cross-sectional view of the penis emphasizing the membranous urethra, the urogenital diaphragm, the bulbourethral or Cowper gland, and the orifices of the bulbourethral glands. (From Applegate EJ: *The anatomy and physiology learning system: textbook,* ed 4, St Louis, MO, 2011, Saunders.)

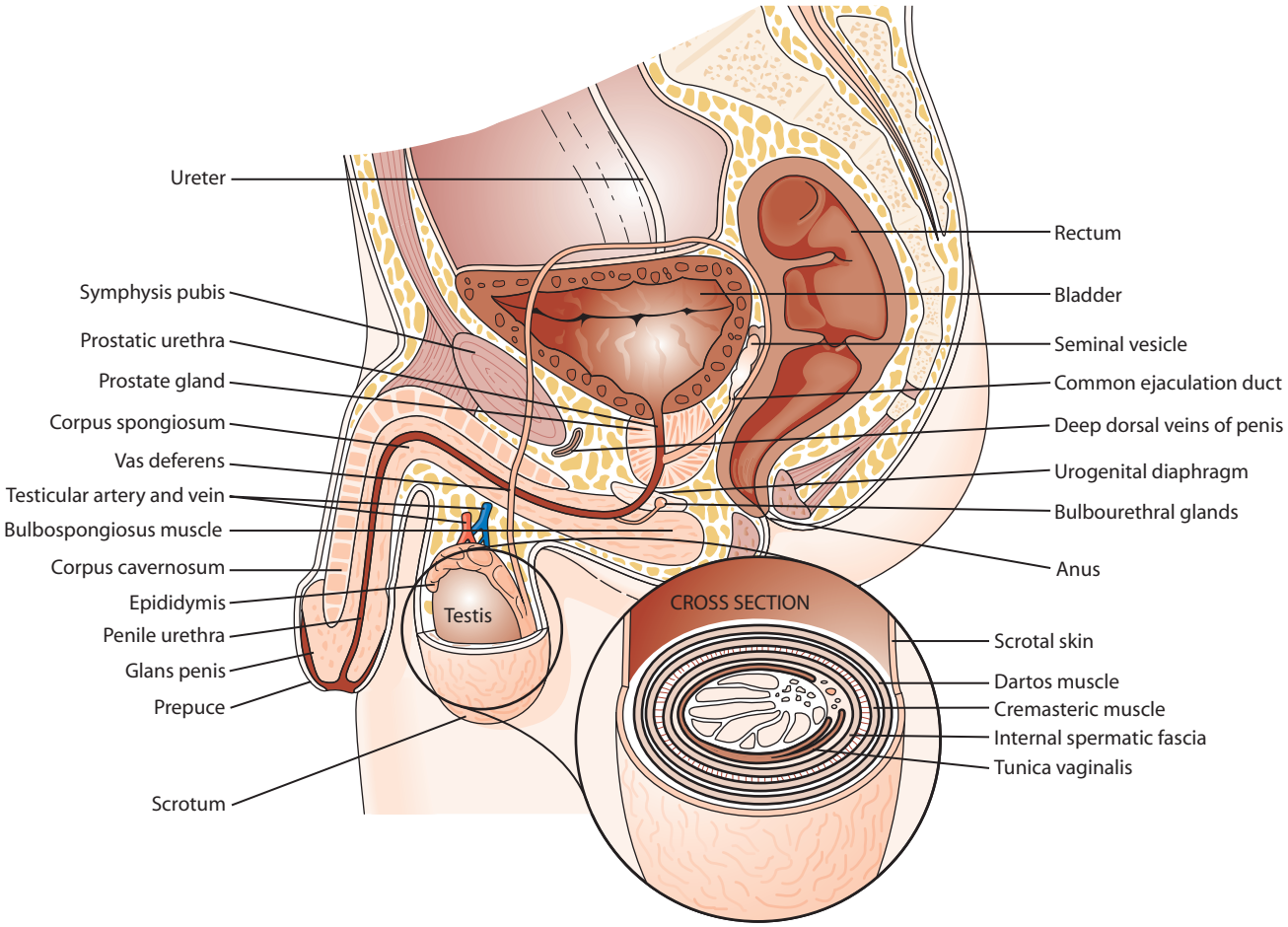


FIG 30.5 Male genitourinary anatomy, including a cross-section of the scrotum and its layers.

The main blood supply of the prostate is derived from the inferior vesical artery, a branch of the hypogastric artery. Besides the prostate, this artery also supplies the distal portion of the ureter, the seminal vesicles, and part of the bladder. A complex venous plexus situated between the prostate and overlying tissue freely communicates with the inferior hypogastric veins and provides venous drainage to the prostate.

Seminal Vesicles

The seminal vesicles are paired organs that lie next to the prostate under the base of the bladder (see Fig. 30.5). Their coiled pouches secrete a fluid important to the survival of spermatozoa.

Bulbourethral Glands

The bulbourethral or Cowper glands are located on each side of the membranous urethra within the urogenital diaphragm. They release a mucoid secretion into the urethra.

External Genitalia

Scrotum

The scrotum (see Fig. 30.5) is a pouchlike sac that lies below the penis and pubic symphysis. A septum of connective tissue divides the sac into two compartments. Each compartment contains a male gonad, or testis, with its associated epididymis, the lower end of the vas deferens and the lower portion of the spermatic cord. The scrotum not only supports the testes but also, by relaxation and contraction of its muscular layer, helps regulate the temperature of the testes.

The scrotal sac consists of several tissue layers. The scrotal skin overlies the dartos muscle layer, whose smooth muscle fibers are embedded in loose connective tissue. The dartos muscle functions to contract the scrotal pouch when cold and expand it when warm. Under the dartos layer are several fascial layers (see Fig. 30.5) that are continuous with the muscular layers of the abdominal wall and also make up the covering of the spermatic cord. The external spermatic fascia is continuous with the external oblique aponeurosis of the abdominal wall. A few slips of skeletal muscle derived from the internal oblique muscle layer make up the cremasteric muscle, which adds to the upper part of the cord. The internal spermatic fascia is a continuation of the transverse fascia of the abdominal wall, with the transversus abdominis muscle not contributing to the cord layers. Finally, the peritoneum provides the tunica vaginalis layers, which are actually separated from the abdominal cavity by obliteration of the processus vaginalis.

The scrotum receives its blood supply from the external pudendal artery, a branch of the femoral artery. In addition, the scrotum receives blood from portions of the internal pudendal artery (a branch of the hypogastric artery) and the cremasteric and testicular arteries that transverse the spermatic cord.

Testes

The testes are the male reproductive organs responsible for sperm production. They average about 4 to 5 cm in length and 2 to 3 cm in thickness. The testes lie within the scrotum and are suspended by the spermatic cord. The testes are covered by a thick fascial layer called the *tunica albuginea*. This layer invaginates posteriorly to form the

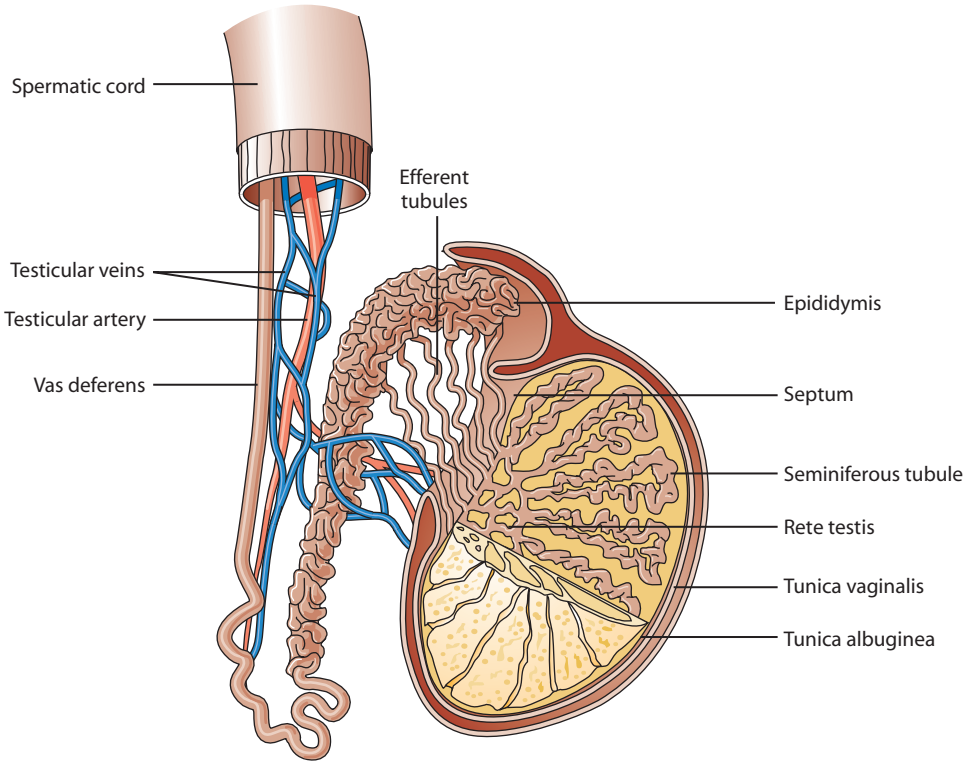


FIG 30.6 Anatomy of the testis and epididymis. Note the numerous compartments of the testis that are filled with seminiferous tubules gathering into the rete testis; they join to form a markedly convoluted tubule that becomes the epididymis, which is continuous with the vas deferens. The epididymis attaches to the dorsomedial aspect of the testis, and the vas deferens joins the other structures of the spermatic cord.

mediastinum testis. This fibrous mediastinum sends fibrous septa into each testis that separate it into many different lobules. Each lobule contains one to four seminiferous tubules that if stretched to full length would measure approximately 60 cm. Spermatozoa production occurs within the epithelial lining of the seminiferous tubules (Fig. 30.6).

The seminiferous tubules have a basement membrane consisting of elastic and connective tissue that supports the seminiferous cells. The seminiferous cells are either Sertoli cells (supporting cells) or spermatogenic cells. Found between the seminiferous tubules and embedded in connective tissue, the interstitial Leydig cells produce and secrete testosterone, a hormone involved in the development of male sexual characteristics (Fig. 30.7). The seminiferous tubules converge on the mediastinum testis. The tubules, which are connected by the straight efferent ducts, drain into the head of the epididymis.

The primary testicular blood supply is derived from the internal spermatic arteries, which arise directly from the aorta below the renal arteries. They course inferiorly through the spermatic cord and anastomose with the cremasteric arteries and the arteries of the vas; these vessels also contribute to the blood supply. The blood from the testis returns through a plexus of veins in the spermatic cord (the pampiniform plexus) that forms the spermatic veins. The left internal spermatic vein enters the left renal vein, which subsequently enters the vena cava. The right internal spermatic vein enters the vena cava directly.

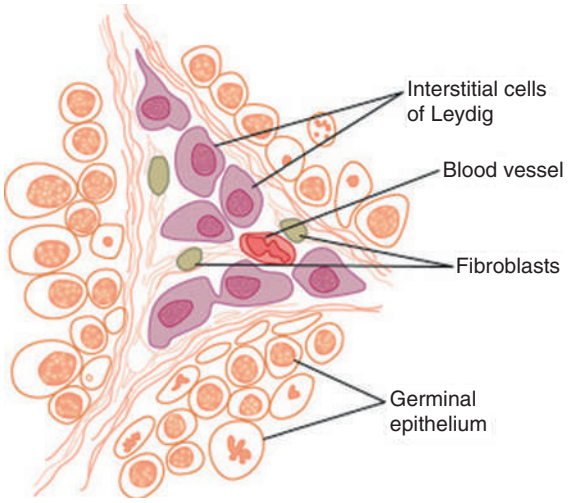


FIG 30.7 The interstitial Leydig cells that secrete testosterone are located in the interstices between the seminiferous tubules. (From Hall JE, editor: *Guyton and Hall textbook of medical physiology,* ed 13, Philadelphia, 2015, Saunders.)

Epididymis and Ductus Deferens

The epididymis is a tightly coiled tube that lies along the top of and behind each testis. It is divided into the head, situated at the upper pole of the testes; the body, lying posterior to the testes; and the tail, which is attached to the inferior pole of the testes (see Fig. 30.6). The body and the tail of the epididymis form one continuous tube that serves as a conduit for maturing spermatozoa. In the epididymis, sperm develop the ability to swim.

As the convoluted tube of the tail leaves its testicular attachments, it increases in diameter to become a thick, muscular tube called the ductus deferens, also called the vas deferens. Leaving the spermatic cord, the vas deferens follows an extraperitoneal course and passes caudally and laterally along the pelvic wall. As it passes medial to the

distal end of the ureter, it bends caudally to reach the midline and lies on the posterior wall of the bladder just medial to the seminal vesicles. It terminates in a dilated ampulla that courses underneath the base of the prostate. At this point the duct of the seminal vesicle joins with the duct of the ampulla, and the ejaculatory duct is formed. The ejaculatory ducts open in the prostatic urethra at the level of the verumontanum.

Penis

The penis is the male organ of copulation and urinary excretion. It is composed of three erectile bodies—two paired corpora cavernosa, which lie dorsally, and the corpus spongiosum, which contains the urethra (Fig. 30.8). Grossly, the penis is divided into three segments. The root of the penis consists of the proximal ends of the corpora cavernosa, which attach to the pelvic bones, and the proximal end of the corpus spongiosum, which connects to the undersurface of the urogenital diaphragm. Together these attachments provide fixation and stability to the penis. The shaft or body of the penis consists of all three erectile bodies: the two cavernous bodies lying on the dorsum, and the corpus spongiosum, which occupies a depression on their ventral surface. Finally, the glans of the penis forms the distal segment of the corpus spongiosum (see Fig. 30.4).

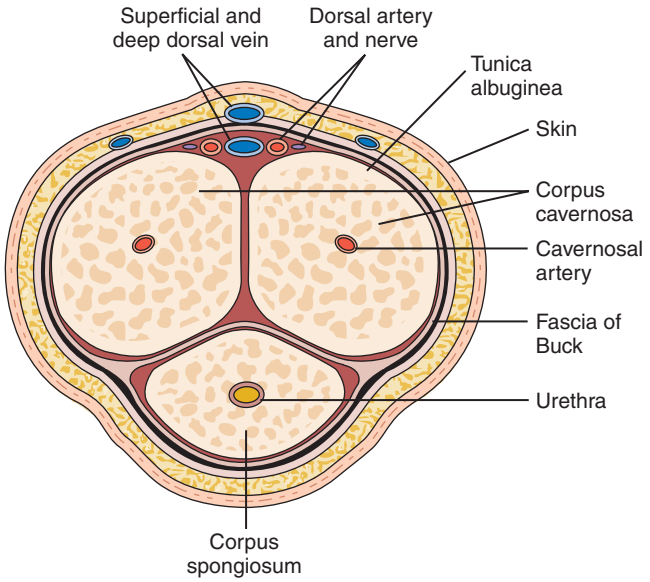


FIG 30.8 Transverse section through the penis. The paired upper structures are the corpora cavernosa. The single lower body surrounding the urethra is the corpus spongiosum.

The three erectile bodies have the capability to become engorged with blood and enlarge considerably with erection. Microscopically, these bodies have an internal spongelike network that consists of endothelium-lined spaces surrounded by smooth muscle.

Each corpus is enclosed in a fascial sheath, the tunica albuginea, and all are subsequently surrounded by a thick fibrous envelope known as the fascia of Buck. The overlying skin of the penis is remarkable for its thinness and looseness of connection with the fascial sheath of the penis. The skin of the penis is folded upon itself to form the prepuce, or foreskin. It is this penile skin overlying the glans that is removed with circumcision.

The arterial blood supply is primarily derived from the paired internal pudendal arteries, which are branches of the hypogastric arteries. Each internal pudendal artery branches several times in the penis. The deep or cavernous artery supplies the entire corpus cavernosum. The urethral artery supplies the corpus spongiosum, and the bulbar artery supplies the bulb of the corpus spongiosum. The dorsal artery continues along the dorsum of the penis and lies below the fascia of Buck and between two dorsal veins. It provides additional supply to the glans (Fig. 30.9).

Venous drainage of the penis is through several channels. The cavernous veins drain the corpora cavernosa, and the circumflex veins join the deep dorsal vein of the penis to also drain the corpora. The superficial dorsal vein drains the glans and part of the distal portion of the corpora. Finally, a bulbar branch drains the bulbous urethra and proximal portion of the corpus spongiosum (see Fig. 30.9). Together these branches coalesce and pass through the urogenital diaphragm into the retropubic venous plexus of Santorini.

The nerve supply of the penis is formed from both parasympathetic and sympathetic components. The parasympathetic fibers arise from S2–S4, and the sympathetic component is derived from the hypogastric plexus. Parasympathetic stimulation from the pudendal nerve results in relaxation of vascular resistance, which increases blood flow to the penis and creates an erection. The pudendal nerve also carries sensory fibers from the penis and enters the sacral spinal cord to contribute to penile erection.

Sympathetic nerve fibers may contribute to erectile capacity, but their role has not been proved conclusively. They do innervate the proximal involuntary sphincter of the bladder neck, where contraction prevents retrograde ejaculation of semen from the prostatic urethra into the bladder. They also innervate the muscles of the seminal vesicles and prostate, which, when stimulated, cause ejaculation of seminal fluid into the urethra.

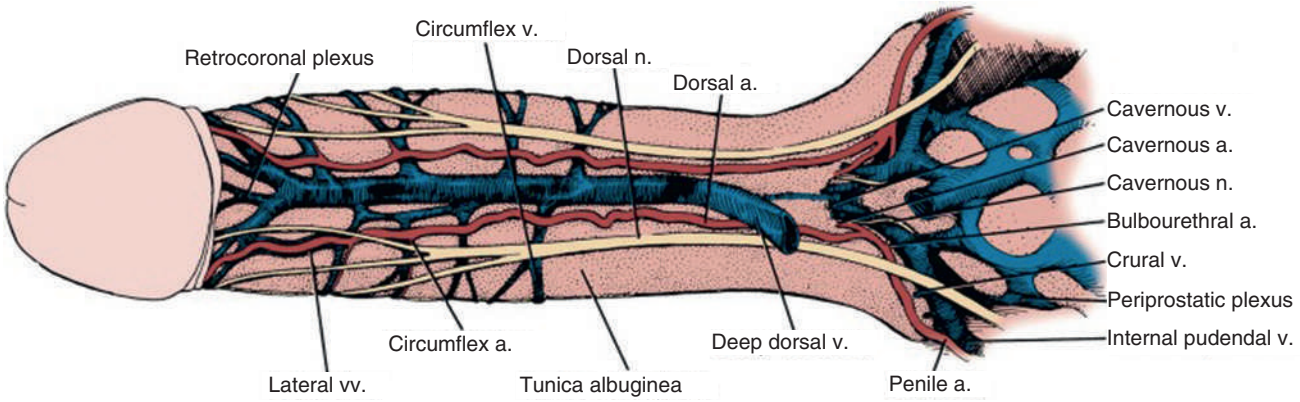


FIG 30.9 Dorsal penile arteries, veins, and nerves. (From Hinman Jr F: *Atlas of urosurgical anatomy,* Philadelphia, 1993, Saunders, p 445.)

KEY POINTS

- The upper genitourinary tract is composed of the kidneys and ureters. The lower genitourinary tract includes the bladder and urethra and the accessory male sexual organs.
- Ureters transport urine from the renal pelvis to the bladder. Ureters have several points of narrowing that predispose to obstruction: ureteropelvic junction, pelvic brim, and ureterovesical junction.
- The adult bladder has a normal capacity of 450 to 500 mL. With overdistention, the bladder may be palpable in the suprapubic region. The bladder is a muscular organ composed of several layers of muscle fibers. An important muscular landmark in the bladder is the trigone. Parasympathetic stimulation of the bladder results in bladder muscle contraction.
- The prostate is a key organ in the male genitourinary system with both reproductive and continence functions. It also causes many pathologic conditions as men age, such as prostatitis, benign hyperplasia, and cancer.
- The urethra extends from the bladder to the meatus at the end of the penis. In addition to transporting urine, the urethra has ducts that receive fluid from the prostate, seminal vesicles, and bulbourethral glands.
- The scrotal sac supports the testes and regulates their temperature. Testes contain several cell types important in sperm production and the development of secondary sex characteristics. Spermatogenic cells produce sperm in the testes. Sertoli cells serve to support and nurture spermatogenesis. Leydig cells produce and secrete testosterone.
- Situated next to the testes, the epididymis serves as a collecting conduit for sperm. The epididymis is continuous with the ductus (vas) deferens. The vas travels along the pelvic wall and joins with the seminal vesicle duct at the prostate to form the ejaculatory duct. The ejaculatory ducts open into the urethra.
- Skin overlying the penis is very loose, which facilitates significant enlargement when the penis is engorged with blood during erection. Parasympathetic fibers forming the pudendal nerve are responsible for erection. Ejaculation is a function of the sympathetic nerve fibers.

EMBRYOLOGY

Developmental processes in the genital and urinary systems are intimately related. To facilitate understanding of this development, the two systems will be discussed in several subdivisions. The urinary system, which is composed of the nephric system and the vesicourethral unit, will be discussed first. The genital system, which is composed of the gonads, the genital ducts, and the external genitalia, will be discussed second.

Nephric System

The nephric system develops progressively through three distinct phases: the *pronephros, mesonephros,* and *metanephros.* The pronephros is the earliest state in humans but corresponds to the mature structure in primitive vertebrates. The pronephros consists of 6 to 10 pairs of tubules connected by a pronephric duct. It grows caudally to join the cloaca, a blind end of the hindgut. The pronephros is a temporary structure and, except for its duct, disappears by the fourth week of intrauterine life (Fig. 30.10).

The mesonephros corresponds to the mature excretory organ of some amphibians. In humans it begins developing at about the fourth to fifth week of gestation. The tubules of the mesonephros are more numerous and form a cuplike outgrowth into which capillaries push to form a primitive glomerulus. The tubules communicate with the mesonephric duct, which is derived from the preceding pronephric duct. The number of mesonephros tubules reaches a maximum by about 8 weeks' gestation and then degenerates.

The final stage of development, the metanephros, begins in the fourth week when the ureteral bud grows out of the mesonephric duct. The bud elongates in a dorsocranial direction, where it meets a mass of mesoderm, the nephrogenic blastema, and begins to differentiate into the ureter and renal collecting system. The metanephros is derived from the nephrogenic blastema and eventually differentiates into the mature mammalian kidney.

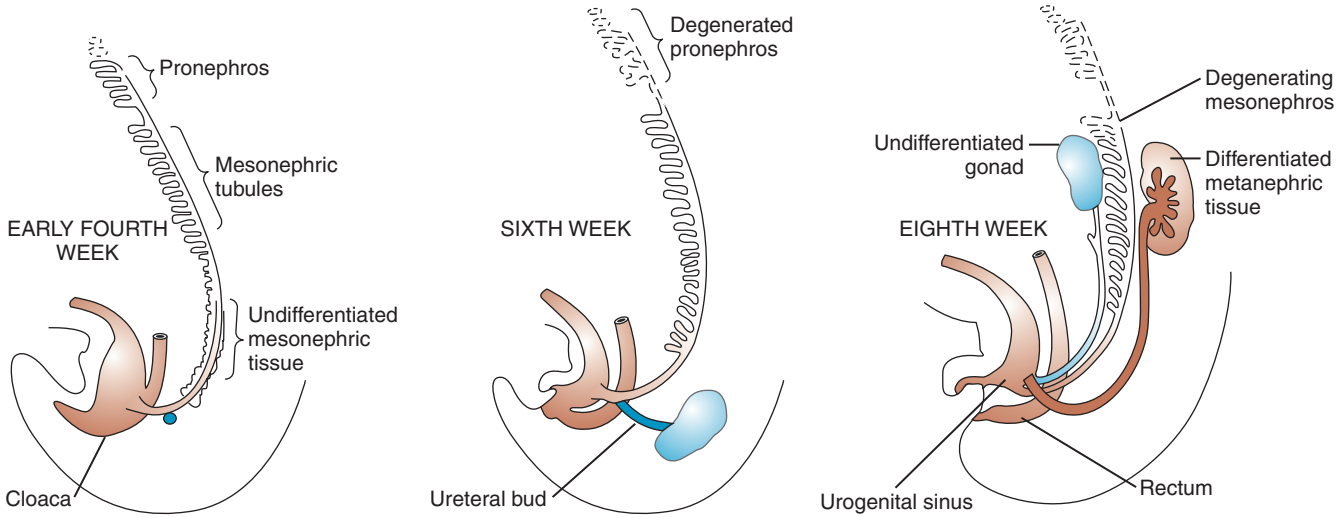


FIG 30.10 Schematic representation of the development of the nephric system. Only a few of the tubules of the pronephros are seen early in the fourth week, whereas the mesonephric tissue differentiates into mesonephric tubules that progressively join the mesonephric duct. The first sign of the ureteral bud from the mesonephric duct is shown. At 6 weeks, the pronephros has completely degenerated and the mesonephric tubules start to do so. The ureteral bud grows dorsocranially and has met the metanephrogenic cap. By the eighth week, cranial migration of the differentiating metanephros can be seen. The cranial end of the ureteric bud expands and starts to show multiple successive outgrowths. (From Tanagho EA, McAninch JW, editors: *Smith's general urology,* ed 13, East Norwalk, CT, 1992, Appleton & Lange, p 18.)

Vesicourethral Unit

The blind end of the caudal hindgut forms the cloaca, which is separated from the outside by a thin membrane of tissue, the urogenital membrane. At about 4 weeks' gestation a septum grows downward and separates the cloaca into a posterior compartment, which will become the rectum, and an anterior compartment, which will form the urogenital sinus.

The urogenital sinus receives the mesonephric duct, which is progressively absorbed into this structure. The mesonephric duct distal to the ureteral bud is absorbed into the sinus, and its mesenchyme subsequently forms the bladder trigone. The ureter, which is derived from the ureteral bud, and the mesonephric duct, which differentiates into the vas deferens, merge into the sinus as well. In a complex pattern of development, the opening of the ureteral bud, which will eventually become the ureteral orifice, migrates upward and laterally. The opening of the mesonephric duct, which will become the ejaculatory duct, migrates downward and medially (Fig. 30.11).

The urogenital sinus can be divided into two main segments. The ventral and pelvic portion, which receives the ureter, forms the bladder, part of the urethra in males, and the whole urethra in females. A phallic or urethral portion will receive the mesonephric ducts and in males will form a second part of the urethra. In females, this portion receives the müllerian ducts, which fuse distally to form the uterus and upper part of the vagina. The lower portion of the female urogenital sinus forms the lower part of the vagina and vaginal vestibule (Fig. 30.12).

Gonads

The undifferentiated and primitive gonads are derived from the urogenital ridge, a dorsal region of thickening from which the primitive kidney also forms. The gonads serve as precursors to the testes in males and the ovaries in females. During the seventh week, an individual gonad begins to assume the characteristics of either a testis or an ovary.

In the presence of testis-determining factor, which is located on the Y chromosome, a gonad develops into a testis. The gland increases in size, and the cells of the epithelium grow centrally into the organ's mesenchyme. These ingrowths become radially arranged, form cords, and begin to converge on the posterior aspect of the testis. The cords eventually differentiate into the seminiferous tubules, which produce spermatozoa. The testes descend behind the abdominal cavity in the retroperitoneal space and into the scrotum, usually by the eighth month of gestation.

In the absence of testis-determining factor, a gonad differentiates into an ovary, and a cortex forms from the germinal epithelium and ultimately gives rise to ovarian follicles containing ova. It descends only partially through the abdominal cavity and eventually lies adjacent to the fallopian tubes.

Genital Duct System

As the embryo develops, two different but related kinds of ducts form beside the undifferentiated gonads. The mesonephric ducts, or wolffian ducts, as previously explained, develop as nephric ducts but will go on to form the male genital ducts. The müllerian ducts develop alongside the mesonephric ducts (paramesonephric) and are genital structures from the start.

Early in development, each of the two müllerian ducts arises laterally to the mesonephric ducts, either directly from the mesonephric ducts themselves or possibly from the adjacent epithelium of the primitive abdominal cavity. Both ducts grow caudally to enter the urogenital sinus.

If a gonad differentiates into a testis, the wolffian ducts subsequently develop into the male duct system consisting of the epididymis, vas deferens, seminal vesicles, and ejaculatory ducts. The müllerian ducts, except for a few rudimentary fragments, rapidly atrophy.

If, on the other hand, a gonad develops into an ovary, the müllerian ducts proceed to form the uterus, fallopian tubes, and upper part of the vagina. The mesonephric, or wolffian, ducts fail to develop further and remain rudimentary (see Fig. 30.12).

External Genitalia

Development of the external genitalia begins at about 12 intrauterine weeks. Before this point, three small protuberances appear on the external aspect of the cloacal membrane. The genital tubercle is located anteriorly, and the genital swellings are situated on either side of the membrane. In the seventh week, rupture of the urogenital membrane gives the urogenital sinus a separate opening on the undersurface of the genital tubercle.

In males, the genital or labioscrotal swellings migrate and fuse centrally to form the scrotum. The fused genital tubercles elongate. The elongated fused tubercles form a cylindric shape with a ventral groove communicating with the urogenital sinus. This groove subsequently becomes covered by folds of tissue and forms the penile urethra (Fig. 30.13).

The female external genitalia closely resemble those of the male until about the eighth intrauterine week. At this time the genital tubercle lags behind in growth and becomes the clitoris. The urogenital sinus shortens and widens somewhat to form the vaginal vestibule, and the genital swellings form the labia majora. The urethral folds become the labia minora.

KEY POINTS

- The fetal gonads can become either a testis or an ovary. The presence of the testis-determining factor on the Y chromosome causes the testis to develop in the male fetus, whereas the absence of this factor allows the gonad to become an ovary in the female. Once the gonad has differentiated, it then is responsible for the cascade of events that lead to the formation of the female or male sexual organs and genitalia.
- During early embryonic development, the genital structures of males and females are similar. Two important ductal systems are the mesonephric (wolffian) ducts and the paramesonephric (müllerian) ducts. The mesonephric ducts develop to form the kidneys and the genital duct in males. In the presence of a testis, the wolffian ducts develop into the epididymis, vas deferens, seminal vesicles, and ejaculatory ducts. In the presence of an ovary, the müllerian ducts develop into the uterus, fallopian tubes, and upper part of the vagina, and the wolffian ducts fail to develop.
- Development of the external genitalia begins at about 12 weeks' gestation. In males the labioscrotal tissue fuses and elongates to form the scrotum and penis. In females, this tissue remains separated and forms the labia minora.

MALE REPRODUCTIVE PHYSIOLOGY

Hypothalamic-Pituitary-Testicular Axis

To fully understand male reproductive function, one must consider the endocrine function of the hypothalamic-pituitary-testicular axis. The components of this system function to maintain a constant level of the circulating hormones responsible for normal male sexual development and behavior, as well as the maturation of sperm necessary for fertility (Fig. 30.14).

The hypothalamus is the integrating center for this hormonal axis. This organ coordinates neural messages from the central nervous system and humoral (bloodborne) messages from the testis to control the secretion of a small peptide hormone: gonadotropin-releasing hormone (GnRH). The pituitary stalk provides the route for GnRH to travel to

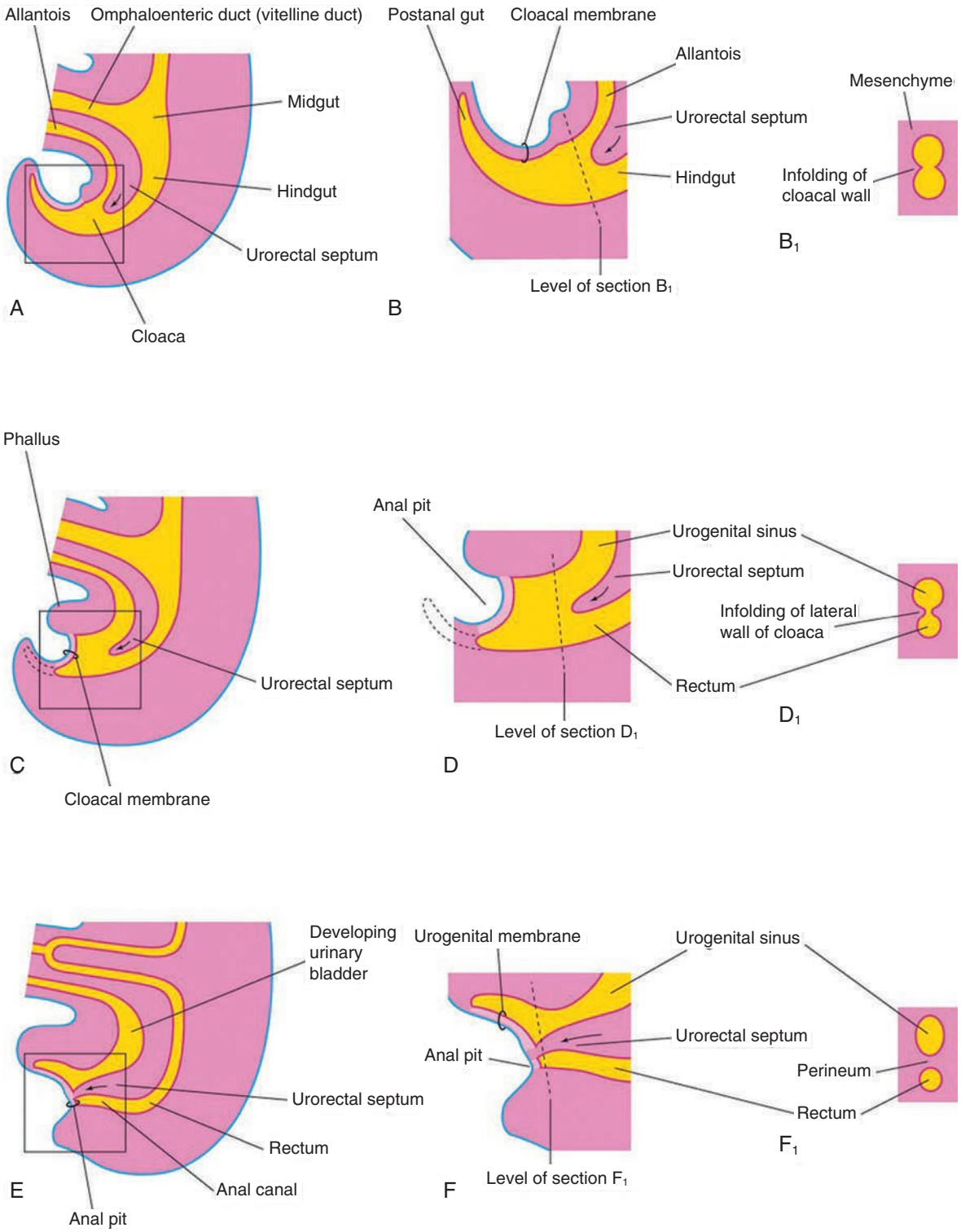


FIG 30.11 Successive stages in the partitioning of the cloaca into the rectum and urogenital sinus by the urorectal septum. **A, C,** and **E,** Views from the left side at 4, 6, and 7 weeks, respectively. **B, D,** and **F,** Enlargements of the cloacal region. **B_1** and **D_1,** Transverse sections of the cloaca at the levels shown in **B** and **D**. Note that the postanal portion (shown in **B**) degenerates and disappears as the rectum forms. (From Moore KL et al, editors: *The developing human: clinically oriented embryology,* ed 10, Philadelphia, 2016, Saunders.)

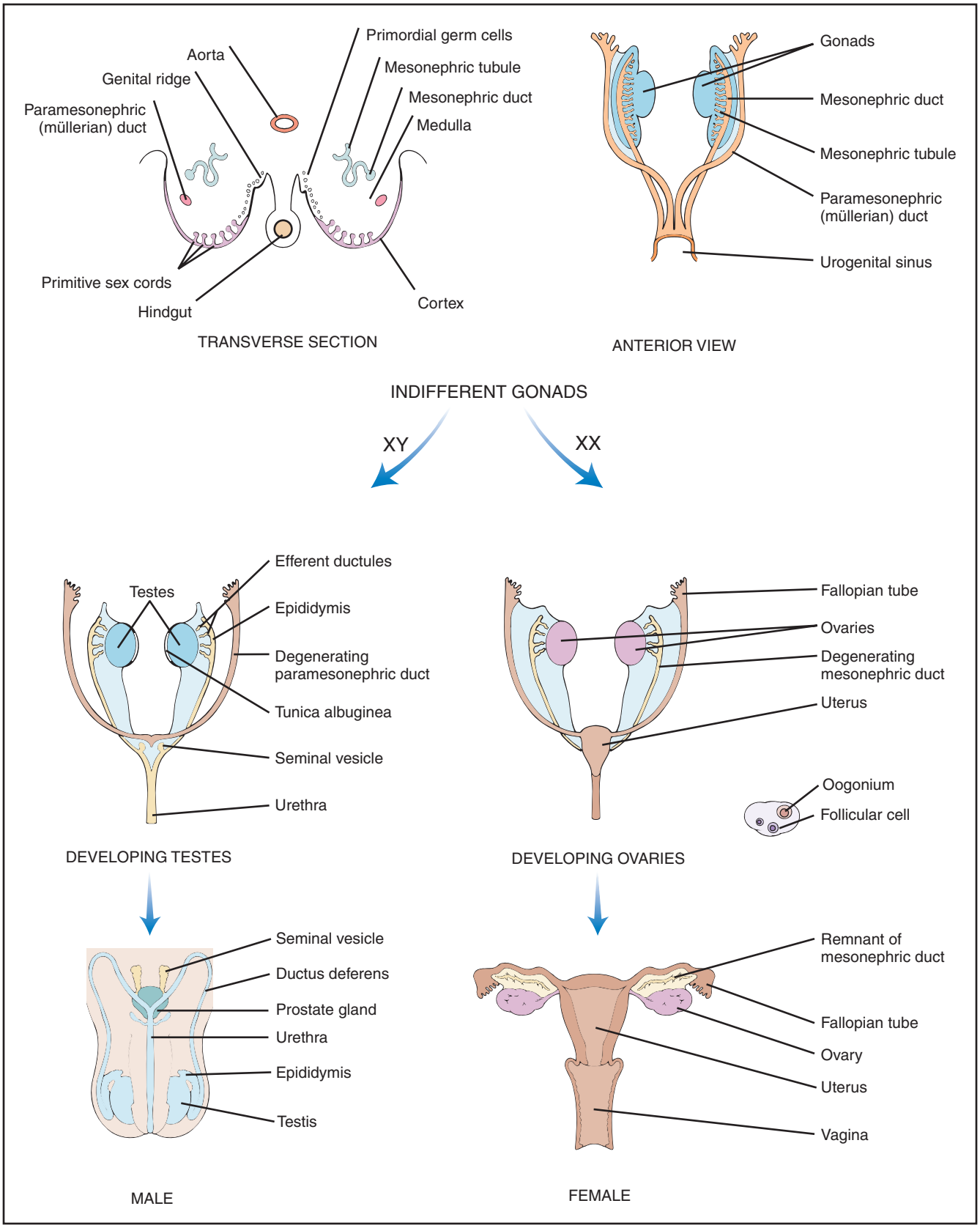


FIG 30.12 Transformation of the undifferentiated genital system into the definitive male and female systems. (From Nichols FH, Zwelling E: *Maternal-newborn nursing: theory and practice,* Philadelphia, 1997, Saunders, p 174.)

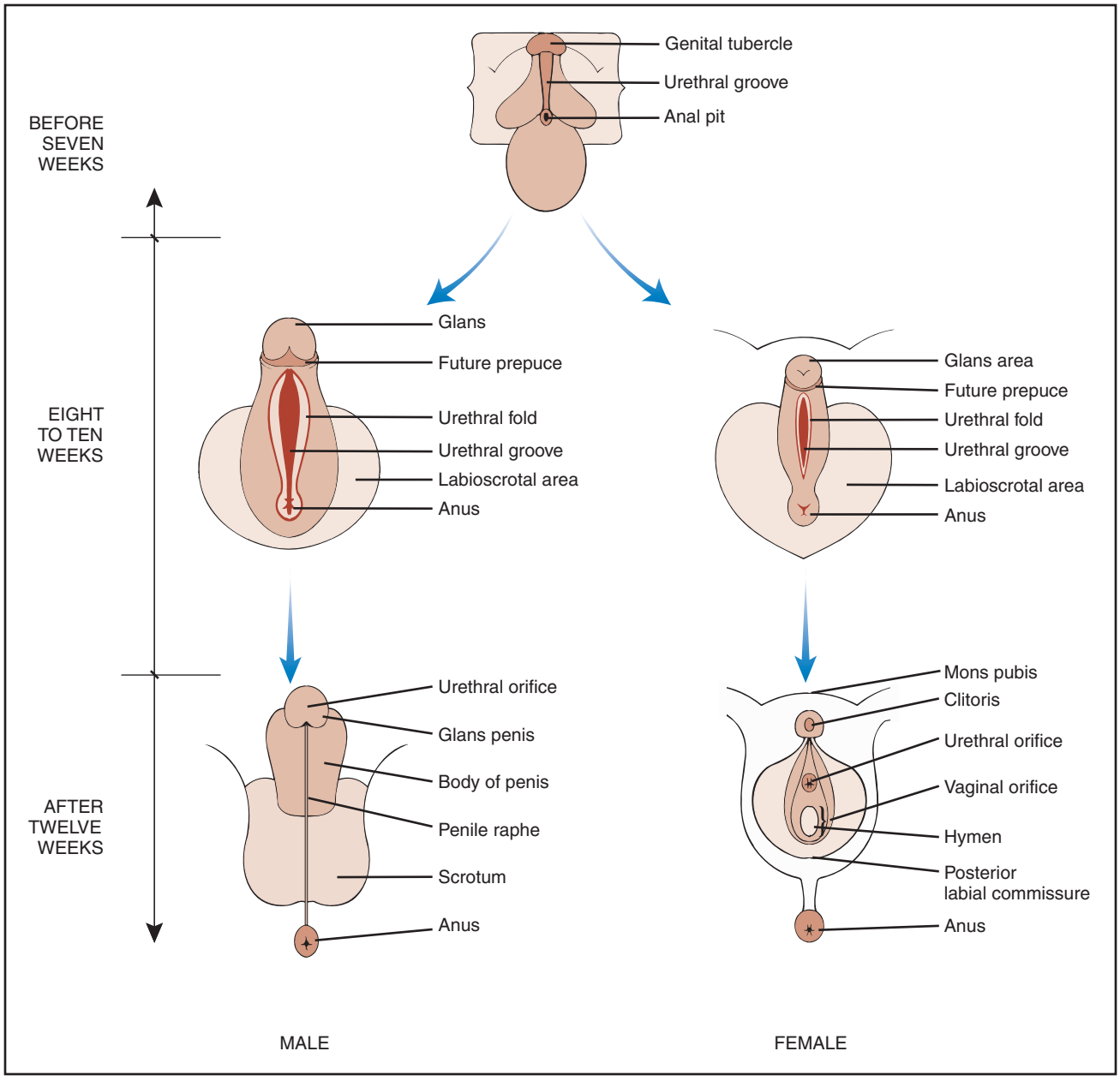


FIG 30.13 Development of the external genitalia from the indifferent stage (before 7 weeks) to fully differentiated stages (after 12 weeks of gestation). (From Nichols FH, Zwelling E: *Maternal-newborn nursing: theory and practice,* Philadelphia, 1997, Saunders, p 175.)

the pituitary gland, which lies caudal to the hypothalamus. A system of veins, the pituitary portal system, traverses the pituitary stalk and transports GnRH to the anterior portion of the pituitary gland.

In response to the secretion of GnRH, the pituitary synthesizes and releases two hormones: luteinizing hormone (LH) and follicle-stimulating hormone (FSH). Although these hormones carry names related to their function in females, they are produced by both genders through a feedback mechanism with GnRH for reproductive purposes.

By binding to receptors on the surface of the testicular Leydig cells, LH mediates testosterone synthesis. Binding of LH produces an increase in the conversion of adenosine triphosphate to cyclic 3,5-adenosine monophosphate. This activity stimulates the production of other intracellular enzymes with subsequent increased synthesis of testosterone. Testosterone is then released into the bloodstream and adjacent seminiferous tubules.

Besides testosterone, other steroid hormones are synthesized, among them dihydrotestosterone, 17-hydroxyprogesterone, and estradiol. Dihydrotestosterone functions to differentiate and mature the male external genitalia and prostate. In early puberty the production of androgen begins to increase, with normal adult plasma levels of testosterone and dihydrotestosterone being 300 to 1200 ng/dL and 30 to 60 ng/dL, respectively.

The function of FSH in male reproduction remains somewhat unclear. However, it appears that the production of sperm in the seminiferous tubules (spermatogenesis) requires the presence of high levels of both androgen and FSH.

A feedback inhibition mechanism controls the secretion of both LH and FSH. Production of LH occurs in response to serum levels of testosterone and estradiol. The regulation of FSH is controlled by inhibin and activin, which are produced by the Sertoli cells. In addition to

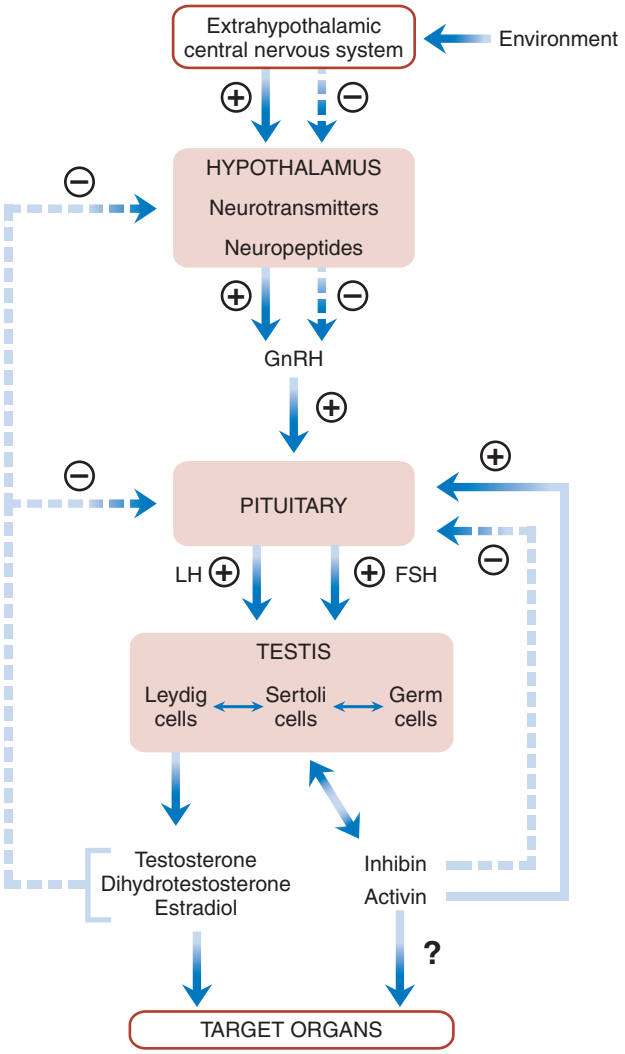


FIG 30.14 Feedback regulation of the hypothalamic-pituitary-testicular axis in males. Stimulatory effects are shown by ⊕ and negative feedback inhibitory effects are shown by ⊖. *FSH,* Follicle-stimulating hormone; *GnRH,* gonadotropin-releasing hormone; *LH,* luteinizing hormone. (From Carroll R, *Problem-based physiology,* Philadelphia, 2010, Saunders.)

inhibin, sex steroids modulate FSH secretion through feedback inhibition on the pituitary.

Spermatogenesis

To understand spermatogenesis, one must briefly consider the histology of the testis and its seminiferous tubules. As previously stated, Leydig cells occur in clusters in the interstitial tissue between the seminiferous tubules. These Leydig cells are responsible for the testicular production of testosterone.

The seminiferous tubules contain both germinal elements and supporting cells, which include the sustaining cells of the basement membrane and the Sertoli cells. The Sertoli cells rest on the basement membrane of the tubule and form a unique impermeable junction with each adjacent Sertoli cell. It is through this junction that the young germinal cells, or primary spermatocytes, migrate and pass from the basal compartment to the basement membrane and then to the central or adluminal compartment of the seminiferous tubule. The junction is also responsible for maintenance of the blood–testes barrier. This barrier ensures that the more mature spermatocytes and spermatids located in the adluminal compartment are behind the barrier and theoretically maintained in a constant intratubular environment to support the development of maturing sperm cells.

The process of sperm production is called *spermatogenesis* and involves several phases (Fig. 30.15). The proliferative phase involves division of the young germinal cells near the basement membrane (spermatogonia) either to replace their numbers or to produce daughter cells that will form spermatocytes. Next, a meiotic phase occurs in which spermatocytes undergo a reduction division. This division reduces the number of chromosomes to the monoploid number of 23 from the diploid number of 46. Finally, haploid spermatids undergo change to form mature spermatozoa.

While sperm cells mature and move from the basement membrane to the adluminal compartment, the Sertoli cells have an important nutritional role in the spermatogenic process. As the spermatid matures, it elongates and develops a tail, or flagellum, that attains a form similar to that of the mature spermatozoon. Mature spermatozoa are released into the tubular lumen and rapidly flow out to the rete testis and into the epididymis. Although each spermatogonium, one of the primitive male germ cells, requires about 70 days to develop into a mature sperm cell, or spermatozoon, within each tubule are spermatozoa in all stages of development. This characteristic allows new spermatozoa to be continuously produced across the male life span. The effects of aging on the male reproductive system are described in the Geriatric Considerations box.

Anatomy of Spermatozoa

The human spermatozoon is approximately 60 µm in length. The oval head contains a nucleus that is highly condensed and stabilized by cross-links between its molecules, which makes it very resistant to physical injury during its passage and storage in the epididymis. An outer membrane, the acrosome, contains the enzymes required for penetration of the female egg before fertilization.

The tail accounts for 90% of the length of the spermatozoon and is divided into a middle piece, principal piece, and end piece. The spermatozoon derives its motile ability from the motor apparatus of the tail, which is called the *axoneme.* The axoneme, which runs the length of the tail, is composed of a central pair of tubules surrounded by a ring of nine pairs of tubules (the 9 + 2 pattern). This ring of tubules is surrounded by a supporting structure of nine noncontractile dense fibers. Within the middle piece, a circular sheath of mitochondria (Fig. 30.16) surrounds these outer dense fibers.

The mitochondria contain the enzymes required for the production of adenosine triphosphate (ATP), the energy source for the cell. Within the axoneme are enzymes and structural proteins. These enzymes convert chemical energy from ATP to the mechanical energy of sperm cell movement to aid in fertilization of the egg.

Transport of Spermatozoa

Once mature spermatozoa are released from the Sertoli cells into the seminiferous tubules, they must pass through approximately 6 m of duct in the male reproductive tract before leaving the urethral meatus and being deposited in the vagina during sexual intercourse. From the seminiferous tubules, the spermatozoa are deposited into the rete testis, a collecting chamber for all the seminiferous tubules. From the rete testis, the sperm travel through the efferent ductules, 12 to 20 channels that pass into a single compact duct, the epididymis. The epididymis is a tightly convoluted duct that is divided into three regions: the caput (globus major), the corpus (body), and the cauda epididymis (tail, or globus minor). Unfolded and stretched, the epididymis would measure 12 to 15 feet.

After leaving the epididymis, the sperm enter the ductus or vas deferens. Embryologically, this duct is derived from the mesonephric

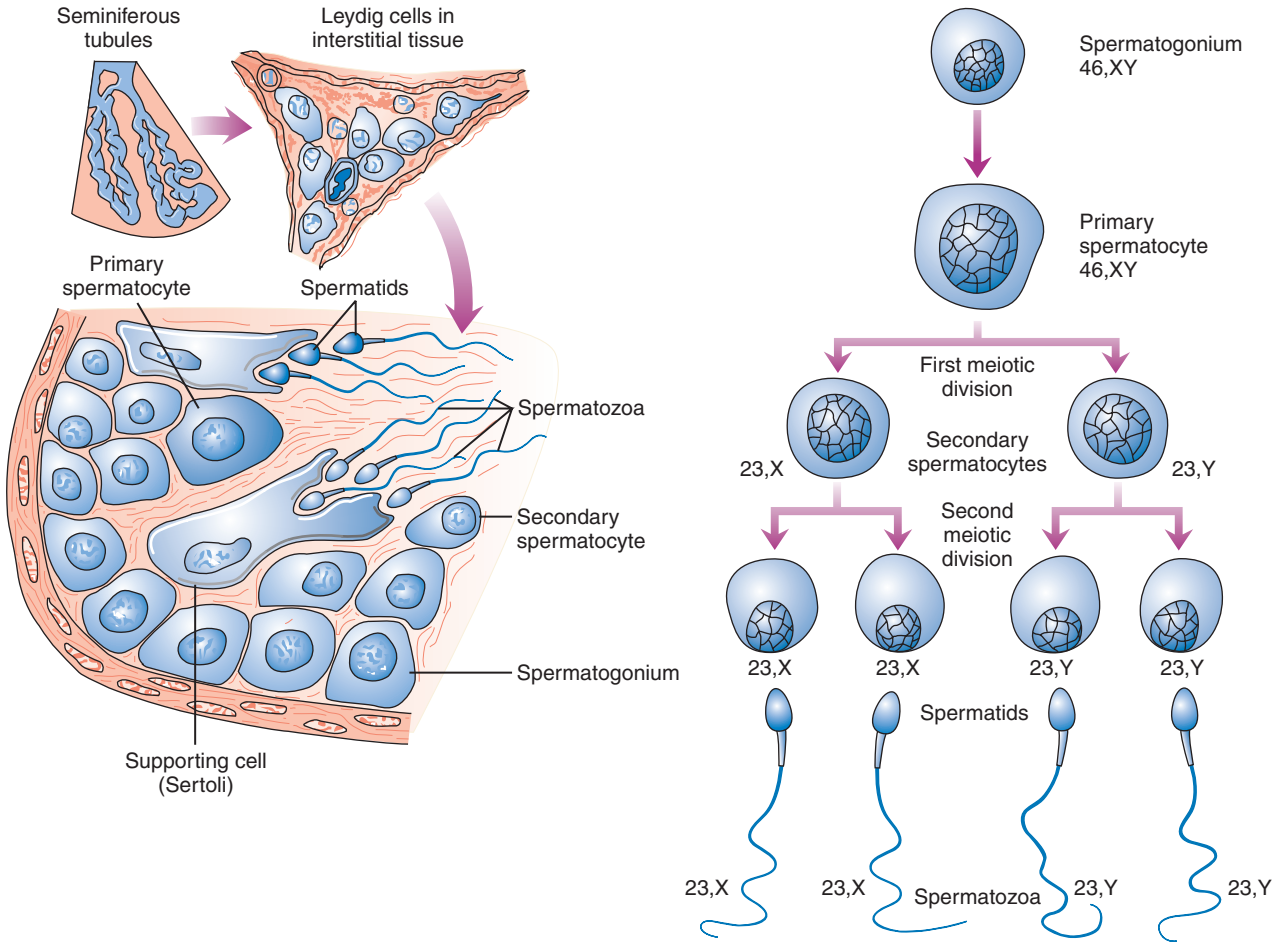


FIG 30.15 Process of meiosis in spermatogenesis.

GERIATRIC CONSIDERATIONS

Male Fertility and Reproductive Organ Function

A decline in male fertility and reproductive organ function usually occurs with aging. However, the magnitude of functional decline of the male reproductive organs is variable. For example, some elderly men maintain their fertility into their 70s and 80s.

Male reproductive organ variability is due to organ-specific tissue changes. Active male germinal cells continue to produce spermatozoa (spermatogenesis), although the number of sperm produced declines proportionally over time. The testes become smaller as a result of the increased amount of connective tissue, fibrosis of the tubules, and decreased numbers of capillaries. The number of active seminiferous tubules declines with aging. The number of Leydig cells that produce testosterone decreases, leading to a decrease in testosterone level with aging.

The arteries and veins in the penis become increasingly sclerotic. The penis itself becomes smaller, with an increase in fibroelastic tissue. Penile sensation is decreased. Sexually, the aging male has a longer refractory period after orgasm and decreased force of ejaculation.

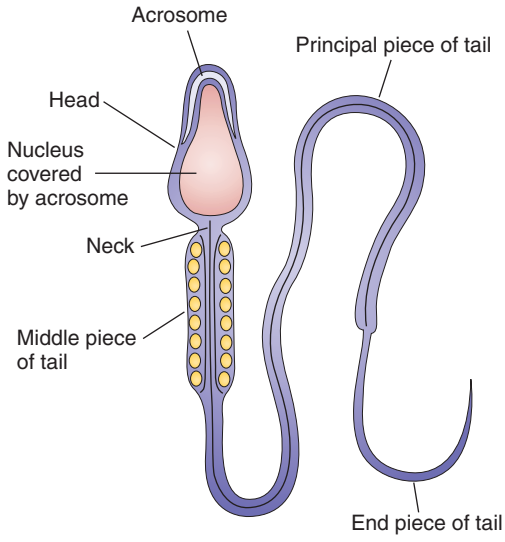


FIG 30.16 Anatomy of a mature sperm cell. (From Moore KL et al, editors: *The developing human: clinically oriented embryology,* ed 10, Philadelphia, 2016, Saunders.)

duct. It passes through the scrotum, traverses the inguinal canal into the pelvis, and then passes behind the bladder to enter the prostatic urethra at the ejaculatory ducts of the verumontanum. The terminal portion of the vas deferens is known as the *ampulla.* It is joined by the ducts of the seminal vesicle before entering the ejaculatory ducts.

As one passes in a proximal-to-distal direction from the efferent ducts to the vas deferens, the thickness of the muscle gradually increases. In the vas deferens, three interconnected smooth muscle layers form a thick muscular wall, with the ratio of wall thickness to lumen being the greatest in any human structure. This thick muscular wall facilitates rapid sperm transport at the time of ejaculation.

Aside from serving as a conduit and storage depot for spermatozoa, the epididymis facilitates the maturational processes. Most studies have demonstrated that sperm taken directly from the testes are incapable of naturally fertilizing eggs. The development of motility and increased fertility are acquired during transit through the epididymis.

Because epididymal sperm are probably immotile, other mechanisms must be involved in their transport. Initially, spermatozoa are carried into the efferent ducts by fluid from the rete testis. Within the efferent ducts, motile cilia within the lumen function to reabsorb testicular fluid and help move spermatozoa into the epididymis. Within the epididymis, the spermatozoa are probably transported by rhythmic contraction of the smooth muscle cells.

Ejaculation accelerates the passage of spermatozoa through the vas deferens and distal end of the epididymis. In young men, approximately 200 million sperm can be found in the reservoir of the epididymis. About 50% are found in the cauda region. With ejaculation, sperm from the distal part of the epididymis and vas deferens are deposited into the prostatic urethra, where they account for less than 10% of the normal ejaculate.

Erection, Emission, and Ejaculation

To penetrate the vagina and deposit sperm, the penis must be erect. The physiology of erection is a complicated interaction of vascular, neurologic, and hormonal factors. Although erection has classically been thought of as a parasympathetic function, it is more complex. Erection may be mediated by either local stimulation, which causes a reflexogenic erection through the sacral spinal cord, or psychological stimulation, which causes a psychogenic erection through cerebral centers. The presence of erections in patients with spinal cord injuries attests to the presence of reflex erections. Such patients have an intact sacral spinal cord and its reflex arc of afferent and efferent nerves below the site of spinal cord injury.

The penis receives sensory innervation from the pudendal sensory nerves entering the sacral spinal cord. The pudendal nerve is a mixed nerve that provides motor innervation to the pelvic floor musculature and penile sensory fibers. The efferent nerve to the erectile tissue is provided by sacral parasympathetic fibers. Although erection is possible in patients with spinal cord injuries, in intact men it is a much more controlled process influenced to a great extent by the cerebral cortex. Impulses may traverse the spinal cord from the cerebral cortex in the lateral columns and exit the spinal cord through sacral parasympathetic and possibly the thoracolumbar sympathetic nerves as well.

During erection, the vascular spaces that make up the spongy vacuous tissues of the corpora cavernosa and corpus spongiosum fill with blood. The relaxation of smooth muscle tone in these structures that allows filling and subsequent penile erection is modulated by nitric oxide. Research indicates that erectile function cannot be fully explained by parasympathetic or sympathetic mechanisms; this observation has led to consideration that nonadrenergic and noncholinergic neuromodulators are also involved in such function.

Ejaculation may be divided into two phases: *emission* and *ejaculation.* During emission, secretions from the periurethral glands, seminal vesicles, and prostate are deposited with sperm from the vasa deferentia and the cauda epididymis into the prostatic urethra. Control of emission is mediated primarily through the sympathetic nerves, which stimulate contraction of smooth muscle in these genital structures.

With ejaculation, the bladder neck or internal sphincter closes. This closure is also mediated through the sympathetic nervous system. Next, the external sphincter relaxes and the perineal and bulbourethral muscles surrounding the bulb of the corpus spongiosum contract and expel the ejaculate from the posterior urethra and through the urethral meatus.

The physiologic function of the secretory products of the accessory sex glands is uncertain. These secretions make up most of the seminal plasma, with the sperm and testicular fluid probably composing less than 10% of the final ejaculated semen volume. Although some investigators have demonstrated that sperm removed directly from the epididymis are capable of fertilization, these secretions most likely optimize conditions for sperm motility, survival, and transport in both the female and the male reproductive tracts.

Capacitation

Capacitation of the spermatozoa refers to the multiple changes that activate the sperm and enhance their ability to participate in the final process of fertilization. Although sperm are anatomically complete and highly motile when ejaculated, the complex process of capacitation is necessary before the sperm are actually capable of fertilizing the egg. The capacitation process occurs over a period of 1 to 10 hours and occurs in sperm only after they have been introduced into the vagina of the female. Once the sperm are inside the female, the uterine and fallopian tube fluids wash away the various inhibitory factors that had suppressed sperm activity in the male genital ducts. During the time that the spermatozoa were in the fluid of the male genital ducts, they were continually exposed to many floating vesicles from the seminiferous tubules containing large amounts of cholesterol. This cholesterol, continually donated to the cellular membrane covering the sperm acrosome, toughens the outside membrane and prevents release of its enzymes. After ejaculation, the sperm that are deposited in the vagina swim away from the cholesterol vesicles upward into the uterine fluid, and they gradually lose much of their excess cholesterol during the next few hours. As the cholesterol is lost, the membrane at the head of the sperm becomes much weaker.

The membrane of the sperm head also becomes much more permeable to calcium ions. Large amounts of calcium enter the sperm to increase the powerful whiplike motion of the flagellum beyond its previously weak, undulating motion. In addition, the calcium ions probably alter the intracellular membrane covering the leading edge of the acrosome, thus making it possible for the acrosome to release its enzymes very rapidly and easily as the sperm penetrates the granulosa cell mass surrounding the ovum. These enzymes are released even more rapidly and easily as the sperm attempts to penetrate the zona pellucida of the ovum itself.

Acrosome Reaction

The head of a sperm is essentially a highly compact package of genetic chromatin material covered by a specialized acrosome and acrosomal (head) cap. Stored in the acrosome of the sperm are large quantities of hydrolytic (water-splitting) enzymes that are released during capacitation. The specialized acrosomal enzymes first break down cervical mucus to allow sperm to pass into the uterus and uterine tubes. If an ovum is present in the female reproductive tract when semen is introduced, continued release of acrosomal enzymes results in digestion of proteins in the structural elements of the outer covering of the egg. A high sperm

count is essential for male fertility because the female ovum, once it is expelled from the ovarian follicle into the abdominal cavity and fallopian tube, contains multiple layers of granulosa cells. Before a sperm can fertilize the ovum, it must first pass through the granulosa cell layer, and then it must penetrate the thick covering of the ovum itself, the zona pellucida. It is believed that the acrosomal enzyme hyaluronidase plays an important role in opening pathways between the granulosa cells so that the sperm can reach the ovum.

On reaching the zona pellucida of the ovum, the anterior membrane of the sperm binds specifically with a receptor protein in the zona pellucida. Then the entire anterior membrane of the acrosome rapidly dissolves, and all the acrosomal enzymes are immediately released. Within minutes, these open a penetrating pathway for passage of the sperm head through the zona pellucida.

The head at first enters the perivitelline space lying beneath the zona pellucida but outside the membrane of the underlying oocyte. Within 30 minutes, the membranes of the sperm head and the oocyte fuse; the sperm genetic material enters the oocyte to cause fertilization, and the embryo begins to develop.

KEY POINTS

- Normal male sexual development and spermatogenesis depend on the appropriate secretion of reproductive hormones. GnRH, secreted by the hypothalamus, induces the anterior pituitary gland to secrete LH and FSH. The bloodstream receives these hormones, which then travel to the testes where they bind to testicular cells.
- Leydig cells in the testes possess LH receptors and respond by increasing production of testosterone. Testosterone and related androgens are necessary for maturation of the male external genitalia. The function of FSH is less well understood, but appears to be necessary for spermatogenesis.
- Spermatogenesis occurs when germinal cells within the seminiferous tubules undergo meiosis to form haploid (23 chromosomes) spermatids. Spermatids then develop into mature spermatozoa with the assistance of Sertoli cells. Sperm require 70 days to mature, and they are continuously produced and released into the epididymis.
- Sperm are well formed to perform their function in that they have a highly stabilized nucleus that is resistant to physical trauma, a mobile tail (axoneme) for swimming, and specialized enzymes to enhance penetration of the egg.
- Sperm traveling from their site of origin in the testes must pass through approximately 6 m of tubules before arriving at the penile meatus. This tubular system includes the seminiferous tubules in the testes, epididymis, vas deferens, and urethra. About 200 million sperm may be stored in the epididymal reservoir. Increased motility and fertility appear to be acquired by sperm as they pass through the epididymis. Sperm account for less than 10% of the ejaculate volume.
- The physiologic process of erection is a complex interplay of vascular, neurologic, and hormonal factors. The sacral parasympathetic nerves provide important innervation to the penis. Acetylcholine from parasympathetic nerves causes relaxation of penile smooth muscle with subsequent engorgement.
- The sympathetic nervous system mediates the process of ejaculation. Sympathetic actions include contraction of the internal sphincter to prevent retrograde ejaculation and relaxation of the external sphincter to allow emission.
- Sperm deposited in the vagina undergo further changes in a process known as *capacitation.* This process improves the chances of sperm successfully producing fertilization of an egg. Enzymes are released (acrosome reaction) to facilitate penetration of the ovum, a process that further increases the chances of successful fertilization.

SUMMARY

The male genitourinary tract may be divided into upper and lower tracts, with the upper tract composed of the kidneys and ureters and the lower tract composed of the bladder and urethra. Auxiliary genital glands that lie adjacent to or surround the urethra include the prostate, seminal vesicles, and bulbourethral glands. The external genitalia of the male consist of the scrotum, testes, epididymis, and penis.

Embryologic development of the male and female genital and urinary systems is closely related. The nephric system develops progressively through three distinct phases: the pronephros, mesonephros, and metanephros. The gonads are derived from the urogenital ridge, from which the primitive kidney also forms. Finally, the genital duct systems develop from two different but related ducts adjacent to the undifferentiated gonads, the müllerian ducts and the mesonephric, or wolffian, ducts.

Male reproductive function depends on an intact hypothalamic-pituitary-testicular endocrine axis. Spermatogenesis takes place in the seminiferous tubules. Spermatozoa mature in their transit through the male reproductive tract. Through erection, emission, and ejaculation, sperm enter the vagina. Through capacitation and the acrosome reaction, spermatozoa acquire the ability to fertilize ova residing in the female reproductive tract.

RESOURCES

Applegate EJ: *The anatomy and physiology learning system: textbook*, ed 4, St Louis, MO, 2011, Saunders.

Black JM, et al: *Medical-surgical nursing: clinical management for positive outcomes*, ed 6, Philadelphia, 2001, Saunders.

Carroll R: *Problem-based physiology*, Philadelphia, 2010, Saunders.

Hall JE: *Guyton and Hall textbook of medical physiology*, ed 13, Philadelphia, 2016, Saunders.

Hinman F, Jr: *Atlas of urosurgical anatomy*, Philadelphia, 1993, Saunders.

Moore KL, et al, editors: *The developing human: clinically oriented embryology*, ed 10, Philadelphia, 2016, Saunders.

Nichols FH, Zwelling E: *Maternal-newborn nursing: theory and practice*, Philadelphia, 1997, Saunders.

Sauerland EK: *Grants dissector*, ed 10, Baltimore, 1991, Williams & Wilkins.

Tanagho EA, McAninch JW, editors: *Smith's general urology*, ed 13, East Norwalk, CT, 1992, Appleton & Lange.

Wein AJ, et al, editors: *Campbell-Walsh urology*, ed 11, Philadelphia, 2016, Saunders.

31

Alterations in Male Genital and Reproductive Function

Marvin Van Every

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- What are the common causes of and clinical findings in priapism?
- What are the common causes of primary and secondary erectile dysfunction?
- What are the usual clinical manifestations and significance of testicular cancer, testicular torsion, cryptorchidism, and hydrocele or spermatocele?
- What clinical manifestations would lead to a suspicion of prostatitis, and how would confirmed prostatitis be treated?
- How can benign prostatic hyperplasia be distinguished from prostate cancer?
- What clinical manifestations are indicative of prostatic enlargement?

CHAPTER OUTLINE

Disorders of the Penis and Male Urethra, 641
Congenital Anomalies, 641
Micropenis, 641
Urethral Valves, 642
Urethrorectal and Vesicourethral Fistulas, 642
Hypospadias, 642
Epispadias, 643
Acquired Disorders, 643
Priapism, 643
Phimosis and Paraphimosis, 644
Peyronie Disease, 644
Urethral Strictures, 644
Erectile Dysfunction, 645
Premature Ejaculation, 645
Infectious Disorders, 645
Neoplastic Disorders, 647
Neoplasms of the Penis, 647
Disorders of the Scrotum and Testes, 647
Congenital Disorders, 647
Cryptorchidism, 647
Acquired Disorders, 648
Hypogonadism, 648
Hydrocele, 648
Spermatocele, 649
Testicular Torsion, 649
Male Infertility, 649
Infectious Disorders, 650
Epididymitis, 650
Fournier Gangrene, 650
Neoplastic Disorders, 650
Neoplasms of the Testis, 650
Disorders of the Prostate, 651
Benign Prostatic Hyperplasia, 651
Prostatitis, 652
Prostate Cancer, 652

The male genital system is susceptible to numerous congenital, acquired, and infectious conditions and, to a lesser extent, neoplasms. These disorders may interrupt the normal functions of urinary excretion and sexual function and fertility and directly affect the quality of life. This chapter will identify and explain the most common conditions that come to the attention of practitioners.

DISORDERS OF THE PENIS AND MALE URETHRA

Congenital Anomalies

Micropenis

Micropenis is defined as a small, normally formed penis with a stretched length more than two standard deviations below the mean. The normal

range in newborns is 2.0 to 3.5 cm, so micropenis may be defined as a stretched length of less than 1.9 cm.

Etiology and pathogenesis. Penile development and growth are both testosterone dependent. Therefore micropenis may result from defects in testosterone production or a deficiency that results in poor growth of the organs that are targets of this hormone.

Diagnoses and treatment. Patients with micropenis must be evaluated for endocrine abnormalities. To check for these, one should measure serum levels of testosterone, luteinizing hormone, and follicle-stimulating hormone (FSH). A karyotype should also be obtained. Depending on the results of these measurements, the problem may be determined to involve the hypothalamic-pituitary axis (Prader–Willi and Kallmann syndromes) or to be some form of a testicular disorder (Klinefelter syndrome).

Treatment depends on administering testosterone, either intramuscularly (IM) or topically, to stimulate penile growth. Such treatment requires caution because skeletal growth may be altered by premature closure of the epiphyseal growth plates in the long bones. In rare cases, when micropenis fails to respond to testosterone, a female sex assignment may be indicated.

Urethral Valves

The vast majority of urethral valves are posterior in location and occur in the distal prostatic urethra. They are the most common cause of urinary obstruction in male newborns and infants. These valves are mucosal folds that resemble thin membranes and cause obstruction when the child attempts to void (Fig. 31.1).

Etiology. Many theories have been given to explain how valves develop. It has been suggested that several different processes may occur to form different posterior valves. Most commonly, posterior valves may result from abnormal insertion and persistence of the distal Wolffian ducts. Less frequently, a persistent urogenital membrane may result in valves and obstruction.

Clinical manifestations. Children with posterior valves may have variable degrees of obstruction. In the most severe cases, intrauterine renal failure may cause oligohydramnios (decreased amniotic fluid), pulmonary hypoplasia (incomplete lung development), and either stillbirth or extreme distress at the time of delivery. More frequently, inability to void is noted shortly after birth (normal voiding occurs within 48 hours after birth), or the infant has abdominal masses representing a thickened palpable bladder or hydronephrotic kidneys. Varying degrees of azotemia and renal failure occur with this scenario.

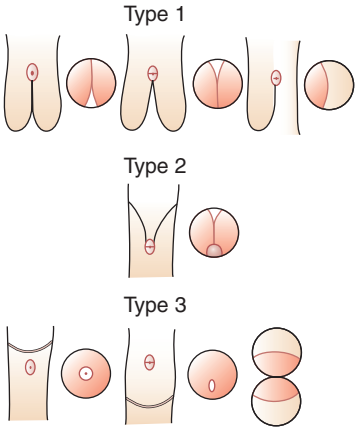


FIG 31.1 Posterior urethral valves. (Redrawn from Young HH, Frontz WA, Baldwin JC: Congenital obstruction of the posterior urethra. *J Urol* 1919;3:289.)

Finally, urinary ascites (extravasated urine in the peritoneum) may result from a urinary leak that is usually difficult to localize. In an infant with abdominal distention, the diagnosis of urethral valves is confirmed by a plain abdominal radiograph showing the bowel "floating" in the center of the abdomen. Prenatal ultrasounds often suggest the diagnosis before birth so immediate evaluation and treatment can take place upon delivery.

Older infants with a urethral valve are less likely to have a palpable kidney or ascites. Rather, urinary tract infection, poor stream with straining to void, or occasionally hematuria may be present. Urethral valves in these older male infants may not produce much obstruction, thus making the diagnosis more difficult.

Treatment. Management of posterior valves involves initial management of the metabolic abnormalities with appropriate fluid management and electrolyte replacement. In patients with a urinary tract infection, drainage of urine with a urethral or occasionally a suprapubic catheter is necessary. Finally, ablation of the valves with an endoscopic resectoscope should be performed. In infants, this step may be delayed and a cutaneous vesicostomy made to temporarily divert and drain the urine. This approach reduces the risk of traumatizing the infant's delicate urethra, which may create urethral stricture disease.

Rarely, urethral valves are located anteriorly in the penile urethra. Valves in this location are a very rare congenital anomaly and most likely represent urethral dilation or a diverticulum proximal to the valve. Endoscopic resection will correct the problem.

Urethrorectal and Vesicourethral Fistulas

Etiology. Urethrorectal and vesicourethral fistulas are rare and almost always associated with an imperforate anus. Failure of the urorectal septum to develop completely leads to persistent communication between the rectum posteriorly and the urogenital tract anteriorly.

Clinical manifestations and treatment. Children with a urethrorectal or vesicourethral fistula may pass fecal material and gas through the urethra. If the anus has formed normally with an external opening, urine may drain through the rectum. The diagnosis is made with cystoscopy and contrast-enhanced radiography to delineate a blind rectal pouch or communication. Surgery is needed to resect the fistula and open the imperforate anus.

Hypospadias

In hypospadias, the urethral meatus is located on the ventral undersurface of the penis or on the perineum (Fig. 31.2). The condition may occur with varying degrees of severity. In the least severe cases, the meatus is located distally on the penis, either at the corona or on the undersurface of the glans. With increasing severity of the condition, the meatus assumes a more proximal location and is more often associated with chordee, or curvature of the penile shaft (Fig. 31.3).

Etiology and treatment. Hypospadias is the result of incomplete fusion of the urethral folds, so the meatus may be found anywhere along the phallus from the perineum to the glans. In the majority of cases hypospadias occurs distally, with about 85% of all cases involving the glans or corona. Because incomplete fusion of urethral folds may indicate insufficient masculinization, it is recommended that the more severe penoscrotal and perineal openings be evaluated for conditions of intersex.

Management of hypospadias involves surgical repair. Many procedures are available, with several repairs indicated for each type of hypospadias. The goal of surgery is a good overall cosmetic appearance that will allow the patient to stand and direct his urinary stream and will also allow normal sexual function.

FIG 31.2 Varying forms of hypospadias. **A,** Glanular hypospadias. **B,** Subcoronal hypospadias. Note the dorsal hood of foreskin. **C,** Penoscrotal hypospadias with chordee. **D,** Perineal hypospadias with chordee and partial penoscrotal transposition. **E,** Megameatal variant of hypospadias diagnosed after circumcision; note absence of hooded foreskin. (From Kliegman RM et al: *Nelson textbook of pediatrics,* ed 19, Philadelphia, 2011, Saunders.)

Epispadias

In epispadias, the urethra opens on the dorsal aspect of the penis at a point proximal to the glans (see Fig. 31.3). Although much less common than hypospadias, it can be considerably more disabling.

Etiology and treatment. The embryogenesis of epispadias is related to another congenital condition: exstrophy of the bladder. In this condition, the abdominal wall fails to form below the level of the umbilicus. At birth, the back wall of the bladder is exposed to the external environment. The development of epispadias is simply a mild degree of exstrophy, with a deficiency of abdominal wall formation present inferiorly. Most commonly, the defect extends proximally to involve the urinary sphincter and results in urinary incontinence. Less commonly, the urethral meatus is located more distally along the dorsum of the penis and is accompanied by urinary continence because the sphincter is not affected.

Management of exstrophy and proximal epispadias with incontinence is difficult and involves staged surgical procedures to reconstruct a continent bladder neck and a functional urethra. The less common distal epispadias is usually managed with tubular reconstruction procedures similar to those used for repair of hypospadias.

Acquired Disorders

Priapism

Priapism may be defined as a painful, persistent erection. The patient usually reports several hours of painful erection in which the corpora cavernosa are tense with congested blood. The corpus spongiosum and glans are characteristically soft and uninvolved.

Etiology and treatment. The causes of priapism are multiple. Most cases are idiopathic, with the next most common cause being sickle cell disease. Other etiologic factors include use of anticoagulant therapy, presence of diabetes mellitus or leukemia, and use of certain

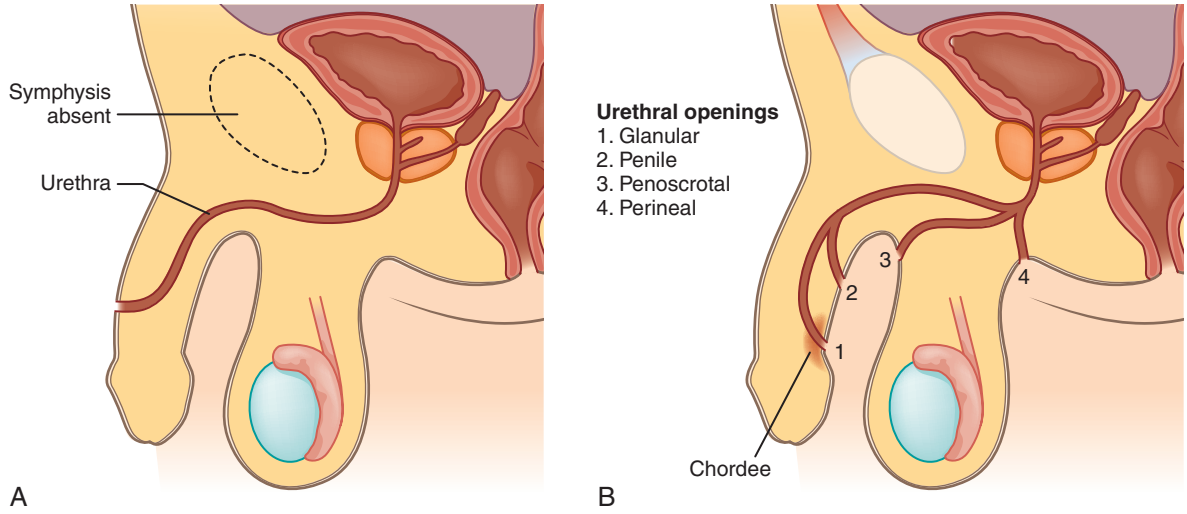


FIG 31.3 Epispadias **(A)** and hypospadias **(B)** showing possible locations of the urethral meatus.

antidepressant medications. Recently intracavernosal injection of vasoactive substances for the management of erectile dysfunction (ED) has been noted to cause priapism. On a rare occasion, oral erectile dysfunction medications can cause priapism. Although multiple causes exist, the common abnormality is probably an obstruction of venous drainage resulting in the buildup of viscous, poorly oxygenated blood in the corpora. If the process is allowed to continue, fibrosis of the corpora cavernosa will eventually occur and may cause ED.

Management of priapism may involve a combination of measures, depending on the cause and duration of the condition. Initial therapy for priapism secondary to sickle cell disease includes sedation and oxygen. For the management of priapism secondary to other causes, initial measures may include aspiration of blood from the corpora, as well as injection of α-adrenergic agents. If the priapism remains refractory to these initial measures, a surgical shunting procedure may be necessary in which a shunt is created between the erect corpora cavernosa and the detumesced corpus spongiosum.

Phimosis and Paraphimosis

Etiology, clinical manifestations, and treatment. Phimosis occurs when the uncircumcised foreskin cannot be retracted over the glans of the penis (Fig. 31.4A). Phimosis is usually the result of chronic inflammation and infection from poor hygiene. Calculi and squamous cell carcinoma may occur, although it is usually the presence of erythema, tenderness of the phimotic foreskin, or a discharge that prompts the patient to seek medical attention. Management involves treating the infection with antifungal agents or antibiotics, followed by circumcision.

Paraphimosis, on the other hand, occurs when a foreskin that has been retracted over the glans up onto the shaft of the penis cannot be replaced in its normal position (see Fig. 31.4B). In this condition, which is usually secondary to chronic inflammation under the foreskin, a constricting ring of skin forms around the base of the retracted glans. The constriction causes venous congestion of the glans, with further swelling and edema making the condition worse. Treatment entails reducing the paraphimotic foreskin back over the glans, which can usually be accomplished by compressing the glans to reduce the edema. Occasionally, a slit or formal circumcision is needed to manage the problem.

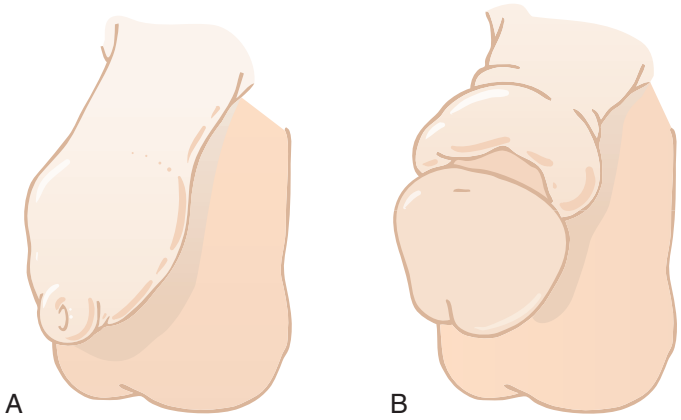


FIG 31.4 **A,** Phimosis. **B,** Paraphimosis.

Peyronie Disease

Etiology and treatment. Peyronie disease refers to the formation of palpable, fibrous plaque on the surface of the corpora cavernosa. This plaque subsequently causes curvature of the penis with painful, incomplete erections. No satisfactory treatment for this disease is available, although some cases may remit with time. Conservative therapies that have had limited success include the use of aminobenzoate potassium (Potaba), colchicine, or pentoxifylline (Trental). Intralesional injection of collagenase *Clostridium histolyticum* with penile modeling can correct the problem in some patients. In addition, several operative procedures have been developed. These procedures involve excising the plaque and repairing the corporal defect with a graft or plicating the corporal bodies.

Urethral Strictures

Etiology. Urethral strictures are fibrotic narrowings of the urethra and are usually composed of scar tissue. Most acquired strictures are due to a prior infection such as gonorrhea, or trauma. Traumatic causes can be both iatrogenic, such as large urethral catheters and instrumentation, and noniatrogenic, such as straddle injuries.

Clinical manifestations and treatment. A decreased urinary stream is the most common complaint. Other common complaints include urethral discharge, infection, and urine retention. Urethral strictures are usually diagnosed by cystoscopy or retrograde urethrography, which would demonstrate a narrowing of the urethra. Management of urethral strictures involves procedures to dilate, incise, or reconstruct the urethra, depending on the extent and duration of the stricture.

Erectile Dysfunction

Erectile dysfunction (ED) is the inability to achieve or maintain an erection sufficient for satisfactory sexual performance. It is highly prevalent in aging men, affecting approximately 50% of men older than 60 years of age. Its prevalence and incidence are highly connected with risk factors such as hypertension; elevated cholesterol level; presence of diabetes mellitus and/or metabolic syndrome; and lifestyle choices such as smoking, obesity, and lack of exercise.

The physiologic process of penile erection is a complex interaction of the vascular, hormonal, and neurologic systems. ED may be primary or secondary. Primary ED refers to the inability to attain an erection throughout life and is often related to deep-seated psychiatric problems of some duration. Occasionally, vascular trauma sustained during early childhood or adolescence may account for primary ED.

Etiology. Far more common than primary ED is secondary ED. An individual with secondary ED is no longer able to achieve normal erections but did have normal erections in the past. The causes of secondary ED are multiple and may be discovered by examining the patient's medical history. Common causes of secondary ED are peripheral vascular disease, the use of certain medications, endocrine problems, trauma, iatrogenic causes (surgery), and psychological causes. To differentiate organic causes from psychogenic ED, one relies on the history, physical, and basic laboratory testing such as measurement of serum glucose and testosterone levels. Penile tumescence testing can also be utilized to make this distinction.

Arterial insufficiency of the penis may occur from obstruction of the arterial supply. Several processes may account for this obstructive arteriosclerosis. Stenosis of the arteries secondary to atheromatous plaque may be the most common etiologic factor. Diabetes mellitus not only may result in occlusion of arterial vessels, but also may cause a neuropathy of the pudendal nerve that might result in ED. Most investigators have suggested that ED may result from excessive venous drainage from the penis. This occurs because the blood is not adequately trapped in the corpora.

The list of medications that may cause ED is long. Several antihypertensive agents, including propranolol, monoamine oxidase inhibitors, and thiazides, have been associated with varying degrees of ED. Other medications linked to ED include phenothiazines, antihistamines, and some antidepressants.

Endocrinopathy accounts for a small percentage of ED cases. Pituitary dysfunction resulting in decreased or no secretion of luteinizing hormone may result in decreased secretion of testosterone. Primary failure of the testes may also cause decreased secretion of testosterone. Finally, excessive secretion of the hormone prolactin by the pituitary gland may result in low testosterone levels.

Trauma to the penis resulting in penile fractures and damage to penile erectile tissue may occasionally lead to partial or complete impotence. More common injuries include pelvic fractures with subsequent damage to the penile vascular and nervous supply. Iatrogenic trauma secondary to several commonly performed operations, including aortoiliac vascular surgery, and radical pelvic cancer operations may also result in ED. Peyronie disease is also associated with ED.

A newer concept in ED is the idea of vascular endothelial damage, which can be diffuse throughout the body. Some researchers believe ED may be an indicator of coronary artery disease.

Finally, it must be remembered that successful sexual function depends not only on intact vascular, hormonal, and neurologic systems, but also on intact psychological and social responses. Several psychological factors may be manifested as problems of low desire, erectile failure, or premature ejaculation. A discussion of the psychological contribution to ED is beyond the scope of this book.

Treatment. Management of ED requires an initial evaluation to differentiate organic causes from psychogenic causes. Further evaluation to distinguish among the various organic causes may then be needed. Once a psychogenic cause has been ruled out, several therapeutic options exist. Surgical options include the insertion of an inflatable or semirigid prosthetic device into the corpora cavernosa. Several investigators have discovered that intracavernous injection of various vasoactive substances can cause an erection. Several of these substances, including papaverine, phentolamine, and prostaglandin E_1, are commonly used and afford a nonsurgical treatment option.

Viagra, the first oral therapy for ED, is the citrate salt of sildenafil, a selective inhibitor of cyclic guanosine monophosphate (cGMP)–specific phosphodiesterase type 5 (PDE5). To understand its clinical pharmacology, a review of some of the physiologic mechanisms of erection follows. Briefly, erection of the penis involves release of nitric oxide in the corpus cavernosum during sexual stimulation. Nitric oxide then activates the enzyme guanylate cyclase, and the subsequently increased levels of cGMP produce smooth muscle relaxation in the corpus cavernosum and allow inflow of blood. Sildenafil has no direct relaxant effect on isolated human corpus cavernosum, but it enhances the effect of nitric oxide by inhibiting PDE5, which is responsible for degradation of cGMP in the corpus cavernosum. When sexual stimulation causes local release of nitric oxide, inhibition of PDE5 by sildenafil causes increased levels of cGMP in the corpus cavernosum, smooth muscle relaxation, and inflow of blood to the corpus cavernosum, which results in erection (Fig. 31.5). Sildenafil citrate at the recommended doses appears to have no effect in the absence of sexual stimulation and affords another nonsurgical treatment option. Other PDE5 inhibitors for ED are Levitra and Cialis. In any particular patient, one of the three available oral medications may work better than the others. None of these agents should be used in conjunction with nitrate medications.

Another nonsurgical alternative entails the use of a vacuum device to sustain an erection. Finally, in specific cases of ED, surgical procedures may be done to revascularize the arterial supply of the penis or ligate the penile venous drainage.

Premature Ejaculation

Premature ejaculation (PE) is the most common male sexual dysfunction and is present in up to 30% of all males. The International Society for Sexual Medicine defines PE as male sexual dysfunction characterized by ejaculation that always or almost always occurs before or within 1 minute of vaginal penetration, as well as the inability to delay ejaculation upon all or almost all vaginal penetrations combined with negative personal consequences, such as distress, bother, frustration, and/or the avoidance of sexual intimacy.

Etiology and treatment. The etiology of PE is not well defined, but can be both biological and psychosocial. The diagnosis is achieved with a careful medical and sexual history and physical examination.

Some medications for treatment are listed in Table 31.1. Further research will elucidate the causes and allow for better treatment in the future.

Infectious Disorders

Sexually transmitted infections (STIs) are common in the male genital system. A number of infections are sexually transmitted and affect the

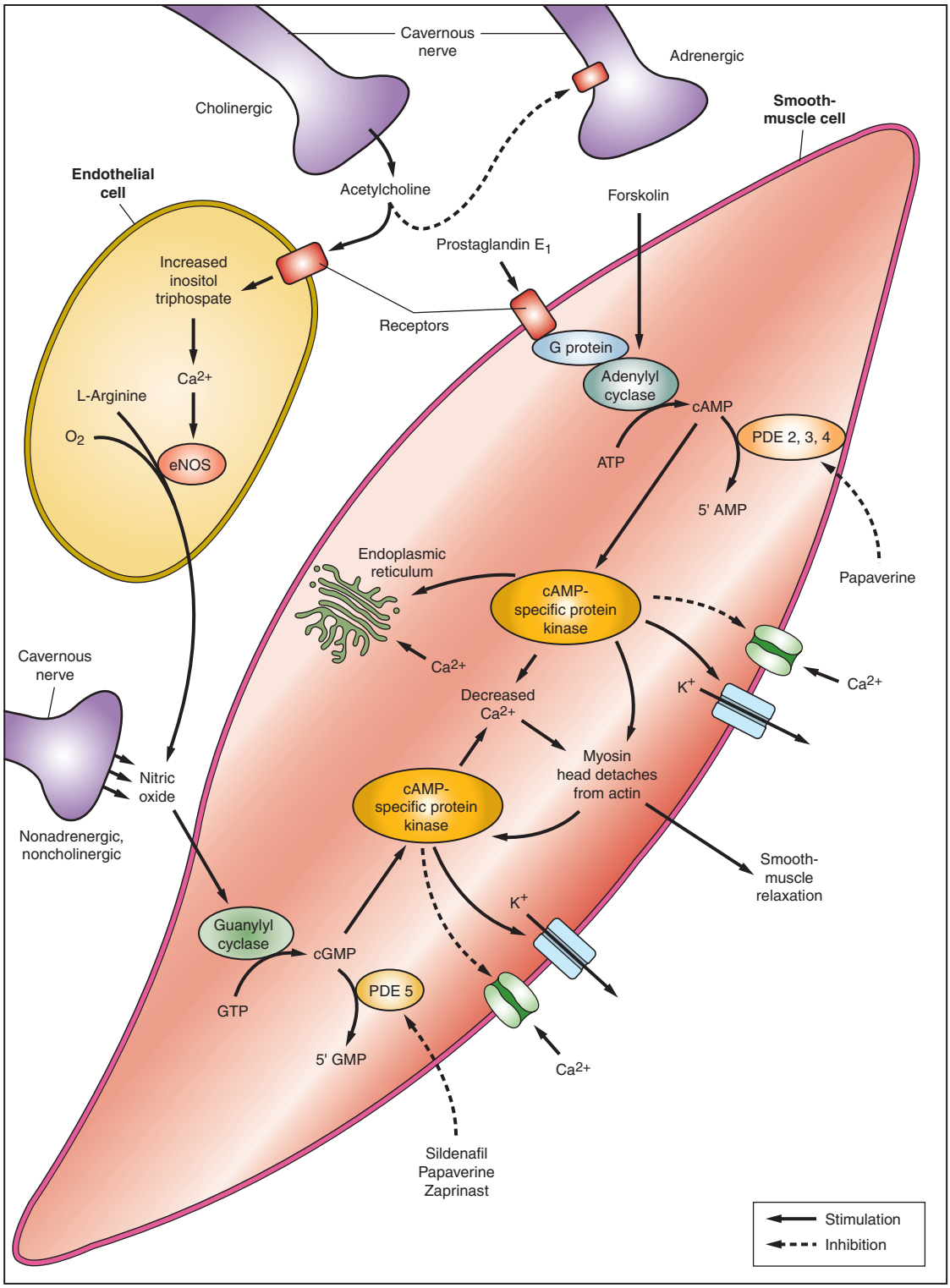


FIG 31.5 Mechanism of corpus cavernosum smooth muscle relaxation. Smooth muscle relaxation in the corpus cavernosum is the underlying mechanism of erection. The principal neurotransmitter is NO acting through cGMP and G-protein. Pharmacologic agents that produce erection act through this pathway by regulation of the intracellular balance of Ca^{2+} and K^+ concentrations. *ATP,* Adenosine triphosphate; *cAMP,* cyclic adenosine monophosphate; *cGMP,* cyclic guanosine monophosphate; *GTP,* guanosine triphosphate; *NO,* nitric oxide; *eNOS,* nitric oxide synthase; *PDE_5,* phosphodiesterase type 5. (Redrawn from Lue TF: Erectile dysfunction. *New Engl J Med* 2000;342:1802–1813.)

TABLE 31.1 Medical Therapy Options for the Treatment of Premature Ejaculation*

Oral Therapies	Trade Names†	Recommended Dose‡§
Nonselective Serotonin Reuptake Inhibitors		
Clomipramine	Anafranil	25–50 mg/day *or* 25 mg 4–24 hr preintercourse
Selective Serotonin Reuptake Inhibitors		
Fluoxetine	Prozac, Sarafem	5–20 mg/day
Paroxetine	Paxil	10, 20, 40 mg/day *or* 20 mg 3–4 hr preintercourse
Sertraline	Zoloft	25–200 mg/day *or* 50 mg 4–8 hr preintercourse
Topical Therapies		
Lidocaine/prilocaine cream	EMLA cream	Lidocaine 2.5%/prilocaine 2.5% 20–30 min preintercourse

*This list does not reflect order of choice or efficacy.
†Trade names listed may not be all-inclusive.
‡Peak plasma concentrations occur 2 to 8 hr postdose, and half-lives range from 1 to 3 days.
§Titrate doses from low to high based on response.

penis and urethra, including gonococcal urethritis, nongonococcal urethritis, syphilis, herpes, and genital warts. Gonococcal and nongonococcal urethritis and syphilis are effectively managed with antibiotics. Herpes and genital warts are associated with viruses and tend to be chronic, with intermittent recurrence. A more in-depth discussion of STIs is provided in Chapter 34.

Neoplastic Disorders

Neoplasms of the Penis

Etiology. Although cancer of the penis is rare in the United States and accounts for less than 0.2% of cancer deaths, its prevalence fluctuates widely among various locations. The causes are poorly understood, but phimosis of the foreskin accompanied by chronic inflammation has been thought to be the primary etiologic factor. The incidence of penile cancer among circumcised men is extremely low.

The majority of penile cancer cases are squamous cell carcinoma (97%). They usually occur on the glans or the inner surface of the foreskin. Metastasis occurs by lymphatic dissemination, with initial involvement of the palpable inguinal lymph nodes. Death from penile carcinoma is a result of uncontrolled lymphatic spread and subsequent necrosis of the overlying skin, debilitation, and sepsis.

The lesion of penile cancer is usually ulcerative and fungating in appearance and may be associated with pain, bleeding, and urethral discharge. Inguinal adenopathy is present in more than 50% of patients at the time of diagnosis, although frequently the adenopathy represents an inflammatory response secondary to the lesion rather than metastasis.

Treatment. Therapy for penile carcinoma depends on the stage of the lesion. Topical chemotherapy and radiation therapy may be considered for certain lesions. Larger distal penile lesions often require partial penectomy, whereas proximal lesions may require total penectomy with creation of a perineal urethrostomy. Finally, removal of the involved inguinal lymph nodes by inguinal lymphadenectomy may be performed in cases of suspected stage III disease. Systemic chemotherapy is also used to treat metastatic disease.

The prognosis of penile carcinoma depends on the stage of disease. The 5-year survival rate for men with tumors localized to the penis is 65% to 90%. With inguinal node involvement, 5-year survival rates drop to about 30% to 50%, and if distant metastases are present, the 5-year survival rate is low.

KEY POINTS

- Congenital disorders of the penis may result from hormonal deficiencies or abnormalities in embryonic development. Micropenis, for example, is usually a result of testosterone deficiency. Urethral valves, fistulas, and malpositioning of the urinary meatus (hypospadias, epispadias) are related to abnormal embryonic development.
- Priapism is a persistent, painful erection, most commonly of unknown cause. Priapism may occur in conditions that cause obstruction of venous drainage, including sickle cell anemia, anticoagulant therapy, diabetes mellitus, certain antidepressant medications, and PDE5 inhibitors such as Viagra.
- Phimosis and paraphimosis are disorders of the foreskin. Phimosis is associated with chronic inflammation and poor hygiene and results in a foreskin that cannot be retracted. Paraphimosis refers to a foreskin that remains retracted and cannot be returned to its normal position.
- Urethral strictures may be congenital or acquired. Most acquired strictures are secondary to gonorrheal infection or urethral trauma. Weak urinary stream, bladder infections, and retained urine are common manifestations.
- Erectile dysfunction (ED) is the inability to achieve a sustained erection. Causes of ED are categorized as primary and secondary. Primary ED is rare and is usually related to adolescent vascular trauma or psychiatric problems. Secondary ED may be due to a variety of factors, including vascular disease, medications, endocrine disorders, trauma, and psychological distress.
- A number of infections are sexually transmitted and affect the penis and urethra, including gonococcal urethritis, nongonococcal urethritis, syphilis, herpes, and genital warts. Gonococcal and nongonococcal urethritis and syphilis are effectively managed with antibiotics. Herpes and genital warts are associated with viruses and tend to be chronic, with intermittent recurrence.
- Penile neoplasms are rare, particularly in circumcised males. Phimosis and chronic inflammation may be important etiologic factors. Like other neoplasms, penile cancer has a better prognosis if managed before dissemination.

DISORDERS OF THE SCROTUM AND TESTES

Congenital Disorders

Cryptorchidism

Cryptorchidism means "hidden testis" and refers to any testis that occupies an extrascrotal position. The cryptorchid testis may be incompletely descended and as such be located intraabdominally, within the inguinal canal, or just external to the canal but above the scrotum. Occasionally the testis may emerge from the external ring of the inguinal canal and be misdirected into an abnormal extrascrotal position. In this situation the testis may be called *ectopic.* An ectopic testis may be located in any of several locations but is most commonly found in a superficial inguinal pouch (Fig. 31.6).

The incidence of cryptorchidism is about 0.7% to 1.0% of male infants at 1 year of age. The cause of the condition is uncertain but may be related to an intrinsic testicular defect or a subtle hormonal deficiency.

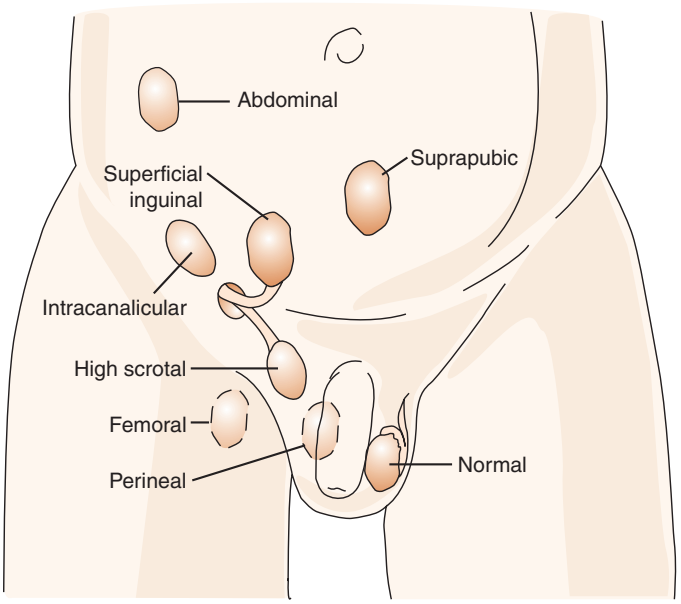


FIG 31.6 Sites of ectopic testes.

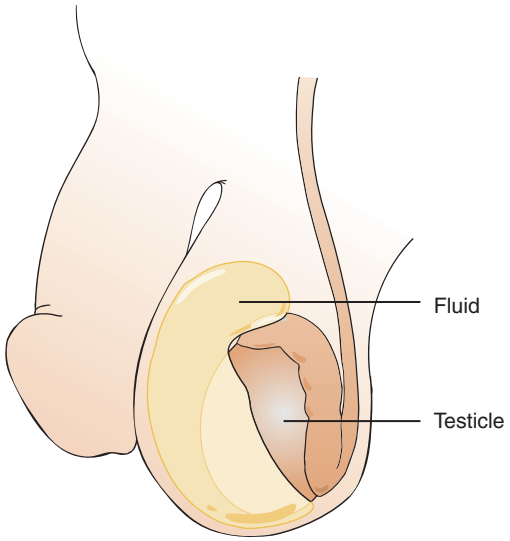


FIG 31.7 Hydrocele.

The incompletely descended, cryptorchid testis undergoes deleterious changes. The tubules become fibrotic, with a deficiency of spermatogenesis and subsequent infertility. More important is the increased incidence of testicular malignancy in cryptorchid testes. Several studies have revealed an increased prevalence of testicular tumors in subjects with a history of cryptorchidism.

Treatment. Because of the increased risk of malignancy and infertility, treatment at an early age to bring the testis into a normal scrotal position is recommended. An operative procedure *(orchiopexy)* is usually required, although in certain situations descent may be stimulated by the administration of human chorionic gonadotropin, which is given in a series of intramuscular injections.

Acquired Disorders

Hypogonadism

Androgen deficiency in the aging male (ADAM)—or andropause—is increasingly recognized as a problem for American men. The U.S. FDA estimates that between 4 and 5 million men suffer from hypogonadism. There are multiple causes (Box 31.1), but primary testicular failure is the most common etiology. The effects of ADAM include ED, loss of muscle tone, osteoporosis, and lipid metabolism changes. It is also associated with metabolic syndrome.

The diagnosis is made by determining the serum testosterone level. Other laboratory tests may include measurement of prolactin and luteinizing hormone (LH) levels.

Treatment. Fortunately, it is quite easy to treat because multiple forms of testosterone can be administered, including IM injections, patches, gels, and subcutaneous pellets. Patients receiving replacement therapy need to be monitored for prostate cancer and polycythemia.

Hydrocele

Etiology and clinical manifestations. A hydrocele consists of a fluid collection surrounding the testicle or spermatic cord and contained within the tunica or processus vaginalis (Fig. 31.7). Scrotal swelling in infants or young boys may indicate a hydrocele. These congenital hydroceles exist because of communication between the abdominal cavity and scrotum through the processus vaginalis. The scrotum is characteristically small and soft in the morning but larger and tense at night as it fills with fluid from the abdominal cavity.

Hydroceles may also develop secondary to scrotal injury, radiation therapy, infection of the epididymis, or testicular neoplasms. More commonly, however, the cause is uncertain, with the hydrocele developing slowly over time and occurring in middle-aged or elderly men. These acquired hydroceles may vary in size and consistency from small and soft to large and tense. The fluid is usually clear and yellow.

Treatment. Because a hydrocele is a benign condition, treatment is required only if the fluid collection becomes uncomfortable for the patient. Occasionally, a tense hydrocele might restrict circulation to the testicle. Management usually involves a surgical procedure to drain the fluid with either resection or plication of the hydrocele sac to prevent reaccumulation of the fluid. Aspiration of the hydrocele may be performed, although fluid often reaccumulates.

BOX 31.1 Etiologies of Adult-Onset Hypogonadism

Primary	Secondary
• Cancer chemotherapy	• Adrenal or testis tumor
• Excessive heat	• AIDS
• Iatrogenic injury	• Hemochromatosis
• Klinefelter syndrome	• Hepatic cirrhosis
• Malnutrition	• Hypercortisolemia
• Mumps orchitis	• Hyperprolactinemia
• Sickle cell disease	• Hypothyroidism
• Testicular radiation	• Medications (glucocorticoids, spironolactone, cimetidine, phenytoin, flutamide, ketoconazole)
• Testicular trauma	• Obesity
	• Occupational exposures
	• Pituitary ablation (surgical or external radiation therapy)
	• Pituitary or hypothalamic tumor
	• Severe systemic illness
	• Uremia

From Carson CC et al: Evolving concepts in the management of androgen deficiency in the aging male. *Urol* 2003;62:105–109.

Spermatocele

Etiology and treatment. Spermatoceles are painless, cystic masses containing sperm. Although they are usually small, they may be quite large and difficult to distinguish from a hydrocele. The cause of spermatoceles is uncertain; they may arise from the tiny tubules that connect the epididymis to the testis (vasa efferentia) or from the epididymis itself. Like hydroceles, spermatoceles need not be treated unless they become large enough to trouble the patient, in which case an operative procedure to excise the spermatocele may be performed.

Testicular Torsion

Torsion of the testicle is described as a twisting of the spermatic cord with subsequent compromise of the testicular vascular supply and testicular ischemia, followed by infarction (Fig. 31.8). Although torsion may occur in the neonatal period, the majority of cases occur in prepubertal boys.

Clinical manifestations. The diagnosis is suggested by the onset of severe pain in one testis, followed by swelling of the scrotum. Lower abdominal pain accompanied by nausea and vomiting may also occur. The condition may be differentiated from epididymitis (inflammation of the epididymis), which is also associated with scrotal swelling, by the presence of vascular echoes detected with a Doppler ultrasound. A testis made ischemic by torsion will not demonstrate blood flow, whereas the inflammation of epididymitis and its hypervascularity show increased vascular flow. Testicular nuclear scanning is another way to diagnose testicular torsion.

Treatment. Management of torsion involves an operation to open the scrotum, untwist the testis, and "pex" (secure) it to the scrotal wall. Because the chance of torsion also involving the contralateral testis is increased, the contralateral testicle is "pexed" to the scrotal wall as well. If detorsion is accomplished within 12 hours of the event, the prognosis for testicular viability is usually good. If torsion has been present for more than 24 hours, viability of the testis is doubtful.

Male Infertility

Etiology and pathogenesis. Approximately 15% of couples are unable to conceive after 1 year of unprotected intercourse. A male factor is solely responsible in about 20% and contributory in another 30% to 40% of infertility cases. An abnormal semen analysis is present in almost all cases of male infertility, but other factors may also be involved. Some causes of male infertility are identifiable and can be corrected, such as ductal obstruction, varicocele, or hypogonadotropic hypogonadism (a hormone abnormality) (Box 31.2). Other problems can be diagnosed but not corrected, such as bilateral testicular atrophy. Many cases are simply idiopathic or unexplained.

Diagnosis and treatment. The purpose of evaluating the male partner in infertile couples is to determine the cause of the infertility, if possible, so appropriate treatment and counseling can be accomplished. In some couples, treatment can result in pregnancy through intercourse. In others, more advanced treatments are necessary.

In some patients there is no effective treatment and they can be appropriately counseled. Other patients will have a risk of transmitting genetic abnormalities and will need to be made aware of that potential. In some patients a life- or health-threatening condition may underlie the infertility and need treatment.

As with all medical evaluations, there should be a complete and specific history and physical examination. Laboratory testing includes two semen analyses (Box 31.3) and usually an endocrine evaluation (testosterone and FSH). More evaluation could include postejaculatory urinalysis, transrectal ultrasound, scrotal ultrasound, specialized semen tests, and karyotyping.

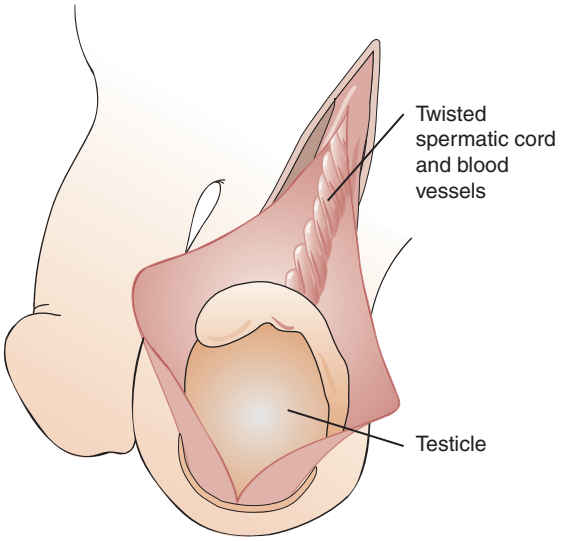


FIG 31.8 Testicular torsion.

BOX 31.2 Correctable Causes of Male Infertility

- Endocrinologic causes
- Testicular abnormalities
  - Retractile testes
  - Cryptorchidism
  - Torsion
- Obstruction
  - Vasal
  - Epididymal
  - Ejaculatory
- Disorders of ejaculation
- Sexual dysfunction
- Varicocele

From Galarneau GJ, Nagler HM: Cost-effective infertility therapies in the '90s: to treat or to cure? *Contemp Urol* 1999;11:32–45.

BOX 31.3 Semen Analysis: Reference Values

On at Least Two Occasions

Ejaculate volume 1.5–5.0 mL
pH >7.2
Sperm concentration >20 million/mL
Total sperm number >40 million/ejaculate
Percent motility >50%
Forward progression >2 (scale 0–4)
Normal morphology:
- >50% normal*
- >30% normal†
- >14% normal‡

And

Sperm agglutination <2 (scale 0–3)
Viscosity <3 (scale 0–4)

*World Health Organization, 1987.
†World Health Organization, 1992.
‡Kruger (Tygerberg) Strict Criteria, World Health Organization, 1999.
From Jarow J, Sigman M, Kolettis PN, Lipshultz LR, McClure RD et al: *The optimal evaluation of the infertile male: AUA Best Practice Statement,* Linthicum, MD, 2010, American Urological Association Education and Research, pp 1–38.

There are multiple treatments for male factor infertility. In men with varicoceles surgical repair or embolization can often improve fertility. Ductal obstruction can sometimes be surgically treated. In other cases more advanced treatment is necessary such as microsurgical epididymal sperm aspiration or percutaneous testicular sperm extraction. This allows for retrieval of sperm that can then be used for in vitro fertilization or intracytoplasmic sperm injection. Some patients will improve their semen analysis with endocrine therapy, and in some couples intrauterine insemination will be successful. For some couples the best option may be donor sperm or adoption.

Infectious Disorders

Epididymitis

Etiology. Epididymitis, or inflammation of the testis, has several causes. It may occur as a result of trauma or the reflux of sterile urine up the vas deferens. However, the majority of cases are probably secondary to a bacterial cause, with both sexually transmitted organisms (*Neisseria gonorrhoeae* and *Chlamydia trachomatis*) and non–sexually transmitted organisms (*Pseudomonas* and *Escherichia coli*) involved.

Clinical manifestations and treatment. With epididymitis the scrotum may be enlarged, reddened, and tender. The pain may radiate along the spermatic cord into the inguinal area. Fever may also occur, as may urethral discharge, cystitis, and cloudy urine. Laboratory testing usually reveals an elevated white blood cell (WBC) count, and urine culture may reveal the infecting organism.

Treatment for the condition involves bed rest, scrotal support, and administration of antibiotics. In advanced cases, incision and drainage with the intravenous administration of antibiotics may be needed to effectively manage a resulting scrotal abscess. On rare occasions, the testicle may need to be removed.

Fournier Gangrene

Etiology, clinical manifestations, and treatment. Fournier gangrene is a severe but rare condition involving gangrenous necrosis of the scrotum. Symptoms are pain and swelling of the scrotum, fever and chills, and sepsis. The diagnosis can often be made by history and physical examination. Additional tests such as ultrasound, computed tomography (CT), or magnetic resonance imaging can also be helpful in diagnosis. Usually, an underlying disease such as diabetes, alcoholism, or another general debility predisposes the patient to such an aggressive infection. Extravasation of infected urine from urethral trauma, a perforated urethral diverticulum, or a non–urinary tract source such as a perirectal abscess may act as the source of infection. Treatment, which must be instituted swiftly, includes incision and drainage of fluctuant areas and debridement of necrotic tissue along with administration of antibiotics. This often is a fatal disease unless it is managed quickly and aggressively.

Neoplastic Disorders

Neoplasms of the Testis

Although testicular tumors are rare, with a prevalence of 3.7 cases per 100,000 population, their peak incidence is in late adolescence to early adulthood. These neoplasms therefore represent the most common solid tumors of U.S. men ages 20 to 34 years. Testicular self-examination is an important tool for early detection because prompt treatment is associated with a higher success rate. Fig. 31.9 shows an ultrasound image demonstrating cancer in a testis.

Etiology. Although the cause of testicular tumors is uncertain, a strong association is seen between cryptorchidism and the subsequent development of malignancy. Nevertheless, the majority of patients with testicular tumors have no history of cryptorchidism, which suggests that several unrecognized factors may be contributing to the pathogenesis.

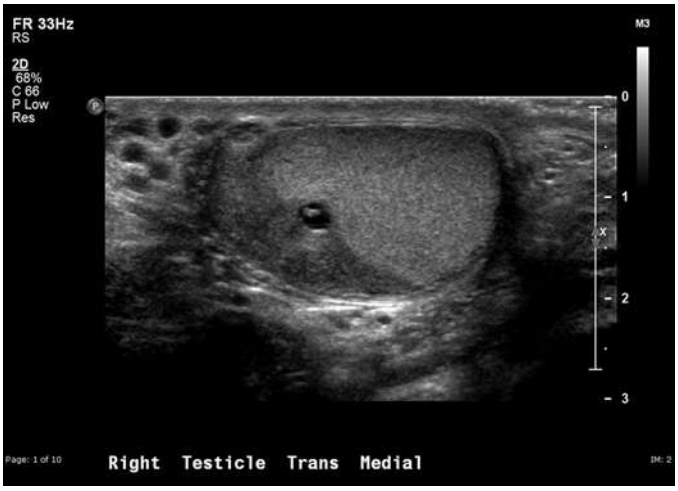


FIG 31.9 Testis ultrasound demonstrating cancer.

Histologically, testicular tumors may be considered in two groups. In the first group are nongerminal neoplasms, including tumors that originate from either the Leydig cells or other stromal tissue cells of the testis. In the second group are germinal neoplasms, which are derived from the germinal cells of the testis. This group accounts for the vast majority (95%) of testicular tumors. Germinal neoplasms may be further subdivided into two groups: seminomas and nonseminomas.

Treatment. Although germinal tumors may consist entirely of one histologic subtype, many contain elements of more than one subtype. Treatment and prognosis vary according to the subtype of germinal tumor. For example, seminoma in its early stages is exquisitely sensitive to and easily cured with radiation therapy. On the other hand, nonseminomatous germ cell tumors in the early stage are usually successfully managed with surgery. The prognosis is also variable. Pure choriocarcinomas are usually first seen at an advanced stage with distant metastases. Treatment is usually less effective for this aggressive lesion. However, the majority of germ cell tumors may be effectively managed even if lymph node metastases are present.

Except for choriocarcinomas, which disseminate by vascular means, testicular germ cell tumors usually metastasize through the lymphatic system. They usually disseminate in a stepwise manner, first involving the retroperitoneal lymph nodes lying adjacent to the great vessels. If unmanaged, the disease may progress to involve other lymph nodes and other organs such as the lungs.

Multiple staging systems have been devised to classify the extent of this disease. Most are a variation of the system proposed by Boden and Gibb in 1951.

Because the management of testicular tumors is complicated and somewhat controversial, a complete discussion is not possible here; however, several issues can be highlighted. After diagnosis of a testicular tumor, an operation to remove the testicle is performed. This procedure involves an inguinal incision with removal of the testis from the scrotum, followed by ligation and removal of the spermatic cord and testicle together. Histologic classification, additional staging studies, and other factors then determine further treatment. This treatment may involve close observation with frequent radiologic studies to determine new progression, surgery to remove the retroperitoneal lymph nodes (Fig. 31.10), chemotherapy, or radiation therapy. Some situations may call for a combination of these measures.

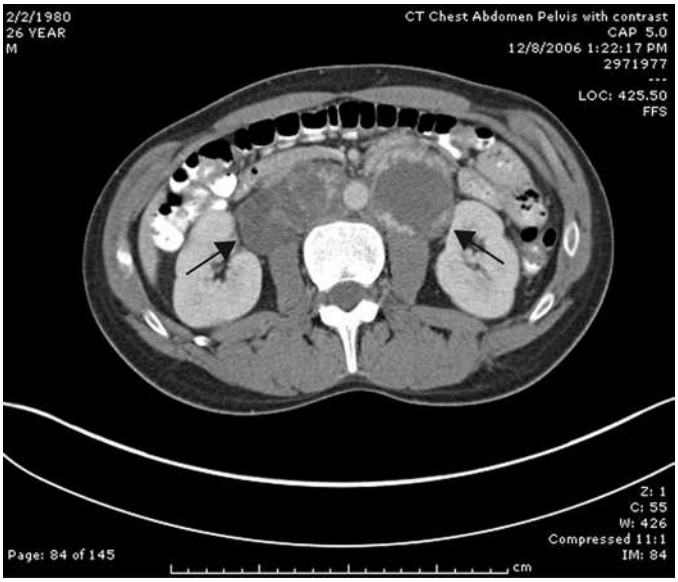


FIG 31.10 CT scan demonstrating large retroperitoneal lymph nodes.

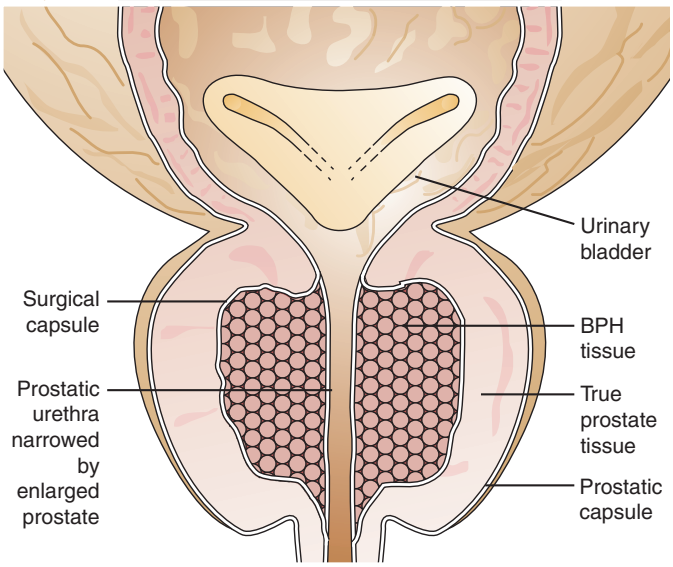


FIG 31.11 Gross appearance of hyperplastic prostatic tissue obstructing the prostatic urethra. *BPH,* benign prostatic hyperplasia.

KEY POINTS

- Cryptorchidism refers to a testis located in a position other than the scrotum. Often the testis has failed to descend completely and is located in the inguinal canal. Undescended testes are associated with infertility and an increased risk of testicular malignancy.
- A hydrocele is a collection of fluid in the testicle or spermatic cord. In the pediatric age group it is associated with a communication between the abdominal cavity and the scrotum (hernia). Hydroceles are benign and treated only if they become uncomfortable. A spermatocele is a cyst that contains sperm. Like hydroceles, they are benign and do not require treatment unless they cause discomfort.
- Testicular torsion refers to a twisting of the spermatic cord with subsequent testicular ischemia and infarction. Sudden onset of severe testicular pain is common. If the torsion is reduced within 12 hours, the testicle may be viable.
- Inflammation of the epididymis, called *epididymitis,* is most commonly associated with infectious agents. Manifestations include a swollen, tender, reddened scrotum with associated bladder infection and cloudy urine. Antibiotics are indicated. Aggressive infections of the scrotum may result in Fournier gangrene manifested by gangrenous necrosis of the scrotum.
- Although rare in the population, testicular cancer is the most common solid tumor in men ages 20 to 34 years. The great majority of testicular neoplasms originate in the germ cells. Most germ cell tumors can be effectively managed even after lymph node metastasis. Management includes surgical removal of the testis and spermatic cord, with irradiation and chemotherapy as indicated.

DISORDERS OF THE PROSTATE

Benign Prostatic Hyperplasia

Benign prostatic hyperplasia, also referred to as benign prostatic hypertrophy (BPH), is a very common disorder. An estimated 80% of men older than 60 years experience some degree of BPH. It is important to recognize that BPH and prostate cancer are not related entities, and no study has conclusively demonstrated that BPH predisposes to the development of prostate cancer.

Etiology. Although the exact cause of BPH is unknown, the occurrence of the disease with aging suggests a relationship to changes in the aging male endocrine system. The process involves hyperplasia of the glands surrounding the prostatic urethra (Fig. 31.11). As this tissue increases in size, it compresses the urethra and produces symptoms of bladder outlet obstruction.

Clinical manifestations. Symptoms of obstruction may be minimal at first but may eventually progress to complete obstruction and urinary retention. A decrease in the force of the urinary stream, hesitancy or difficulty in initiating a urinary stream, and interruption of the stream may occur. Because the bladder may fail to empty completely, infection associated with residual urine may occur. Fig. 31.12 illustrates possible complications of benign prostatic enlargement.

The diagnosis of BPH usually involves recognition of the characteristic symptoms. Rectal examination disclosing an enlarged prostate, urethral catheterization or bladder scanning to document a large postvoid urinary residual, and radiographic evidence of hypertrophy and obstruction on CT scan (Fig. 31.13) or abdominal ultrasound are some of the measures that may be used to make the diagnosis. Urodynamics are sophisticated measures of urinary function that can help confirm the diagnosis.

Treatment. The first line of treatment is with medication such as α-blockers or 5α-reductase inhibitors. The former treatment decreases the tension in the prostate by relaxing the muscle fibers in the gland. This reduces pressure in the bladder neck and urethra, allowing easier flow. The latter treatment actually decreases the size of the gland by blocking the conversion of testosterone to dihydrotestosterone. This can shrink the gland by up to one-third and thereby reverse the years of growth that have caused impingement of the urethra.

The next forms of treatment are the minimally invasive treatments that are usually performed in the physician's office. These include microwave therapy, needle ablation, and some other forms of treatment to decrease the tissue squeezing the urethra closed. They are usually well tolerated but may not cause total relief of the obstruction, and the long-term results are variable.

The last form of treatment is surgery. The classic operation for BPH is called a *transurethral resection of the prostate (TURP).* This procedure uses a resectoscope that is passed through the penis and into the prostate. An electric wire is then used to resect chips of tissue from the interior of the gland to form an open channel for urination.

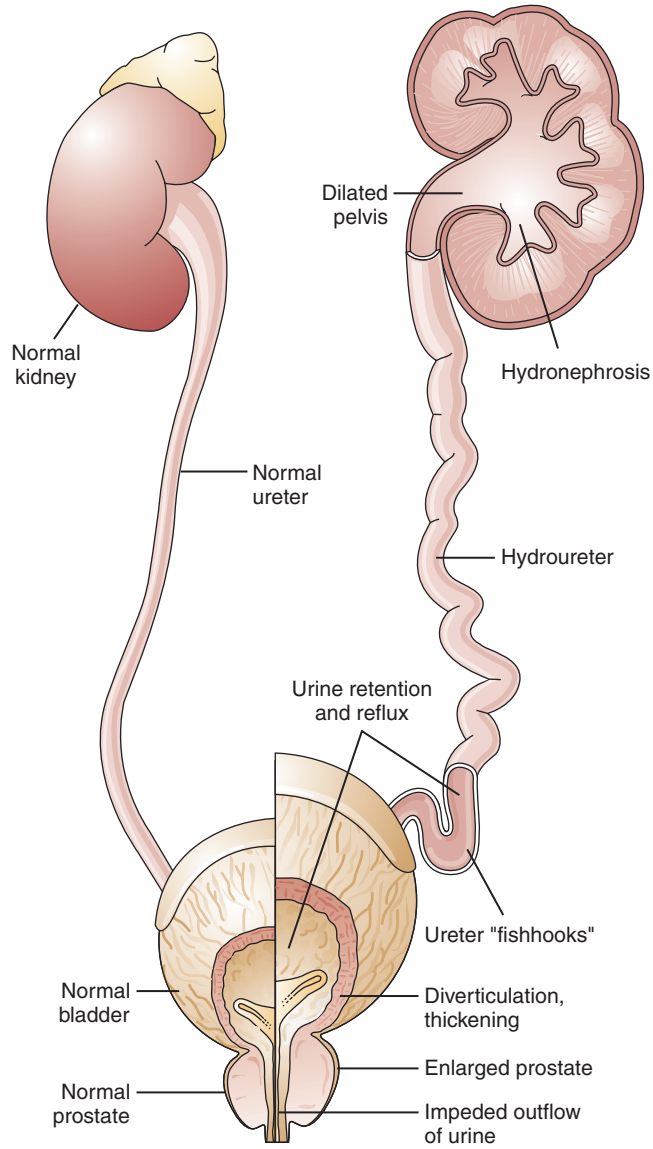


FIG 31.12 Sites for potential complications caused by benign prostatic enlargement *(right)* are compared with a normal kidney, ureter, bladder, and prostate *(left)*.

A newer alternative to the TURP is laser therapy. In this approach, a scope is passed via the urethra into the prostate, and the tissue is vaporized with a laser fiber to create an open channel.

In some patients with extremely large glands or other bladder pathology, an incision is made in the lower abdomen and an open simple prostatectomy can be performed to remove the adenoma (inner portion of the gland). This creates a large cavity and channel for the urine to pass through.

All of the treatments have various side effects and risks. Therefore each patient should have his treatment tailored to his specific situation.

Prostatitis

Prostatitis, or inflammation of the prostate, has several causes and encompasses several syndromes. A common classification of prostatitis proposed by the National Institutes of Health has three categories of prostatitis with further subcategories to define the disease.

The causative organism in bacterial prostatitis is usually *E. coli,* with species of *Proteus, Klebsiella, Enterobacter, Pseudomonas, Serratia, Staphylococcus,* and *Enterococcus* occurring less commonly. Possible routes of infection include ascending infection up the urethra, reflux of infected urine into the prostatic ducts, hematogenous infection, and invasion of rectal bacteria by direct extension or lymphogenous spread (Fig. 31.14). Many cases of prostatitis result from periurethral infection associated with an indwelling urethral catheter.

Clinical manifestations, diagnosis, and treatment. Acute bacterial prostatitis is characterized by the onset of fever; chills; low back pain; and the voiding symptoms of frequency, urgency, and dysuria. Rectal examination usually reveals a tender, swollen prostate, and subsequent urinalysis may show the presence of WBCs and bacteria.

The diagnosis of bacterial prostatitis is usually suggested by the initial symptoms and signs. Microscopic inspection of the urine and expressed prostatic secretions may reveal WBCs and bacteria. A urine culture with sensitivity testing for the offending organism is recommended to direct therapy with an appropriate antibiotic. In the event of high fever and an elevated WBC count, intravenous antibiotics are recommended.

Chronic bacterial prostatitis may be associated with variable symptoms. Although some men with chronic bacterial prostatitis may report a history of acute bacterial prostatitis, many have no history of this problem. Most men complain of voiding symptoms with pain localized to various areas, including the perineum, back, suprapubic area, and, occasionally, the testis. High-grade fever and chills are uncommon with this entity, as opposed to acute bacterial prostatitis.

In chronic bacterial prostatitis, pathogenic organisms may persist in prostatic tissues unaltered by the administration of several antibiotics. Because most antibiotics accumulate poorly in prostatic secretions, discontinuation of antibiotic use often results in reinfection and recurrence of symptoms. It is this occurrence of relapsing infections, often caused by the same organism, that is typical of chronic bacterial prostatitis. Several antibiotic agents, such as trimethoprim-sulfamethoxazole (Septra, Bactrim) or ciprofloxacin, when used for a prolonged period (4 to 6 weeks) have a better cure rate because of their capability to penetrate prostatic tissue.

Prostatitis may also occur secondary to nonbacterial inflammation. In fact, nonbacterial prostatitis probably accounts for the majority of cases of prostatitis. The symptoms of this entity are variable but usually include irritative voiding; symptoms of urgency, frequency, and nocturia; and occasional perineal and suprapubic pain. Although these symptoms are similar to those of bacterial prostatitis, patients have no history of positive urine cultures or urinary tract infections. Treatment may include a course of antibiotics, oral antiinflammatory agents (e.g., ibuprofen), α-blockers, prostatic massage, and, occasionally, sitz baths. Symptoms are often intermittent, and patients should be reassured that the disease is not contagious and does not predispose to the development of cancer or other serious disease.

The final classification of prostatitis, chronic pelvic pain syndrome, is typified by symptoms of prostatitis but no history of urinary tract infection and no evidence of inflammation in prostatic secretions. The cause of this entity is uncertain and may involve spasm of the pelvic floor musculature. Treatment may involve the use of α-adrenergic receptor–blocking agents, antibiotics, or multiple other therapies.

Prostate Cancer

Other than skin cancers, prostate cancer is now recognized as the most prevalent form of cancer in men. About 160,000 cases are diagnosed annually in the United States, with approximately 27,000 deaths annually attributed to the disease. Prostate cancer ranks as the second-leading cause of cancer death among men. Cancer of the prostate rarely occurs

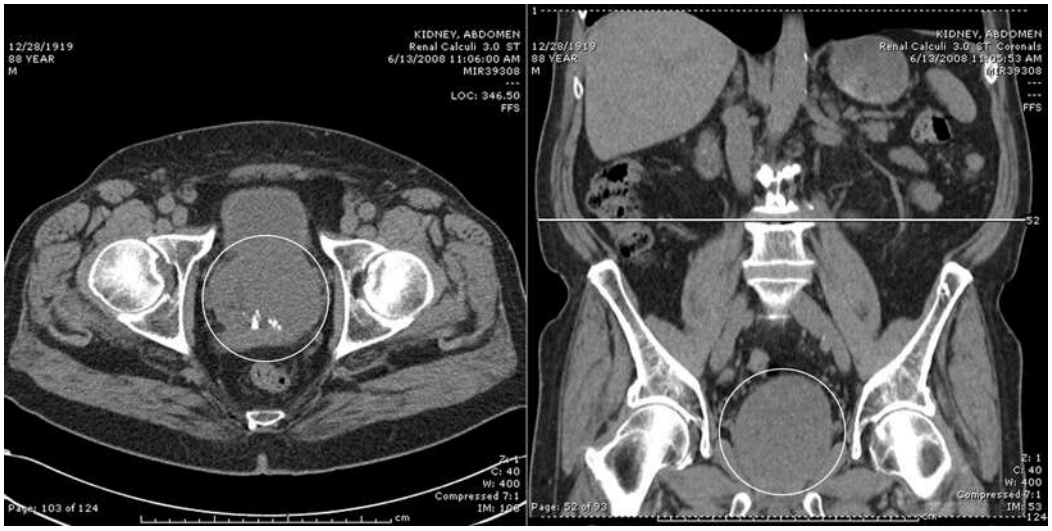


FIG 31.13 CT scan demonstrating BPH *(circled areas)*.

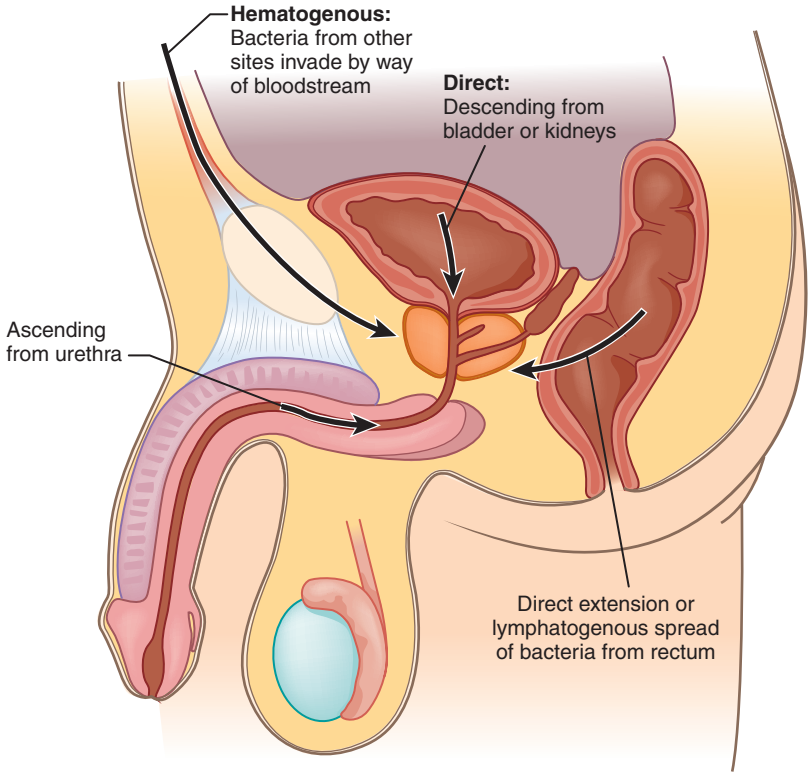


FIG 31.14 Postulated pathways of infection to the prostate gland. (From Black JM, Matassarin-Jacobs E: *Medical-surgical nursing: clinical management for continuity of care,* ed 6, Philadelphia, 2001, Saunders, p 963.)

in men younger than 50 years, and its incidence increases with age. The majority (95%) of prostate cancers are adenocarcinomas with abnormal proliferation of prostatic glandular structures.

Etiology. The precise cause of prostate cancer is undetermined, although genetic, hormonal, dietary, and viral factors have all been suggested. Varying degrees of aggressiveness of prostate cancer have been recognized, with different tumors expressing different malignant potential and ultimately carrying a different prognosis. Classification of prostate cancers into different groups considers the structure and internal architecture of tumor cells and their pattern of proliferation. For example, cells of the more aggressive or poorly differentiated prostate cancers have more indistinct cell borders, larger nuclei, and loss of acinar (gland) formation.

Diagnosis. The diagnosis of prostate cancer may involve several clinical scenarios. The disease may be diagnosed after microscopic inspection of prostate tissue removed for the management of presumed BPH. Prostate cancer may also be detected on rectal examination in patients with or without voiding symptoms. Occasionally, patients have urinary retention or even azotemia and renal failure secondary to obstructive nephropathy. Much interest has focused on the search for effective measures to detect prostate cancer in its early and most easily manageable stages. Two techniques, a blood test for serum

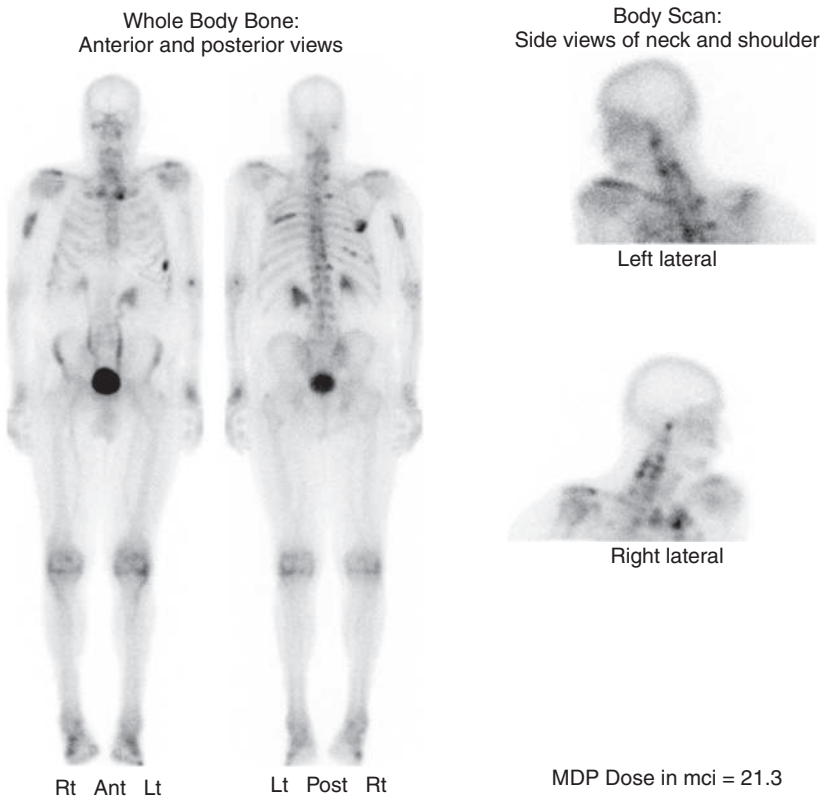


FIG 31.15 Bone scan demonstrating skeletal metastases.

prostate-specific antigen (PSA) and transrectal ultrasonography, have shown efficacy in the early detection of prostate cancer.

The diagnosis of adenocarcinoma of prostate cancer is usually made by checking the PSA level. This simple blood test has been a major factor in the early detection of adenocarcinoma of the prostate. Although it is not very accurate, it is helpful in identifying men who are at risk of having cancer. Since the adoption of its widespread use, most patients have been diagnosed before becoming symptomatic. Many patients will have voiding symptoms because they also have benign enlargement of the gland, which is often also present in men with prostate cancer. If a patient has symptoms from his cancer, it is often too late for cure because the symptoms usually indicate extensive disease. These symptoms include hematuria, weight loss, malaise, anorexia, and back pain.

Most patients will be diagnosed on the basis of an abnormal PSA blood test result, which usually leads to a transrectal ultrasound and biopsy of the prostate.

Because most patients are now diagnosed at an early stage, there is little need for further staging tests. However, if there is concern for more extensive cancer, then they can be evaluated with a bone scan (Fig. 31.15) or CT scan of the abdomen and pelvis.

Treatment. Management of prostate cancer depends on several factors, including the stage of the tumor, as well as the age and health of the patient. Debate exists over treatment for men with localized disease, with options ranging from active surveillance to aggressive therapy. Active surveillance may be an option for men with early-stage prostate cancer. Patients with more worrisome cancer may be candidates for a more aggressive approach, such as surgery to remove the prostate and surrounding tissue (radical prostatectomy), radiation therapy, or cryotherapy (freezing the gland). It is important to note that approximately 70% to 90% of men who undergo radical prostatectomy will experience erectile dysfunction. Urinary incontinence may also occur. Given the effects of surgical intervention, it is important that patients have access to preoperative and postoperative counseling about issues arising from their diagnosis and the effect of various treatments and potential complications.

In addition to radical prostatectomy or radiation therapy, in some cases lymph nodes in the pelvis are removed (pelvic lymph node dissection). Advanced disease may respond to hormonal manipulation. Orchiectomy, oral administration of estrogens, or IM injection of LH–releasing hormone agonist and antagonist may reduce the patient's serum testosterone level. Many prostate cancers are androgen sensitive and may be temporarily controlled with androgen ablation. In more advanced cases that are no longer hormonally responsive, palliative measures such as spot radiation treatment of painful areas of bone metastasis and analgesics may be required.

In the last several years there have been advances in chemotherapy and other treatments such as abiraterone and enzalutamide for advanced and castrate-resistant prostate cancer. Another treatment is sipuleucel-T. This therapy is based on autologous cellular immunotherapy to induce the patient's immune system to attack the prostate cancer cells. A great deal of research is being conducted to identify vaccines and other therapies to prevent and cure this highly prevalent cancer.

GERIATRIC CONSIDERATIONS

Prostate Cancer

Cancer of the prostate rarely occurs in men younger than 50 years, and its incidence increases with age. Symptoms vary depending on the stage of the disease. Men who have early-stage prostate cancer may be monitored closely without any treatment. Men with more advanced-stage disease are usually treated more aggressively.

KEY POINTS

- Symptoms of benign prostatic hyperplasia include diminished force of the urinary stream, hesitancy, and poor bladder emptying. Transurethral resection of the obstructing prostatic tissue is the usual treatment.
- Inflammation of the prostate, or prostatitis, is characterized by low back pain, urinary frequency, urgency, and dysuria. Fever and chills may also be present with acute bacterial prostatitis. *E. coli* is the most commonly associated organism. Prostatitis may also occur in the absence of infection.
- Prostate cancer is usually detected as a lump or enlargement of the prostate gland. As with other cancers, early, accurate diagnosis is important for effective therapy. Surgical resection, radiation therapy, and hormone therapy (to reduce androgen levels) may be used. The choice of treatment depends on the grade and stage of the disease and the individual's age, general health, and life expectancy.

SUMMARY

Disorders of the penis and male urethra may be grouped into congenital and acquired anomalies, infections, and neoplasms. Common congenital anomalies include urethral valves and hypospadias. Common acquired disorders involve phimosis, urethral strictures, and ED. Sexually transmitted diseases are some of the most common infections involving the penis and urethra; they include gonococcal urethritis, nongonococcal urethritis, syphilis, genital herpes, and genital warts. Neoplasms of the penis and urethra are relatively rare.

Congenital disorders of the scrotum and testes include cryptorchidism. This condition is one of the most common problems seen by pediatric urologists. Testicular torsion and Fournier gangrene are two of the more immediate urologic emergencies. Finally, neoplasms of the testes, although rare, may afflict younger men in the prime of life.

Disorders of the prostate account for a majority of the visits to a practicing urologist. Briefly, these disorders can be divided into problems of BPH, prostatitis, and prostatic cancer. Prostate cancer is the most frequently diagnosed cancer in men, with more than 160,000 cases diagnosed and approximately 27,000 deaths yearly.

RESOURCES

American Urological Association: Contemporary management of advanced prostate cancer. *AUA Update Series* 34:268, 2015.

Armstrong BK, et al: PSA testing for men at average risk of prostate cancer. *Public Health Res Pract* 27(3):2017. doi:10.17061/phrp2731721. pii: 2731721.

Bansal UK, et al: The efficacy of tadalafil daily versus on demand in the treatment of erectile dysfunction: a systematic review and meta-analysis. *Urology* 2017 Sep 4. doi:10.1016/j.urology.2017.08.031. pii: S0090-4295(17)30905-6.

Caldwell BT, et al: Current management for pediatric urologic oncology. *Adv Pediatr* 64(1):191–223, 2017. doi:10.1016/j.yapd.2017.04.001.

Cohen T, et al: Management of bilateral undescended bilobed testes and review of the literature. *Urology* 2017. doi:10.1016/j.urology.2017.08.026. [Epub ahead of print]; Aug 28. pii: S0090-4295(17)30888-9.

Falcone M, et al: The management of the acute ischemic priapism: a state of the art review. *Actas Urol Esp* 2017 May 17. doi:10.1016/j.acuro.2017.02.004. [Epub ahead of print]; pii: S0210-4806(17)30045-1.

Gaither TW, et al: Evaluation and management of erectile dysfunction in the hypertensive patient. *J Sex Med* 14(9):1071–1078, 2017. doi:10.1016/j.jsxm.2017.07.010.

Hall SJ, et al: Integrated safety data from a randomized, double-blind, controlled trial of autologous cellular immunotherapy with sipuleucel-T patients with prostate cancer. *J Urol* 186(3):877–881, 2011.

Hatipoğlu N, Kurtoğlu S: Micropenis: etiology, diagnosis and treatment approaches. *J Clin Res Pediatr Endocrinol* 5(4):217–223, 2013.

Hudnall M, et al: Advances in the understanding of priapism. *Transl Androl Urol* 6(2):199–206, 2017. doi:10.21037/tau.2017.01.18.

Jarow J, et al: *The optimal evaluation of the infertile male: AUA best practice statement*. Linthicum, MD: American Urological Association Education and Research, Inc; 2010. http://www.auanet.org/documents/education/clinical-guidance/Male-Infertility-d.pdf. (Accessed 22 August 2017).

Khan FU, et al: Comprehensive overview of prostatitis. *Biomed Pharmacother* 94:1064–1076, 2017. doi:10.1016/j.biopha.2017.08.016. [Epub ahead of print].

Lee CL, Kuo HC: Pathophysiology of benign prostate enlargement and lower urinary tract symptoms: current concepts. *Ci Ji Yi Xue Za Zhi* 29(2):79–83, 2017. doi:10.4103/tcmj.tcmj_20_17.

Locke JA, et al: Treatment of varicocele in children and adolescents: a systematic review and meta-analysis of randomized controlled trials. *J Pediatr Urol* 2017 Aug 9. doi:10.1016/j.jpurol.2017.07.008. [Epub ahead of print]; pii: S1477-5131(17)30299-1.

Nehra A, et al: Peyronie's disease: AUA guideline. *J Urol* 194(3):745–753, 2015. doi:10.1016/j.juro.2015.05.098.

Patel JP, et al: Evaluation and management of erectile dysfunction in the hypertensive patient. *Curr Cardiol Rep* 19(9):89, 2017 Aug 24. doi:10.1007/s11886-017-0889-z.

Paulis G, et al: Recent pathophysiological aspects of peyronie's disease: role of free radicals, rationale, and therapeutic implications for antioxidant treatment-literature review. *Adv Urol* 2017:4653512, 2017. doi:10.1155/2017/4653512. [Epub 2017 Jul 4].

Rancati T, et al: Understanding urinary toxicity after radiotherapy for prostate cancer: first steps forward. *Tumori* 2017 Sep 7:0. doi:10.5301/tj.5000681. [Epub ahead of print].

Serefoglu EC, et al: An evidence-based unified definition of lifelong and acquired premature ejaculation: report of the Second International Society for Sexual Medicine Ad Hoc Committee for the Definition of Premature Ejaculation. *J Sex Med* 2(2):41–59, 2014. doi:10.1002/sm2.27.

Wein AJ, et al, editors: *Campbell-Walsh urology*, 11th ed, Philadelphia, 2016, Elsevier.

32

Female Genital and Reproductive Function

Rosemary A. Jadack

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- What are the major structures of the internal and the external female reproductive tract?
- What are the major hormonal events of the female reproductive cycle?
- Which hormones are involved in breast development during pregnancy and lactation, and what are their specific functions?
- What physiologic changes are associated with pregnancy?
- What gestational events occur in the fetus during each of the three trimesters of pregnancy?
- What hormonal changes lead to menopause?
- What physiologic changes and complications may result from menopausal hormone deficiencies?

CHAPTER OUTLINE

Reproductive Structures, 656
Organization of the Female Reproductive Organs, 656
Ovaries, 657
Oviducts, 658
Uterus, 658
Vagina, 658
External Genitalia, 658
Menstrual Cycle, 659
Breast, 662
Structure of the Breast, 662
Breast Development, 662
Lactation, 663
Pregnancy, 663
Early Human Development, 663
Implantation, 663
Fetal Membranes and Placenta, 663
Development of the Human Embryo and Fetus, 664
First Month, 664
Second Month, 665
Third Month, 665
Second Trimester, 665
Third Trimester, 665
Parturition, 665
Hormonal Changes, 665
Mechanical Changes, 667
Response of the Mother's Body to Pregnancy, 667
Metabolism During Pregnancy, 667
Changes in the Female Reproductive Organs, 667
Changes in the Circulatory System, 667
Changes in the Respiratory System, 668
Changes in the Urinary System, 668
Weight Gain and Nutrition During Pregnancy, 668
Menopause, 668

The female reproductive system is complex both in structure and in function. From birth to senescence, the organs of the female reproductive system function in concert with each other, with the brain, and with other endocrine organs. This integrated functioning constitutes some of the most intricate and elegant processes of the human body. This chapter presents an overview of these functions, beginning with the development of the female reproductive tract.

The major processes related to the reproductive tract throughout life, including the menstrual cycle, pregnancy, lactation, and menopause, are then described with an emphasis on recent research findings. Health care providers must also consider the developmental, cognitive, functional, social, and financial aspects of women's reproductive lives. Because the functioning of the female reproductive system has an enormous impact on the life of the individual woman, increased importance has been placed on the active involvement of women in understanding their own health care needs. Health care professionals are encouraged to include women as collaborators in decisions about their reproductive health.

REPRODUCTIVE STRUCTURES

Organization of the Female Reproductive Organs

The internal organs of the female reproductive system include the ovaries, oviducts (fallopian tubes), uterus, cervix, and vagina (Fig. 32.1). These

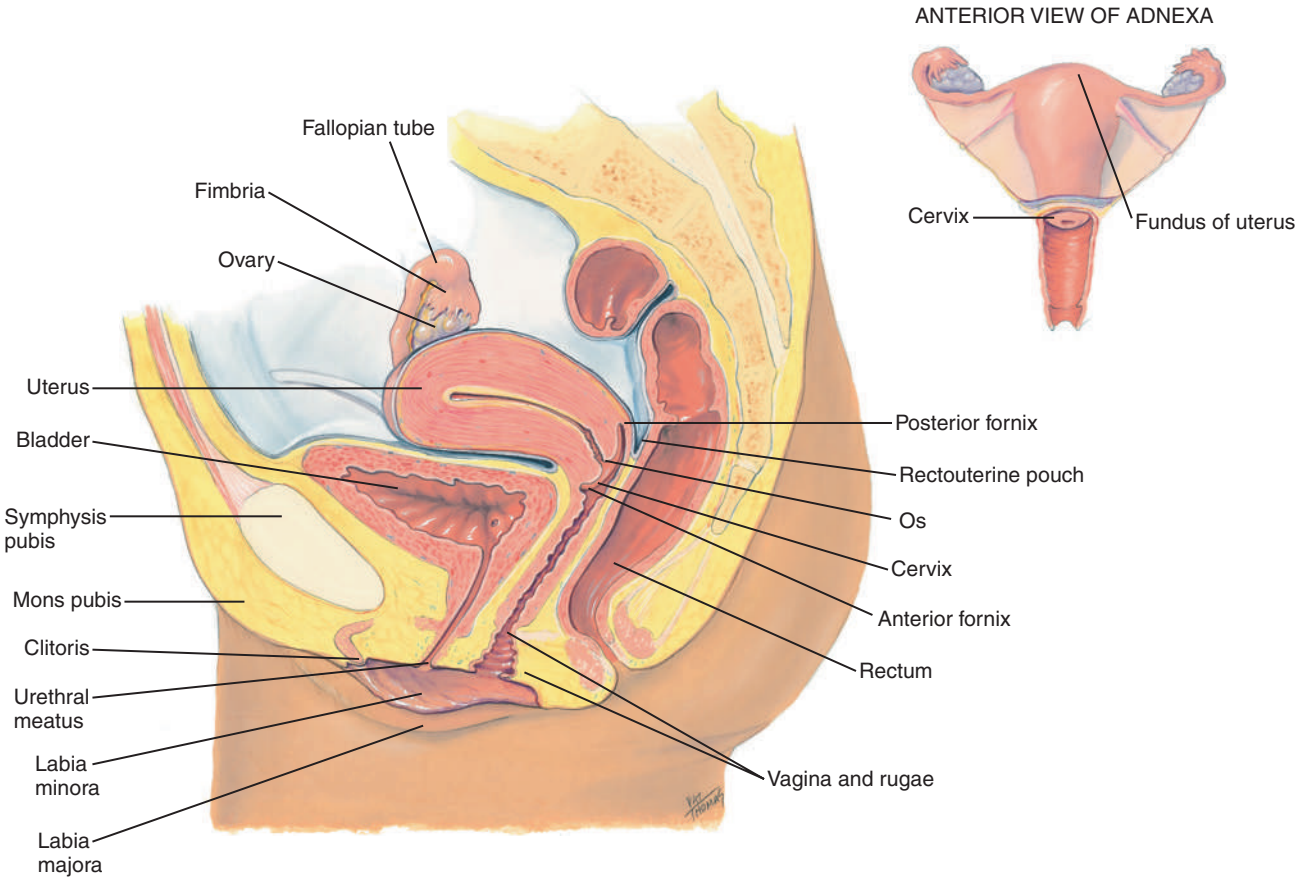


FIG 32.1 Cross-sectional view of the female genitourinary system. (From Jarvis C: *Physical examination and health assessment,* ed 7, St Louis, 2016, Elsevier, p 738.)

organs are situated in the pelvic cavity and are supported and anchored in place by a series of ligaments (Fig. 32.2).

Ovaries

The two ovaries, which are the female gonads, are located close to the lateral walls of the pelvic cavity. When the ovary is in its normal position, its long axis is nearly vertical with respect to the horizontal axis of the body. The size of the ovary varies with age and with the stage of the menstrual cycle. It is somewhat larger before than after pregnancy and further reduces in size with the aging process.

The ovary is covered with a single layer of epithelium. Underneath the epithelium is a layer of dense fibrous connective tissue called the *tunica albuginea.* The tunica albuginea constitutes the outer portion of the cortex of the ovary. The remainder of the cortex consists of connective tissue called the *stroma,* which contains ova in various stages of maturation. The innermost part of the ovary, the medulla, consists of loose connective tissue that is richly supplied with blood and lymph vessels and nerve fibers.

Before birth, hundreds of thousands of oogonia (cells that develop into ova) are present in the ovaries. Thus the entire lifetime supply of ova is established during embryonic development; no new oogonia arise after birth. Each oogonium is surrounded by a cluster of granulosa cells. The oogonium and its granulosa cells constitute a follicle. During prenatal development, the oogonia increase in size and become primary oocytes. By the time of full gestational development, the primary oocytes are in the prophase of the first meiotic division. Ovarian follicular development is shown in Fig. 32.3. During childhood and into adult life, the oocytes enter a nonactive phase. After puberty, a few of the oocytes develop in follicles each month in response to follicle-stimulating

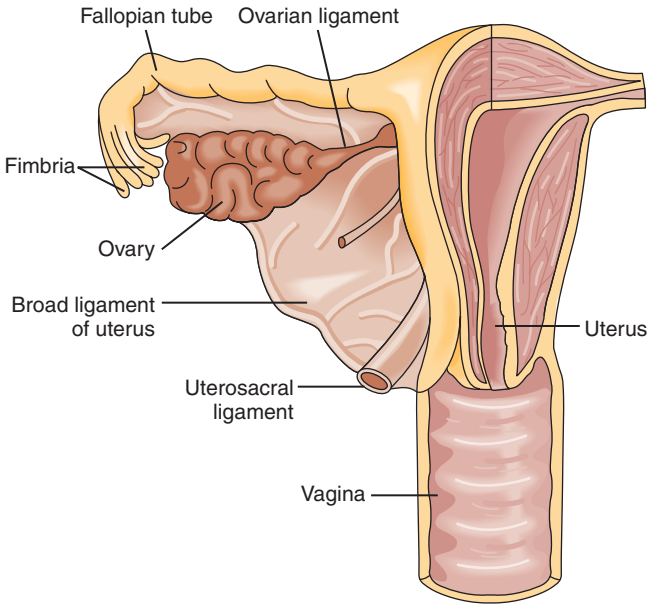


FIG 32.2 View of the female pelvis showing the ovarian and uterine ligaments.

hormone (FSH) secreted by the anterior pituitary gland. The vast majority of follicles and their oocytes die by atresia. Generally, each month, only one mature follicle will develop to eject an oocyte through the wall of the ovary in the process of ovulation, which is described in more detail in the Menstrual Cycle section.

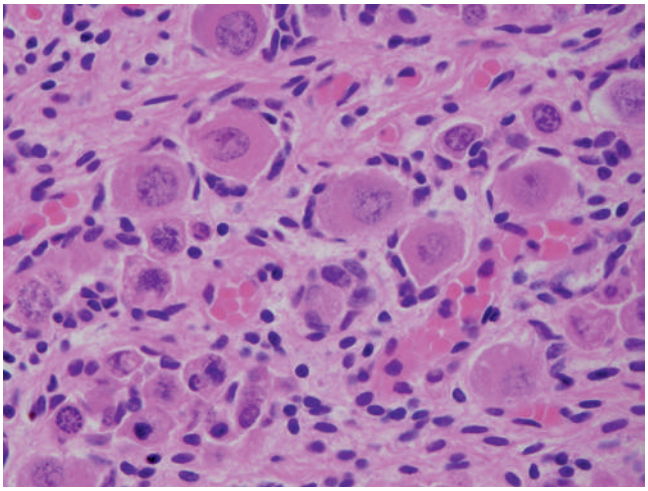


FIG 32.3 Ovary at 36 weeks' gestation showing primordial follicles. (From Mutter GL, Prat J. *Pathology of the female reproductive track,* ed 3, Philadelphia, 2014, Churchill Livingstone, Elsevier.)

Oviducts

The two oviducts, also called the *fallopian* or *uterine tubes,* are each about 10 cm long and are located in the upper margin of the broad ligament. Each oviduct runs laterally from the uterus to the uterine end of the ovary. The free end of the oviduct adjacent to the ovary is called the *infundibulum.* It is shaped like a funnel with long, finger-like projections termed *fimbriae.* The ampulla, the longest part of the oviduct, has an inner lining consisting of ciliated mucous membrane arranged in longitudinal folds. Beneath this ciliated lining is a double layer of smooth muscle with a thick outer layer of peritoneal serosa. The oviduct has an active role in propelling the ovum toward the uterus; the current created by the beating cilia and the peristaltic contractions of the muscular wall are powerful forces that move ova along the oviduct. Once inside the oviduct, the ovum is moved through the ampulla to the isthmus (the short, narrow portion near the uterus) and finally through the intramural passageway to the uterus. Fertilization of the ovum occurs in the upper third of the oviduct, and the zygote (fertilized ovum) begins developing as it moves through the oviduct. If no fertilization occurs, the ovum undergoes degeneration in the oviduct.

Uterus

The uterus varies in size, shape, location, and structure during various phases of a woman's life and reproductive status. In the nonpregnant state, the uterus is about 8 cm long, 4 cm wide in its upper part, and 2 cm thick. The rounded part of the uterus, which lies above and in front of the openings of the oviducts, is called the *fundus;* the main portion of the uterus is the *corpus,* or body. The lower, narrow portion of the uterus is the *cervix,* which extends downward to the opening within the vagina. The cervix contains a narrow canal that joins the uterine cavity at the internal os and opens into the vagina at the external os.

The wall of the body and fundus of the uterus consists of three layers: endometrium, myometrium, and serosa (Fig. 32.4). The outermost layer of the uterus, the serosa, consists of a single layer of mesothelial cells supported by a thin layer of loose connective tissue. The middle layer, the myometrium, consists of three layers of smooth muscle with the muscle fibers arranged in a different direction in each layer. The innermost lining of the uterus, the endometrium, consists of two layers: a thin deep layer called the *basilar layer* and a thick superficial layer referred to as the *functional layer.* During a woman's reproductive years, the endometrium displays a constant cyclic activity of alternate proliferation and sloughing of the functional layer in response to estrogen and progesterone secretion. These changes will be discussed in more detail in the Menstrual Cycle section.

Vagina

The vagina is the sexual organ that enfolds the penis during sexual intercourse, serves as an exit for discarded endometrium, and forms the lower end of the birth canal. It is located anterior to the rectum and posterior to the urethra and urinary bladder. The vagina surrounds the cervix at one end and opens to the vestibule at its other end. The vagina is a highly elastic muscle that is capable of considerable distention. Two longitudinal ridges run along the anterior and posterior walls, with numerous transverse folds called *rugae.* The vagina is lined by a mucous membrane of stratified squamous epithelium overlying a layer of connective tissue. The vaginal wall is subject to thinning with aging; this and other age-related changes in the female sexual organs are described in the Menopause section.

External Genitalia

The external female genital structures include the mons pubis, labia majora, labia minora, clitoris, and vestibule of the vagina (Fig. 32.5). The stages of development of the female external genitalia are depicted in Fig. 32.6. The mons pubis is a rounded elevation in front of the pubis symphysis. It consists primarily of an accumulation of fat. After puberty, the skin over it is covered by coarse hair. The labia majora, which are homologous (i.e., corresponding in structure) with the scrotum of the male, are folds of skin that run downward and backward from the mons pubis to the area behind the vaginal opening. After puberty, the labia majora become pigmented and covered with hair. The labia minora are two small folds of skin located between the labia majora on either side of the vaginal opening. The vestibule of the vagina is the cleft between the labia minora and contains the openings of the vagina, the urethra, and the ducts of the greater vestibular glands (also called *Bartholin glands*). These glands, along with the lesser vestibular or Skene glands, secrete mucus to provide lubrication during sexual intercourse.

The clitoris is a body of erectile tissue that projects from the anterior end of the vulva at the anterior junction of the labia minora. It is about 2 cm long and 0.5 cm in diameter and is covered by a fold of tissue called the *prepuce,* which is formed by the merging of labial tissue. The glans of the clitoris is the rounded elevation on the free end of the body and is highly sensitive to stimulation. During sexual arousal, the erectile tissue of the clitoris becomes engorged with blood.

KEY POINTS

- Organs of the female reproductive tract include the ovaries, oviducts, uterus, cervix, and vagina. Ovaries contain a lifetime supply of ova at birth. After puberty, a few of the ovarian follicles develop about every 28 days in response to secretion of follicle-stimulating hormone (FSH).
- The oviducts (fallopian tubes) actively propel the ovum toward the uterus by ciliary action and peristaltic contractions. The uterine lining undergoes a cyclic process of proliferation and then sloughing in response to estrogen and progesterone.
- External genitalia in the female include the mons pubis, labia majora, labia minora, clitoris, and vestibule of the vagina. The urinary meatus, vaginal opening, and vestibular gland ducts are located in the vaginal vestibule.

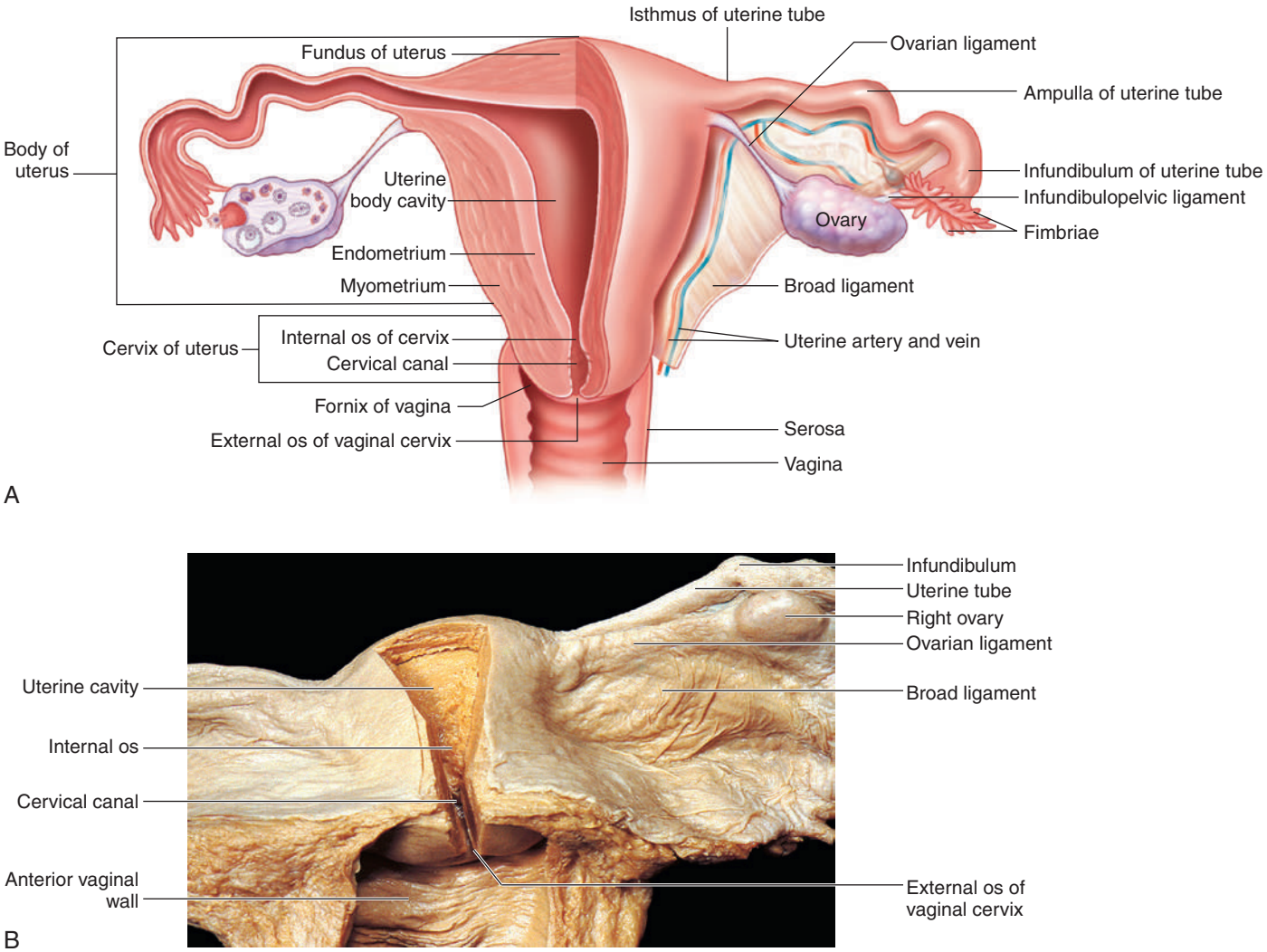


FIG 32.4 Internal female reproductive organs (posterior view). **A,** Diagram shows the left side of uterus and upper portion of the vagina and the left uterine tube and ovary in a frontal section. The broad ligament has been removed from the posterior surface of the uterus and adjacent structures. **B,** Cadaver dissection showing uterine cavity and cervical canal, exposed by removal of parts of their posterior walls. Note that the uterine wall consists of an epithelial lining from which uterine glands extend through the full thickness of the mucosa. Beneath the endometrium, a portion of myometrium is shown. (A, From Patton KT, Thibodeau GA: *Human body in health and disease,* ed 7, St Louis, 2018, Elsevier, p 630. B, From Gosling J et al: *Human anatomy,* ed 4, Philadelphia, 2005, Mosby.)

MENSTRUAL CYCLE

From menarche onward, the normal reproductive years of the female are characterized by rhythmic changes in hormonal secretion and corresponding changes in the sexual organs, which are called the *target organs* of the female hormones. This rhythmic pattern is called the menstrual cycle (Fig. 32.7). Two significant results of the menstrual cycle are stimulation of the production of an ovum and preparation of the uterine endometrium for the implantation of a fertilized ovum at the appropriate phase of the cycle.

Although considerable variation can be found in human females, an average menstrual cycle is 28 days long, with cycles as short as 20 days or as long as 45 days occurring in normal women. The first day of menstruation is considered the first day of the menstrual cycle. Ovulation occurs approximately 14 days before the next cycle begins; thus in a 28-day cycle, ovulation occurs on about day 14 of the cycle.

The release of hormones and the accompanying response of the female sexual target organs are depicted in Fig. 32.7. The principal female reproductive hormones are summarized in Table 32.1. As shown in Fig. 32.7, the events of the menstrual cycle require precise synchronization between the activities of the pituitary gland, ovary, and uterus. Beginning at the first day of the menstrual cycle, or the first day of menstruation, these events can be summarized as follows. The thickened functional layer of the endometrium of the uterus is gradually sloughed off, and about 35 mL of blood is lost. During this phase of the menstrual cycle, FSH is released by the pituitary gland and stimulates a group of follicles to develop in the ovary.

In the preovulatory phase, also called the *proliferative phase,* theca and granulosa cells in the developing follicles in the ovary secrete estrogen, which stimulates growth of the uterine endometrium once again. At about the midpoint of the cycle, an increase in estrogen secretion from the follicles occurs. This increase in estrogen level is thought to render the anterior pituitary more responsive to luteinizing hormone(LH)–releasing

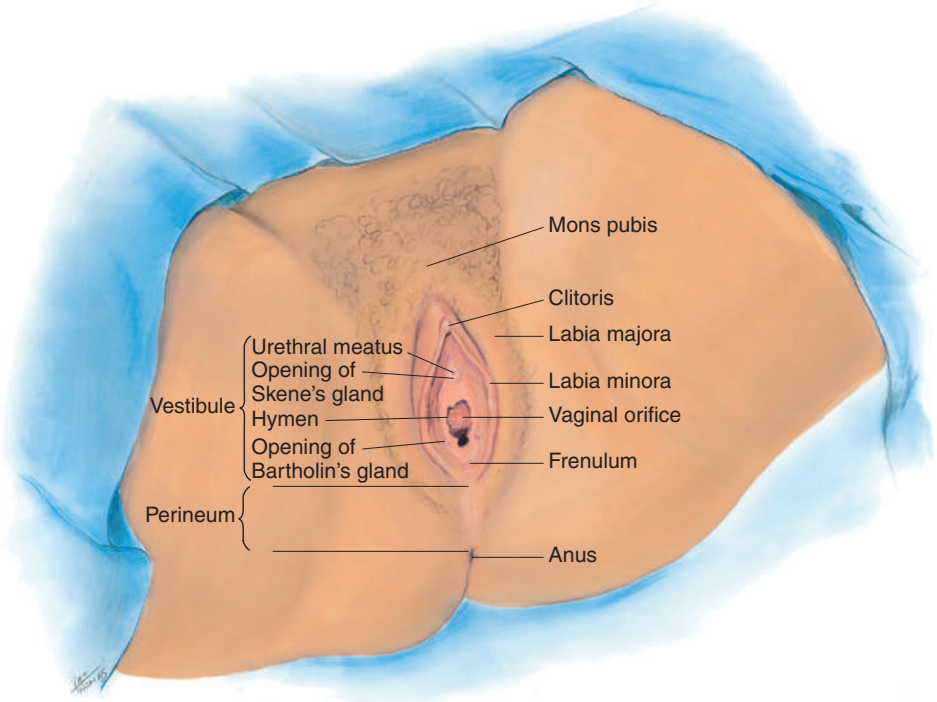


FIG 32.5 External female genitalia. (From Jarvis C: *Physical examination and health assessment,* ed 7, St Louis, 2016, Elsevier, p 737.)

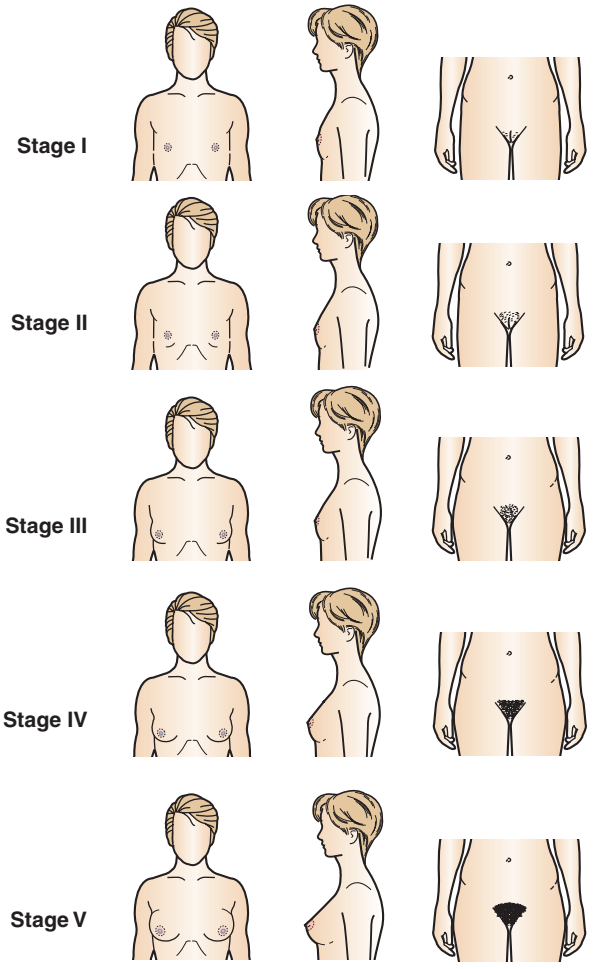


FIG 32.6 Five Tanner stages of sexual maturity ratings (SMRs) in girls.

TABLE 32.1 Principal Female Reproductive Hormones

Hormone	Target Organs	Significant Actions
Estrogen	Multiple sites throughout body, including reproductive structures, bone, fat, and muscle tissues	Development of reproductive organs during puberty Development of secondary sex characteristics, including breast maturation, widening of pelvis, and distribution of fat and muscle tissues in a distinctively female pattern Cyclic preparation of endometrium for implantation of an ovum
Progesterone	Primarily uterus and breasts	Cyclic preparation and maintenance of endometrium for implantation of an ovum Stimulation of development of breast lobes and alveoli
Follicle-stimulating hormone	Ovary	Stimulates ovarian follicle development; with luteinizing hormone, stimulates secretion of estrogen and ovulation
Luteinizing hormone	Ovary	Stimulates final development of ovarian follicle, process of ovulation, and development of corpus luteum

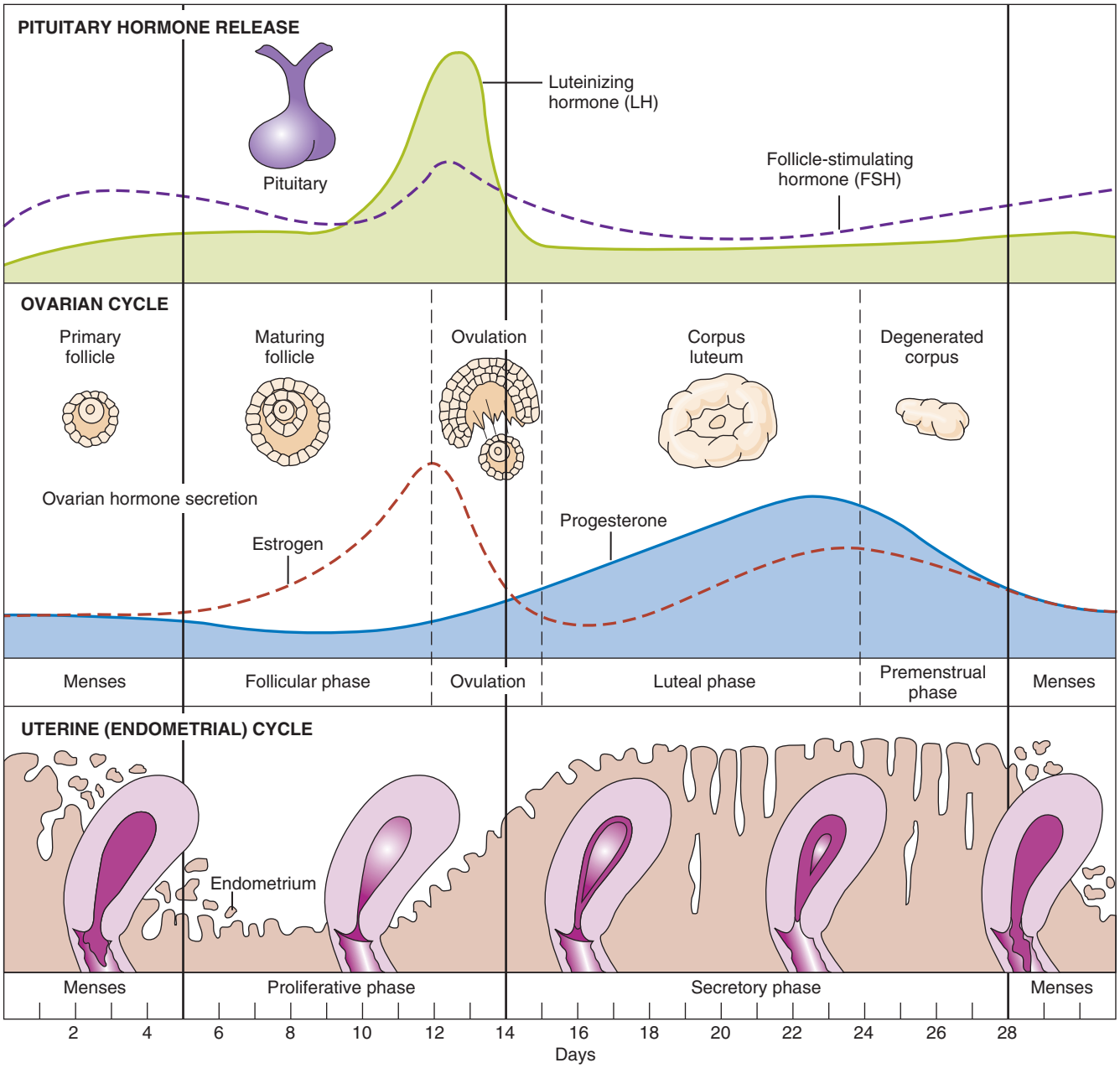


FIG 32.7 Menstrual cycle. The events that take place within the pituitary, ovary, and uterus are precisely synchronized. When fertilization does not occur, the cycle repeats itself about every 28 days.

hormone secreted by the hypothalamus. The anterior pituitary gland then produces a burst of LH. The FSH level also increases about twofold at the same time, and these two hormones act synergistically to cause the extremely rapid swelling of the follicle that culminates in ovulation. During the process of ovulation, the secondary oocyte is ejected through the wall of the ovary into the peritoneal cavity. The free end of the oviduct is strategically located so that the ovum enters its fimbriated end almost immediately.

After ovulation, the postovulatory phase (also called the *luteal phase*) begins. During the luteal phase the site of the ruptured follicle becomes a corpus luteum (Latin for "yellow body"), which secretes estrogen and progesterone. These hormones stimulate continued thickening of the uterine endometrium. The cells of the corpus luteum become greatly enlarged and develop lipid, or fatty, areas that give the cells a distinctive yellow color. In a normal cycle, the corpus luteum grows to approximately 1.5 cm, with maximal development attained about 7 to 8 days after ovulation. If pregnancy does not occur, the corpus luteum begins to degenerate, and progesterone and estrogen levels in the blood fall markedly. Constriction of the spiral arteries located in the uterine wall occurs, and the portion of the endometrium supplied by these arteries becomes ischemic. As the cells in the endometrium die, tissue is sloughed off and menstruation begins again. It is presently thought that prostaglandins liberated in the endometrium may have a role in stimulating the sloughing of endometrial tissue.

If fertilization of the ovum occurs, the embryo arrives in the uterus on about the fourth day of development. Small glands in the endometrium stimulated by progesterone produce a nutritive fluid for the developing embryo. On approximately the seventh day after fertilization, the embryo implants itself in the thick endometrium of the uterus, and development of the placenta occurs. The placenta secretes the hormone

human chorionic gonadotropin (hCG), which in turn signals the corpus luteum to continue to function. Subsequent events in pregnancy are described later in this chapter.

KEY POINTS

- The monthly reproductive cycle averages about 28 days. Beginning on the first day of menses, the important events of the cycle are as follows:
  1. The endometrial layer is sloughed.
  2. The ovarian follicles are stimulated by pituitary follicle-stimulating hormone (FSH).
  3. Estrogen is secreted from the developing follicles.
  4. Proliferation of the endometrium occurs in response to estrogen.
  5. At the midpoint of the cycle, a burst of LH and a doubling of FSH secretion from the pituitary gland stimulate ovulation.
  6. The ruptured follicle changes into a corpus luteum and secretes estrogen and progesterone.
  7. In the absence of pregnancy, secretion of estrogen and progesterone drops rapidly and the endometrial lining sloughs off again to complete the cycle.
- With fertilization and implantation of the ovum, the developing placenta secretes hCG, which in turn stimulates the corpus luteum to continue to secrete estrogen and progesterone and thus prevent endometrial sloughing.

BREAST

The breast is an important accessory organ in sexual function and human reproduction. Although its primary physiologic function is lactation (production of milk) to nourish the human infant, the significance of the breast as a symbol of feminine sexuality in contemporary Western culture must also be recognized.

Structure of the Breast

The breasts are located anterior to the pectoralis major muscle and are separated from it by a layer of fat. The position of the breasts is maintained by fibrous bands called *Cooper ligaments*, which are easily stretched, especially if the breasts are large. Lymph drainage from the breasts is mainly toward the axillary lymph nodes, with some drainage toward the substernal and diaphragmatic lymph nodes.

Each breast consists of 15 to 20 lobes of glandular epithelial tissue and a ductal system embedded in interstitial tissue and fat. The secretory cells that constitute the glandular epithelium are arranged in grapelike clusters called *alveoli* (Fig. 32.8). Ducts or openings from each alveolus unite to form a single duct from each lobe. These main ducts then enlarge slightly into ampullae immediately before opening onto the surface of the nipple. The nipple, located at the center of the adult female breast, is composed of bundles of smooth muscle fibers with erectile properties. The areola that surrounds the nipple has a diameter of 1.5 to 2.5 cm. The openings from the lactiferous ducts are arranged radially under the areola; thus 15 to 20 small openings are located on the surface of each nipple through which milk flows in a lactating female.

Breast Development

The five Tanner stages of development of the female breast are depicted in Fig. 32.6. As shown in this figure, the breasts contain only rudimentary glands during childhood. At puberty, estrogen and progesterone, in the presence of growth hormone and prolactin, promote the development

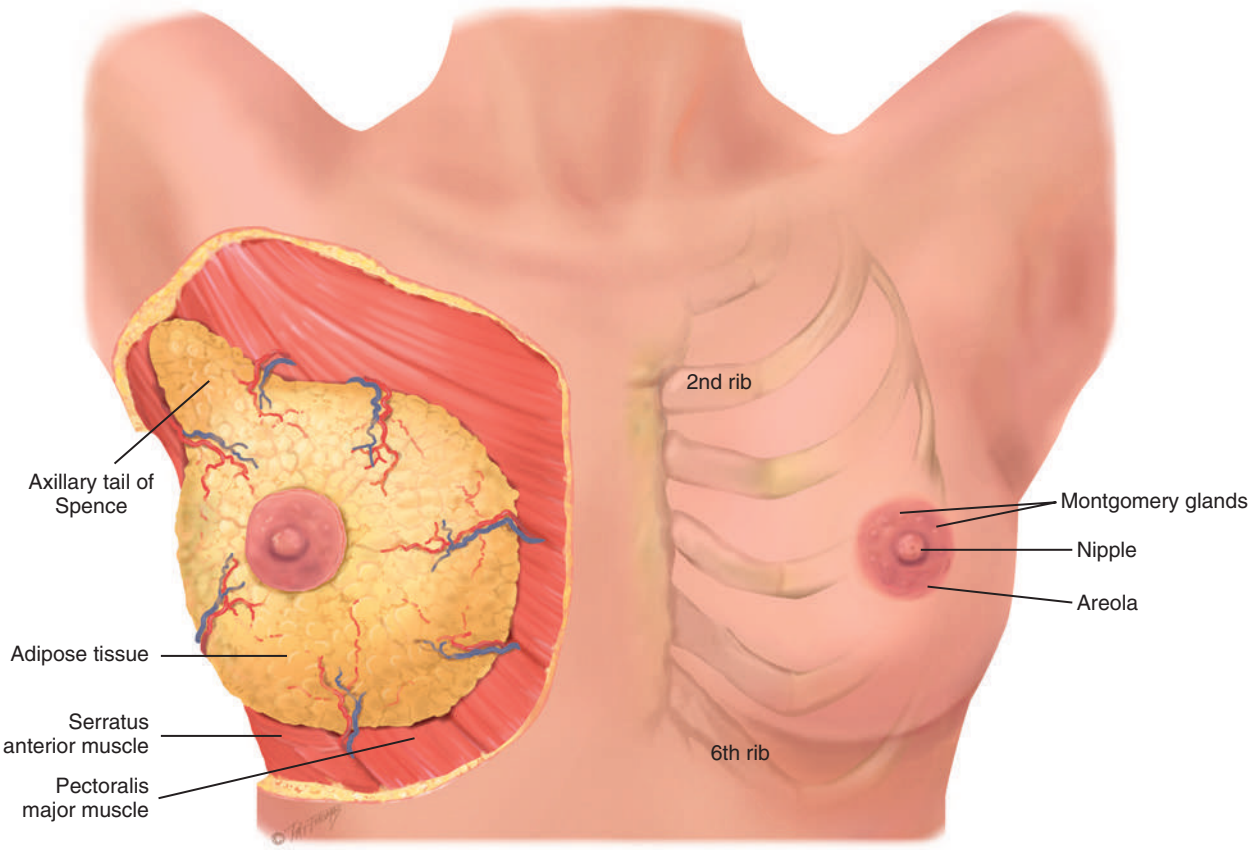


FIG 32.8 Mature female breast. (From Jarvis C: *Physical examination and health assessment,* ed 7, St Louis, 2016, Elsevier, p 385.)

of glandular tissue and ducts and the deposition of fat characteristic of the adult female breast. Throughout the reproductive years, some women note swelling of the breast around the latter part of each menstrual cycle before the onset of menstruation. The water retention and subsequent swelling of breast tissue during this phase of the menstrual cycle are thought to be due to high levels of circulating progesterone stimulating the secretory cells of the breast.

Lactation

During pregnancy, high concentrations of estrogen and progesterone produced by the corpus luteum and the placenta stimulate the development of glands and ducts in the breast. During the first trimester of pregnancy, the ducts proliferate; in the second trimester, the ducts group together to form large lobules with new alveoli formation. In the third trimester, the existing alveoli dilate in preparation for lactation. Toward the end of pregnancy and until 1 to 3 days after childbirth, the mammary glands form colostrum, which contains protein and lactose but little fat. After birth of the infant, the hormone prolactin secreted by the mother's anterior pituitary gland stimulates milk production, and milk is produced by the third day after delivery. The initiation and maintenance of lactation are a complex neuroendocrine process involving sensory nerves in the nipples and breast tissue, the spinal cord and hypothalamus, and the pituitary gland. The suckling movements of the infant on the breast stimulate the release of prolactin from the anterior pituitary gland and oxytocin from the posterior pituitary gland. These hormones in turn stimulate lactation and ejection of milk from the alveoli into the ducts, where it is accessible to the infant. Oxytocin then promotes the actual release of milk, called the *milk ejection reflex.*

KEY POINTS

- At puberty, breast development occurs in response to estrogen and progesterone in cooperation with growth hormone and prolactin. During pregnancy, high estrogen and progesterone levels stimulate further development of the mammary glands and ducts.
- Milk production and release are stimulated by the pituitary hormones prolactin and oxytocin in response to suckling.

PREGNANCY

During the 9 months of human gestation, the single-celled zygote gives rise to an infant with a complex set of physiologic systems. The fertilized ovum contains the entire genetic complement—or encoded genetic instructions—to develop into a fully functioning term infant, given adequate nutrition and time. Three basic developmental processes—growth, morphogenesis, and cellular differentiation—are involved in this transformation. Growth denotes the proliferation of new cells by mitosis, a necessary but not sufficient process for development. The arrangement of cells in a particular order is called *morphogenesis* and is essential to the elaboration of higher forms of life. In addition to growth and morphogenesis, cellular differentiation is needed for cells to specialize structurally and biochemically in a myriad of ways. This section describes the sequence of events in which growth, morphogenesis, and cellular differentiation function to transform a human zygote with encoded genetic information into a human infant. In addition, this section will describe the response of the mother's body to pregnancy. Information on genetic control of inheritance and genetic disorders is contained in Chapters 5 and 6, respectively, and the reader may wish to refer to these chapters for specific content in these areas.

Early Human Development

Fertilization of the ovum occurs in the oviduct. Within 24 hours after fertilization, the zygote begins a series of divisions by the process of mitosis; this process is referred to as *cleavage* (Fig. 32.9). From a two-cell entity the zygote soon divides multiple times, and its cytoplasm begins to be partitioned into specific cells that will serve as the building blocks of the embryo. As more cleavage takes place, the embryo is transported through the oviduct to the uterus. This process takes about 4 days. The embryo receives nutrition during this time from secretions released by the epithelial cells lining the oviduct. After the embryo enters the uterus, the zona pellucida, the membrane surrounding the embryo, dissolves. About day 4, the embryo arrives in the uterus and floats freely while receiving nutrition from secretions from the endometrial glands stimulated by progesterone.

At this point the cells of the embryo have arranged themselves into a hollow spherical structure called the *blastocyst* (Fig. 32.9). The outer cells of the blastocyst, called the *trophoblast,* will ultimately become the protective and nutritive membranes (chorion and placenta) that surround the developing embryo. The inner cell mass, a small cluster of cells that projects into the cavity of the blastocyst, will develop into the structures of the embryo itself. If at this point the inner cell mass divides into two separate groups of cells, identical twins with an identical genetic complement will result. Fraternal twins develop when two ova are fertilized by two sperm cells and do not have an identical genetic complement.

Implantation

On approximately day 7 after fertilization, the embryo attaches to the uterine lining and then implants itself in the endometrium (Fig. 32.10). Enzymes secreted by the trophoblast erode a small portion of the uterine lining, and by day 10 of development the embryo has completely penetrated the endometrium. The opening in the uterine lining is closed, initially by a blood clot and then by regeneration of uterine epithelium; all subsequent development of the embryo occurs in the wall of the uterus.

Fetal Membranes and Placenta

Fetal membranes protect the developing embryo or fetus and provide needed substrates for growth and development, particularly oxygen and nutrition. In addition, they serve the purpose of elimination of waste products of metabolism. All terrestrial vertebrates have four fetal membranes: amnion, yolk sac, chorion, and allantois. In the developing human, the yolk sac is usually thought to be a vestigial structure, although it serves as an important temporary center for the formation of blood cells between the second and sixth weeks. The allantois is also considered vestigial, although its blood supply contributes to formation of the umbilical vessels.

The amnion begins to develop at a very early stage and eventually expands to surround the entire embryo. The space between the amnion and the embryo is called the amniotic cavity. It is filled with a clear amniotic fluid that keeps the embryo moist and provides a measure of protection against mechanical injury.

The placenta serves two basic functions. It is the organ of exchange between the developing fetus and the mother; it also provides nutrients to the fetus and removes wastes. It is also an endocrine organ and produces several hormones, most notably hCG. The placenta develops from both the chorion and the maternal uterine tissue. After implantation, the chorion develops rapidly and forms highly vascularized villi while the embryonic circulation develops. The umbilical cord develops and connects the embryo with the placenta. Two umbilical arteries arise in the umbilical cord and

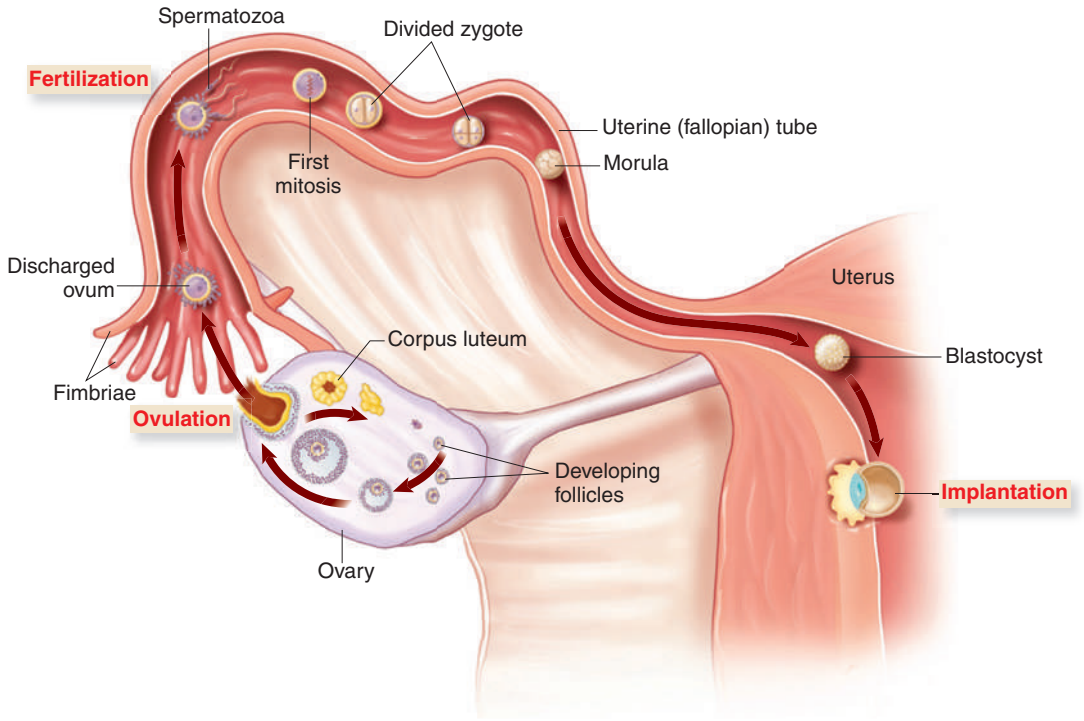


FIG 32.9 Early human development. Drawings illustrate cleavage of the zygote and formation of the blastocyst. At ovulation, an ovum is released from the ovary and begins its journey through the uterine tube. While in the tube, the ovum unites with a sperm to form the single-celled zygote. After a few days of rapid mitotic division, a ball of cells called a morula is formed. After the morula develops into a hollow ball called a blastocyst, implantation occurs. (From Patton KT, Thibodeau GA: *Human body in health and disease,* ed 7, St Louis, 2018, Elsevier, p 654.)

connect with a rapidly proliferating network of capillaries in the villi. The umbilical vein, also located in the umbilical cord, carries blood from the villi back to the fetus.

The placenta eventually consists of the portion of the chorion in which villi develop, along with the uterine tissue between the villi that contains maternal capillaries and small pools of maternal blood. The placenta brings maternal blood adjacent to fetal blood, although the two circulatory systems are completely separate from each other. Thus oxygen and nutrient substrates pass from the maternal blood through the placental tissue and diffuse into the blood of the fetus, where these substances can be used for growth and development of various body tissues. Waste products of fetal metabolism from fetal blood then pass through the placenta into the maternal blood supply and are eventually transported to the maternal kidneys for disposal.

The placenta, like the corpus luteum, secretes both estrogen and progesterone during pregnancy. These hormones serve a variety of purposes in pregnancy. Estrogen promotes enlargement of the uterus and growth of the ductal structure of the breast, as well as alters the elasticity of various pelvic ligaments and the symphysis pubis to allow passage of the infant through the pelvic structures during delivery. In addition to its role in providing early nutrition for the embryo, progesterone has the special effect of decreasing contractility of the gravid uterus, thus preventing spontaneous abortion. In addition, progesterone may have a role in preparing the breasts for lactation, as described earlier in the Lactation section.

Of major importance in the role of the placenta as an endocrine gland is its production of hCG. From the time of implantation, the trophoblastic cells begin to secrete hCG, which sends a signal to the corpus luteum that a pregnancy has begun. The corpus luteum responds by increasing its size and its secretion of estrogen and progesterone, which then promote continued development of the endometrium and the placenta. In the absence of hCG, the corpus luteum would disintegrate, as it does in a nonfertilized menstrual cycle, and the endometrium would deteriorate and be sloughed off along with the embryo. Thus hCG is an essential element in continuation of the pregnant state.

Development of the Human Embryo and Fetus

From fertilization to the end of the eighth week, the developing organism is referred to as an *embryo;* from the ninth week until birth, the developing baby is referred to as a *fetus.* Development of the fetus proceeds in an orderly sequence of complex events. With recent developments in fetal physiology, it is possible to predict which structures will begin their development or function on a particular day of development after conception. Table 32.2 depicts some important developmental events from the time of fertilization to birth. Detailed information on the development of organ systems during fetal life is contained in the chapters in this book that focus on these organ systems; for example, Chapter 35 contains a description of the development of the gastrointestinal tract.

First Month

Rapid growth, morphogenesis, and cell differentiation occur early in development of the human embryo. By 2.5 weeks of development, the notochord and neural plate are formed; these structures eventually give rise to the central nervous system. In addition, the tissue that will form the heart has differentiated. By the end of the first month, an S-shaped heart beats about 60 times per minute, and the three primary vesicles of the brain have formed.

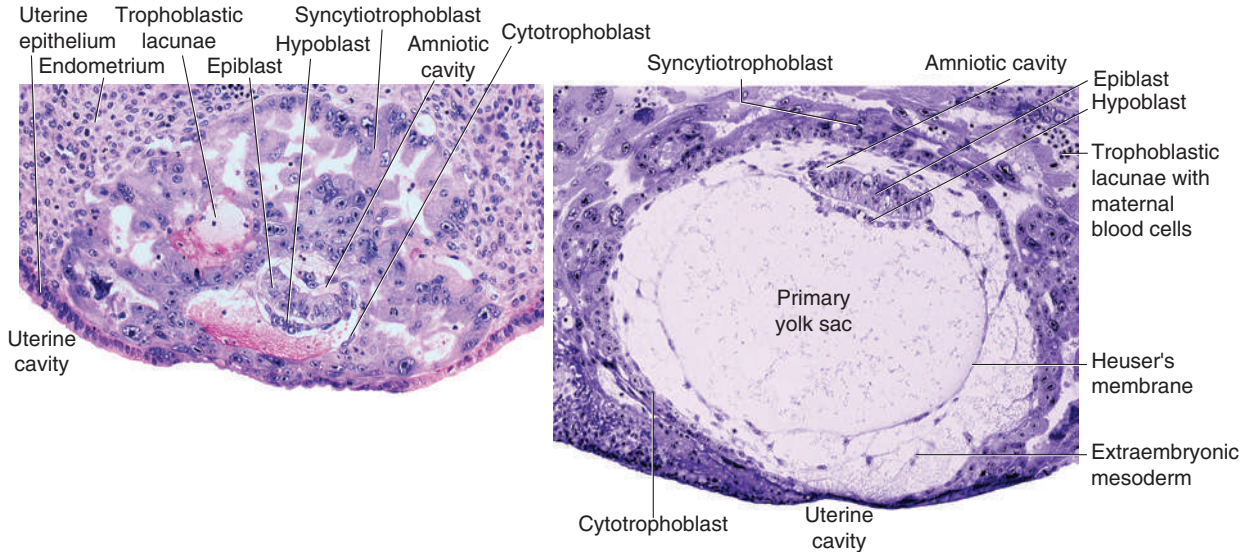


FIG 32.10 Nine-day human embryo at the stage of amnion formation. (From Schoenwolf GC, Bleyl SB, Brauer PR, Francis-West PH. *Larsen's human embryology,* ed 5, Philadelphia, 2015, Churchill Livingstone.)

Second Month

Until the sixth week of gestation, the gonads in both genders are *bipotential,* which means that the gonads present in the embryo may become either testes or ovaries. Beginning about the seventh week, the so-called *indifferent gonad* begins to develop into either a male or a female derivative. Recent research has demonstrated that *SRY* (sex-determining region of the Y chromosome) is the gene that influences the indifferent gonad to organize into a testis. In a genetically female embryo, the gonad organizes into an ovary under the influence of one or more ovary-determining genes, which have not yet been well characterized. The cortex of the gonad accumulates nests of cells that differentiate into ovarian follicles, each containing a primary oocyte. The Wolffian ducts, the primordial structures that are precursors to the male internal reproductive organs, begin to disappear, and the Müllerian ducts, the structures that will develop into the female internal reproductive organs, become dominant.

The external genitalia of both the male and the female are identical until the eighth week of gestation. Like the gonads, the genitalia are bipotential until this time, with the capability of developing into organs of either gender. In a genetically male embryo, dihydrotestosterone, a metabolite of testosterone, binds to androgen receptors in the external genitalia and affects the differentiation of these structures into the male external genitalia. Without the influence of dihydrotestosterone, the bipotential external genitalia will spontaneously develop into female external genitalia.

Fig. 32.11 shows an embryo on day 49. All of the organs continue to develop during the second month, and the embryo becomes capable of movement. The major blood vessels assume their final positions, and the heart assumes its final shape. The brain begins to transmit impulses to regulate function of the organ systems, and a few reflexes are now present. At the end of the second month, the rudiments of all organs are present and the embryo is referred to as a *fetus.*

Third Month

During the third month, the ears and eyes approach their final positions, and some of the bones become distinct. The fetus performs breathing movements consisting of moving amniotic fluid in and out of the lungs and can carry on sucking movements. By the end of the third month, the fetus is almost 56 mm in length and weighs about 14 g (Fig. 32.12).

Second Trimester

A *trimester* refers to a period of 3 months during pregnancy. During the second trimester, or months 4 to 6 of development, the fetus achieves independent mobility and can move freely through the amniotic cavity. The heartbeat of the fetus is now audible through a stethoscope and averages 150 beats per minute. By the fifth month of development, the fetus measures 250 mm (10 inches) in length, which is half its total length at birth. Fig. 32.13 shows a fetus in the second trimester at 4 months of development.

Third Trimester

By far the greatest growth of the fetus occurs during the third trimester. The weight of the fetus almost doubles during the last 2 months. In addition, final differentiation of tissues and organs takes place. Survival of infants born prematurely during this time has increased markedly in the past few years because of an enhanced ability to sustain vital functions such as respiration and regulation of body temperature in neonatal intensive care settings.

Parturition

Parturition refers to the process by which the infant is born. Toward the end of pregnancy, the uterus becomes progressively more excitable until it begins strong rhythmic contractions that ultimately expel the infant. At the present time, the exact cause of the increased uterine activity remains unknown. However, two sets of effects have been suggested as contributing to the increased excitability of uterine musculature at this time: progressive hormonal changes and progressive mechanical changes.

Hormonal Changes

During the latter part of pregnancy, large amounts of estrogen, which has a definite tendency to increase uterine contractility, are secreted. Concurrent with this enhanced estrogen release, the secretion of progesterone, which inhibits uterine contractility, remains constant or may decrease slightly. Thus it is hypothesized that the increased ratio of

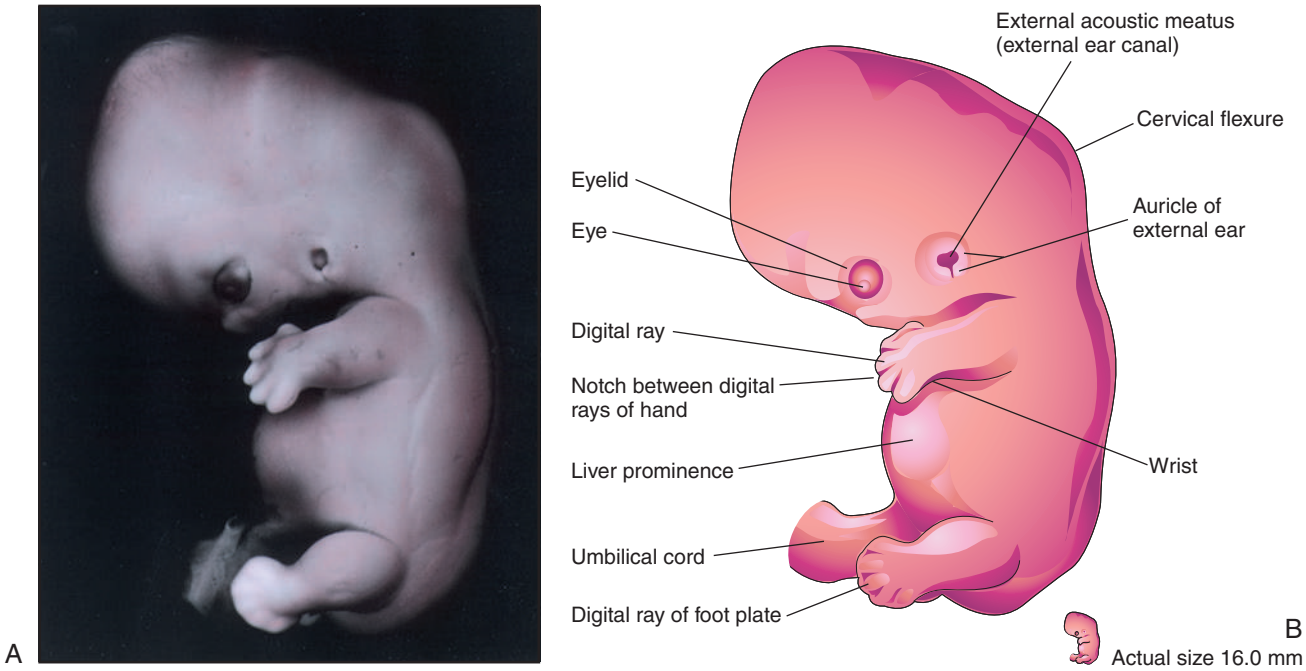


FIG 32.11 Human embryo in the seventh week of development. (A, From Moore KL et al: *Color atlas of clinical embryology,* ed 2, Philadelphia, 2000, Saunders. B, From Moore KL, Persaud TVN, Torchia MG: *The developing human: clinically oriented embryology,* ed 10, Philadelphia, 2016, Saunders.)

TABLE 32.2 Summary of Developmental Events in Human Fetal Life

Time From Fertilization	Key Events
36 hours	Embryo has achieved two-cell stage
4 days	Embryo reaches uterus
7 days	Implantation of embryo in uterine wall
2.5 weeks	Differentiation of heart tissue
	Blood cell formation in yolk sac and chorion
	Formation of notochord and neural plate
3.5 weeks	Formation of neural tube
	Heart tubes begin to beat
	Primordial eye and ear visible
	Respiratory system begins development
	Liver bud differentiates
	Blood vessels established
4 weeks	Formation of three primary brain vesicles
	Limb buds appear
2 months	Embryo capable of movement
	Cerebral cortex differentiating
	Gonad identifiable as testis or ovary
	Bones begin ossification and muscles are differentiating
	Major blood vessels in final positions
3 months	Fetus performs breathing and sucking movements
	Gender is clearly identifiable
5 months	Heartbeat is audible with a stethoscope
	Fetus moves freely through amniotic cavity
6 to 9 months	Rapid growth with final differentiation of tissues and organs
266 days	Birth

FIG 32.12 Photograph of the human fetus at 11 weeks of development. (From Moore KL, Persaud TVN, Torchia MG: *The developing human: clinically oriented embryology,* ed 10, Philadelphia, 2016, Saunders.)

FIG 32.13 **A,** Side view of human fetus at 17 weeks. **B,** Frontal view of 17-week fetus. (A, From Moore KL et al: *Color atlas of clinical embryology,* ed 2, Philadelphia, 2000, Saunders. B, Courtesy Dr. Robert Jordan, St. Georges University Medical School, Grenada.)

estrogen to progesterone secretion in the latter part of pregnancy may promote the increased contractility of the uterus.

Oxytocin is a hormone secreted by the posterior pituitary gland that specifically causes uterine contraction and is thought to have a major role in promoting increased uterine contractility during parturition. The rate of oxytocin secretion is considerably increased at the time of labor (see the following Mechanical Changes section), and the uterus displays increased responsiveness to a given dose of oxytocin at this time.

Mechanical Changes

Stretching smooth muscle organs increases their contractility; in addition, intermittent stretching of smooth muscle can elicit contraction. Thus it is hypothesized that the stretch or irritation of the fetal head against the cervix begins a reflex action that causes the uterus to contract. As the cycle of stretching and contraction is repeated again and again, increased contractions result. In addition, stretching of the cervix causes the release of oxytocin from the posterior pituitary. Oxytocin then stimulates additional uterine contractions, thus initiating another positive feedback cycle of stretching and contraction.

Response of the Mother's Body to Pregnancy

The presence of a developing fetus in the uterus creates an extra physiologic load for the pregnant woman, with resulting effects on her basal metabolism and specific organ systems. Normal physiologic responses to pregnancy are described here; complications of pregnancy are discussed in Chapter 33.

Metabolism During Pregnancy

As a result of increased secretion of many hormones, including thyroxine, adrenocortical hormones, and the sex hormones, the basal metabolic rate increases by about 15% during the latter half of pregnancy. This increase in metabolism results in alterations in many organ systems, including the circulatory, respiratory, and urinary systems.

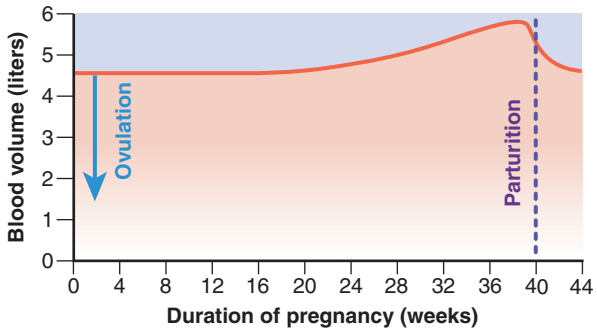


FIG 32.14 Effect of pregnancy on the mother's blood volume.

Changes in the Female Reproductive Organs

The hormones secreted during pregnancy, either by the placenta or by the endocrine glands, directly promote alterations in body structures. In particular, the organs of the female reproductive tract increase markedly in size, with the uterus increasing from 30 to 1100 g and the breasts approximately doubling in size. Concurrently, the vagina enlarges with a widening of the vaginal introitus.

Changes in the Circulatory System

In the latter stages of pregnancy, about 625 mL of blood flows through the maternal circulation of the placenta each minute. This factor, along with a general increase in metabolism, causes an increase in maternal cardiac output to 30% to 40% above normal by week 27 of pregnancy. However, for reasons not understood at the present time, cardiac output decreases to a little above normal during the last 8 weeks of pregnancy, although the high uterine blood flow continues. As shown in Fig. 32.14, an increase in maternal blood volume occurs mainly during the latter half of pregnancy. This increase is mainly due to

hormonal factors. Both aldosterone and estrogens, which are greatly increased in pregnancy, promote increased fluid retention by the kidneys. In addition, bone marrow increases its activity to produce an excess of red blood cells to accompany the excess vascular volume. At the time of parturition, the mother has an additional 1 to 2 extra liters of blood in her circulatory system.

Changes in the Respiratory System

The increased basal metabolic rate and size of the pregnant woman result in an increase in oxygen utilization, with utilization of oxygen being 20% above normal at the time of birth. Concurrently, a commensurate amount of carbon dioxide is formed. In addition, the growing uterus is pressing upward against the abdominal organs, which in turn press against the diaphragm and cause a decrease in diaphragmatic excursion. The net result of these changes is an increase in minute ventilation of approximately 50% and a decrease in arterial P_{CO_2} to slightly below normal.

Changes in the Urinary System

Because of an increased load of excretory products, the rate of urine formation in pregnancy is usually slightly increased. In addition, other alterations in urinary function occur. Renal tubule reabsorption of sodium, chloride, and water is increased as a result of increased production of steroidal hormones by the placenta and adrenal cortex. Concurrently, the glomerular filtration rate often increases by as much as 50%, a change that serves to increase the rate of water and electrolyte loss in the urine. These two events tend to balance each other out, with the result that only a moderate excess of water and salt accumulation occurs under normal circumstances. However, in the condition of toxemia of pregnancy, excess water and salt accumulation may occur with life-threatening consequences.

Weight Gain and Nutrition During Pregnancy

The average weight gain during pregnancy is about 24 lb, with most of this gain occurring during the last two trimesters. Approximately 7 lb of this weight gain is the fetus; 4 lb of the increased weight is amniotic fluid, placenta, and fetal membranes; 2 lb represents an increase in uterine tissue; and another 2 lb of the weight gain is an increase in breast tissue. Thus an average 9-lb increase in weight occurs in the remainder of the woman's body. Approximately 6 lb of fluid may be excreted during the days after birth, after loss of the fluid-retaining hormones of the placenta.

Appetite may be greatly increased during the latter part of pregnancy, in part because of fetal removal of food substrates from the mother's blood and partly because of hormonal factors. The developing fetus assumes priority in regard to many of the nutritional substrates of the mother's body fluids and will continue to grow even when maternal nutrition is inadequate. However, although fetal length may increase normally in the absence of adequate maternal nutrition, fetal weight will be considerably decreased, and abnormal bone formation and decreased size of many bodily organs of the fetus may result.

If the intake of nutritional elements during pregnancy is inadequate, a number of deficiencies can be present in the mother. In particular, deficiencies of calcium, phosphates, iron, and vitamins may be present. As an example, approximately 375 mg of iron is needed by the fetus to form its blood, and an additional 600 mg is needed by the mother to form her own extra blood supply. Because the normal store of nonhemoglobin iron in the mother at the beginning of pregnancy is often about 100 mg and seldom greater than 700 mg, anemia will develop in a pregnant woman without sufficient iron intake in her food. Important also is adequate folic acid intake, which has been shown to help prevent neural tube defects.

KEY POINTS

- At about the seventh day after fertilization, the embryo attaches to the uterine lining. The placenta is the fetal lifeline that provides nutrients and oxygen and eliminates wastes. The placenta also secretes hCG, which is important in maintaining pregnancy.
- Normal gestation is about 9 months. Each 3-month period is called a trimester. By the end of the first trimester, fetal structures and organ systems are present. During the second and third trimesters, the fetus grows in size and weight.
- Near the end of the third trimester, an increase in estrogen production and mechanical stretching of the uterus and cervix are thought to induce parturition. Cervical stretching stimulates the release of oxytocin from the pituitary gland. Oxytocin stimulates uterine contractions.
- Pregnancy is associated with many physiologic changes, including an increased basal metabolic rate (15%), increased cardiac output (30% to 40%) and blood volume (1 to 2 L), increased oxygen consumption (20%) and minute ventilation (50%), increased glomerular filtration rate and tubular reabsorption of sodium and water, and increased body weight (24 lb).

MENOPAUSE

Although menopause is defined specifically as the last menstrual period in a woman's reproductive life, the term is often used to denote the entire period of years before and after this event in which the function of the ovaries is in transition. The terms *climacteric* and *perimenopause* are used in the health care literature to describe this transitional period. At about 45 to 52 years of age the supply of ovarian follicles declines, with the majority becoming atretic or degenerated. With the depletion of ovarian follicles, secretion of estrogen and progesterone by the ovaries declines, and the menstrual cycle becomes irregular. When too little estrogen is secreted to cause endometrial growth, menstrual periods stop permanently.

The decline in ovarian hormone production that occurs in the perimenopausal period causes important physiologic changes in a woman's body. The decline in plasma estrogen levels may result in a number of distressing symptoms, although some women experience no symptoms during this time. Hot flushes, described by women as an unpleasant sensation of sudden warmth sweeping upward over the abdomen, chest, neck, and face, are experienced by nearly 75% of postmenopausal women. Although the precise cause of hot flushes is unknown, it is thought that decreased estrogen levels have an effect on the temperature-regulating center in the hypothalamus. Hot flushes are often accompanied by other symptoms of autonomic nervous system instability such as tachycardia, palpitations, and feelings of faintness. Other distressing symptoms, including pain and stiffness in the joints, sleep pattern disturbances, and changes in gastrointestinal function, have been noted by women in the perimenopausal period. These symptoms are presently the focus of many nursing research projects examining the health of aging women. Although such psychological symptoms as increased nervousness have been reported in the medical literature as being related to the hormonal imbalance in menopause, it has been established that psychological symptoms are not directly related to estrogen deficiency. There is an increasing awareness of the importance of the role of culture and other social factors in the understanding of how women view and experience menopause.

With the decline in estrogen level associated with perimenopause, many structural changes occur in various organs. These changes are summarized in Geriatric Considerations: Changes in the Female Reproductive System. The epidermis of the skin becomes thinner and less elastic

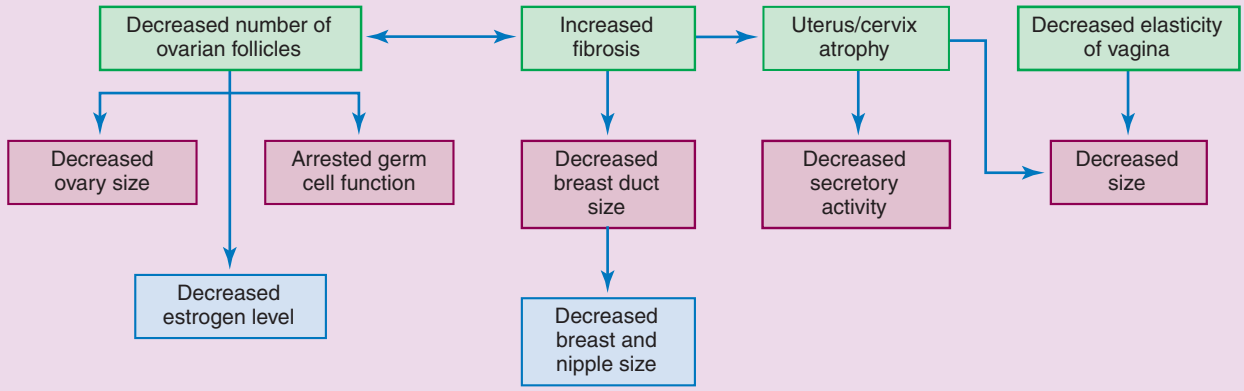
throughout the entire body. The breasts may decrease in size; the labia may also lose their underlying fat and become thinner. The vaginal epithelium may become thin and atrophied, with the result that sexual intercourse may be painful. The decline in estrogen level also leads to osteoporosis and decreased bone density, particularly in white women, with resulting bone fractures. Exercise and supplemental calcium and vitamin D are recommended for postmenopausal women to prevent accelerated bone loss. At present, most authorities recommend estrogen therapy during the perimenopausal period only to prevent and relieve symptoms such as hot flushes and vaginal atrophy. This is due to strong evidence showing that supplemental estrogen and progestin therapy has been associated with an increased risk of breast cancer and cardiovascular disease. Newer treatments that have been developed for osteoporosis include bisphosphonates, calcitonin nasal spray, and raloxifene. These medications typically fall into two categories: drugs that slow bone loss, and medications that increase rate of bone formation. Work continues on the development of selective estrogen receptor modulators and other drugs that may retain some of the beneficial effects of estrogen while avoiding the negative effects. Women in perimenopause may wish to discuss the risks and potential benefits of hormone replacement therapy and other menopausal therapies with their health care providers before making an informed decision about these medications.

KEY POINTS

- Menopause begins at 45 to 52 years of age and denotes the cessation of menstruation. A declining supply of ovarian follicles with decreased estrogen and progesterone production results in irregular menses and then complete cessation of menstruation.
- Declines in estrogen production are associated with hot flushes, tachycardia, palpitations, faintness, joint pain, and sleep disturbances. Structural changes associated with menopause include osteoporosis, thinning of the skin, and atrophy of the vaginal structures and breast tissue.

GERIATRIC CONSIDERATIONS

Changes in the Female Reproductive System



Female reproductive system function declines with organ-specific tissue changes. The number of active female germ cells declines over time with variable function before they are arrested in menopause. The ovaries become smaller and increasingly fibrotic and have fewer ovarian follicles.

The secretion of estrogen by the ovaries stops at menopause, resulting in a marked estrogen level decrease. The ovarian follicles become insensitive to gonadotropins (FSH and LH). However, the peripheral conversion of androgens to estrogen causes a small maintenance level of estrogen to persist at 10% to 30% of previous levels. The androgen-producing ovarian cells (hilar and thecal) continue to secrete testosterone in postmenopausal women.

The follicles, uterus, and cervix undergo atrophy with a decrease in size and secretory action. The vagina is reduced in size with a loss of elasticity and atrophy of the vaginal epithelium. The vascular supply to the vaginal walls decreases with reduced amounts of glycogen and mucopolysaccharide. The pH of Bartholin gland secretions is increased (i.e., more alkaline) because of the loss of estrogen.

The breasts decrease in size. Breast ducts become smaller and are replaced by fat tissue. Some fibrosis and calcification may occur within the ducts. The nipples are smaller with less nipple pigmentation. The aging female nipple may be normal or retracted.

SUMMARY

This chapter has described the major processes related to the human female reproductive tract, including the menstrual cycle, pregnancy, lactation, and menopause. In approaching this material, the reader must view the information presented within the current context of social change in which women are taking an active role in meeting their health care needs. In addition, recent research in the area of reproductive endocrinology has yielded a rapidly expanding understanding of the reproductive structures and their function.

The female reproductive structures are a complex set of organs with multiple, integrated functions. Careful review of the section on reproductive structures, including their embryologic development, will assist the reader in understanding the various alterations in these structures that occur throughout a woman's life. Although the hormonal and structural changes occurring in the female reproductive organs may at first seem overwhelmingly complex to the student, some basic concepts will help in organizing this material. First, the menstrual cycle has two significant results: production of an ovum and preparation of the uterus for implantation of the fertilized ovum. Second, the fertilized ovum contains the entire encoded genetic instructions to produce a unique human individual. Third, pregnancy consists of three basic

developmental processes—growth, morphogenesis, and cellular differentiation—to bring about this transformation, which will also result in multiple changes in the body of the mother. The breast, with its function of lactation, is also a component of the reproductive system and is subject to alterations throughout a woman's life span. Finally, menopause is not a discrete event but rather a process during which the supply of ovarian follicles declines. A review of these concepts will prepare the student for a better understanding of women's health concerns and provide a basis for approaching the next chapter, which considers alterations in reproductive functioning.

RESOURCES

Reproductive Anatomy and Physiology

Apter D: Development of the hypothalamic-pituitary-ovarian axis. *Ann N Y Acad Sci* 17(816):9–21, 1997.

Banister EM, Begoray DL, Daly LK: Responding to adolescent women's reproductive health concerns: empowering clients through health literacy. *Health Care Women Int* 32(4):344–354, 2011.

Barrett KE, Barman SM, Boitano S, Brooks HL: *Ganong's review of medical physiology*, ed 25, New York, 2015, McGraw-Hill.

Genazzani AR, et al: Neuroendocrinology of the menstrual cycle. *Ann N Y Acad Sci* 17(816):143–150, 1997.

Hacker NF, Gambone JC, Hobel CJ: *Hacker & Moore's Essentials of obstetrics and gynecology*, ed 6, Philadelphia, 2016, Elsevier.

Hall JE: *Guyton and Hall textbook of medical physiology*, ed 13, Philadelphia, 2015, Elsevier Saunders.

Hall SK, Moreau C, Trussell J: Determinants of and disparities in reproductive service use among adolescent and young adult women in the United States, 2002-2008. *Am J Public Health* 102(2):359–367, 2012.

Jacobs J, Stanfors M: Racial and ethnic differences in U.S. women's choice of reversible contraceptives, 1995-2010. *Reprod Health* 45(3):139–147, 2013.

Jayes FL, Burns KA, Rodriguez KF, et al: The naturally occurring luteinizing hormone surge is diminished in mice lacking estrogen receptor Beta in the ovary. *Biol Reprod* 90(2):1–9, 2014.

Lopez LM, Tolley EE, Grimes DA, et al: Theory-based interventions for contraception. *Cochrane Database Syst Rev* (8):CD007249, 2013.

Sokalska A, Valentin L: Changes in ultrasound morphology of the uterus and ovaries during the menopausal transition and early postmenopause: a 4-year longitudinal study. *Ultrasound Obstet Gynecol* 31(2):210–217, 2008.

Solomon EP: *Introduction to human anatomy and physiology*, ed 4, Philadelphia, 2015, Saunders.

Tucker EE: Shared decision-making and decision support: their role in obstetrics and gynecology. *Curr Opin Obstet Gynecol* 26(6):523–530, 2014.

Pregnancy and Fetal Development

AlSaif S, Mumtaz S, Wray S: A short review of adipokines, smooth muscle and uterine contractility. *Life Sci* 125:2–8, 2015.

Barboza-Arguello M, Umana-Solis L, Azofeifa A, et al: Neural tube defectsin Costa Rica, 1987-2012: origins and development of birth defect surveillance and folic acid fortification. *Matern Child Health J* 19(3):583–590, 2015.

Blackburn ST: *Maternal, fetal, and neonatal physiology*, ed 4, Philadelphia, 2012, Saunders.

Capel B: Sex in the 90s: SRY and the switch to the male pathway. *Annu Rev Physiol* 69:497–523, 1998.

Cumbers MR, Chung ST, Wakerley JB: A neuromodulatory role for oxytocin within the supramammillary nucleus. *Neuropeptides* 41(4):217–226, 2007.

Fanchin R: Assessing uterine receptivity in 2001: ultrasonographic glances at the new millennium. *Ann N Y Acad Sci* 943:185–202, 2001.

Hermoso M, Vollhardt C, Bergmann K, Koletzko B: Critical micronutrients in pregnancy, lactation, and infancy: considerations on vitamin D, folic acid, and iron, and priorities for future research. *Ann Nutr Metab* 59(1):5–9, 2011.

Kashimada K, Koopman P: Sry: the master switch in mammalian sex determination. *Development* 137:3921–3930, 2010.

Polin RA, Abman SH, Rowitch D, Benitz WE: *Fetal and neonatal physiology*, ed 5, Philadelphia, 2016, Elsevier.

Wilson CA, Davies DC: The control of sexual differentiation of the reproductive system and brain. *Reproduction* 133(2):331–359, 2007.

Menopause

ACOG Practice Bulletin No. 141: Management of menopausal symptoms. *Obstet Gynecol* 123(1):202–216, 2014.

Brotman R, Shardell MD, Gajer P, et al: Association between the vaginal microbiota, menopause status, and signs of vulvovaginal atrophy. *Menopause* 21(5):45–458, 2014.

Hunter MS: Psychological aspects of the menopause. In Panay N, Briggs P, Kovacs G, editors: *Managing the menopause: 21st century solutions*, Cambridge, 2015, Cambridge University Press.

Lewiecki EM, Miller PD, Harris ST, et al: Understanding and communicating the benefits and risks of denosumab, raloxifene, and teriparatide for the treatment of osteoporosis. *J Clin Densitom* 17(4):490–495, 2014.

National Osteoporosis Foundation: *Clinician's guide to prevention and treatment of osteoporosis*, Washington, DC, 2010, Author.

Nelson HD: Menopause. *Lancet* 371(9614):760–770, 2008.

North American Menopause Society: Management of osteoporosis in postmenopausal women: 2010 position statement of the North American Menopause Society. *Menopause* 17(1):25–54, 2010.

North American Menopause Society: The 2012 hormone therapy position statement of the North American Menopause Society. *Menopause* 19(3):257–271, 2012.

North American Menopause Society: Management of symptomatic vulvovaginal atrophy: 2013 position statement of the North American Menopause Society. *Menopause* 20(9):888–902, 2013.

Reginster J, Neuprez A, Beaudart C, et al: Antiresorptive drugs beyond bisphosphonates and selective oestrogen receptor modulators for the management of postmenopausal osteoporosis. *Drugs Aging* 31(3):413–424, 2014.

Writing Group for the Women's Health Initiative Investigators: Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA* 288:321–333, 2002.

33

Alterations in Female Genital and Reproductive Function

Rosemary A. Jadack

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- What are the differentiating factors of the common menstrual disorders?
- What are the common etiologic factors leading to uterine prolapse, uterine retrodisplacement, cystocele, and rectocele?
- How can the pain of endometriosis be differentiated from that of dysmenorrhea?
- What is the rationale for routine Papanicolaou testing for cervical cancer?
- What factors contribute to the high mortality rate of ovarian cancer?
- What clinical findings would indicate the development of pregnancy-induced hypertension, placenta previa, and abruptio placentae in a pregnant woman?
- How can benign and malignant breast lumps be clinically differentiated?

CHAPTER OUTLINE

Menstrual Disorders, 672
Amenorrhea, 672
Abnormal Uterine Bleeding Patterns, 673
Dysmenorrhea, 673
Alterations in Uterine Position and Pelvic Support, 674
Uterine Prolapse, 674
Retrodisplacement of the Uterus, 675
Cystocele, 675
Rectocele, 676
Inflammation and Infection of the Female Reproductive Tract, 676
Pelvic Inflammatory Disease, 676
Vulvovaginitis, 677
Bartholinitis, 677
Benign Growths and Aberrant Tissue of the Female Reproductive Tract, 678
Uterine Leiomyomas, 678
Ovarian Cysts, 678
Endometriosis, 679
Cancer of the Female Genital Structures, 679
Cancer of the Cervix, 679
Endometrial Cancer, 680
Ovarian Cancer, 680
Vaginal Cancer, 680
Cancer of the Vulva, 680
Disorders of Pregnancy, 681
Pregnancy-Induced Hypertension, 681
Hyperemesis Gravidarum, 681
Placenta Previa and Abruptio Placentae, 681
Spontaneous Abortion, 681
Disorders of the Breast, 682
Reactive-Inflammatory Breast Disorders, 682
Mammary Duct Ectasia, 682
Breast Abscess, 682
Fat Necrosis, 682
Reactions to Foreign Material, 683
Benign Breast Disorders, 683
Fibrocystic Breast Disease, 683
Specific Benign Neoplasms, 683
Malignant Disorder of the Breast, 684
Cancer of the Breast, 684

The complex functioning of the female reproductive system described in Chapter 32 may be subject to alterations in structure and function throughout a woman's life that can have far-reaching effects on her health and well-being. This chapter is a survey of these alterations and describes the pathophysiologic basis of the most common disorders of the female reproductive system. In addition, current therapeutics for these alterations, including pharmacologic therapy, will be summarized. The information presented here is an introduction to these complex areas, and the reader may wish to consult in-depth gynecology and obstetrics texts for more detailed information.

Perhaps no other function of the human body is so closely linked to psychological, social, and spiritual concerns as reproductive function.

Any alteration in reproductive status (or the perceived threat of such an alteration) may have profound effects on an individual. Clinicians caring for women experiencing alterations in functioning of the reproductive system should bear in mind the profundity of such alterations for the individual woman and must also maintain an awareness of the context in which women seek help for such problems. The therapeutic environment for women seeking help for reproductive concerns requires a clinical approach in which information is freely shared and mutual decisions are reached between caregiver and client. Previous clinical approaches in which women's concerns were labeled as unimportant or merely psychogenic often resulted in anger, frustration with health care providers, and withdrawal from the health care delivery system. Women are now seeking active involvement in their own care, and clinicians who care for women experiencing the alterations described in this chapter need to approach women's health concerns with sensitivity and openness.

MENSTRUAL DISORDERS

Alterations in the normal functioning of the menstrual cycle include amenorrhea (no menses), abnormal uterine bleeding patterns, and dysmenorrhea (painful menstruation). Although many pathologic conditions can cause these alterations, an obvious cause is often not found.

Amenorrhea

Etiology and pathogenesis. Amenorrhea is the absence or suppression of menstruation in a female age 16 years or older; it occurs if a woman misses three or more consecutive periods. Amenorrhea is categorized as either primary or secondary. Primary amenorrhea is the failure to begin menses by the age of 16 years. Secondary amenorrhea is the cessation of established, regular menstruation for 6 months or longer. Fig. 33.1 shows causes of primary and secondary amenorrhea.

Amenorrhea is normal before menarche (the first menstrual period at the time of puberty), after menopause, and during pregnancy and lactation. At other times, it is considered pathologic and may result from a wide range of pathophysiologic causes (see Fig. 33.1). In the majority of cases, amenorrhea is due to an abnormal pattern of hormonal functioning that interrupts the normal sequence of events in which the endometrial tissue lining the uterus proliferates and then is sloughed. The endometrial tissue must be stimulated and regulated by the correct quantity and sequence of the female sex hormones estrogen and progesterone and

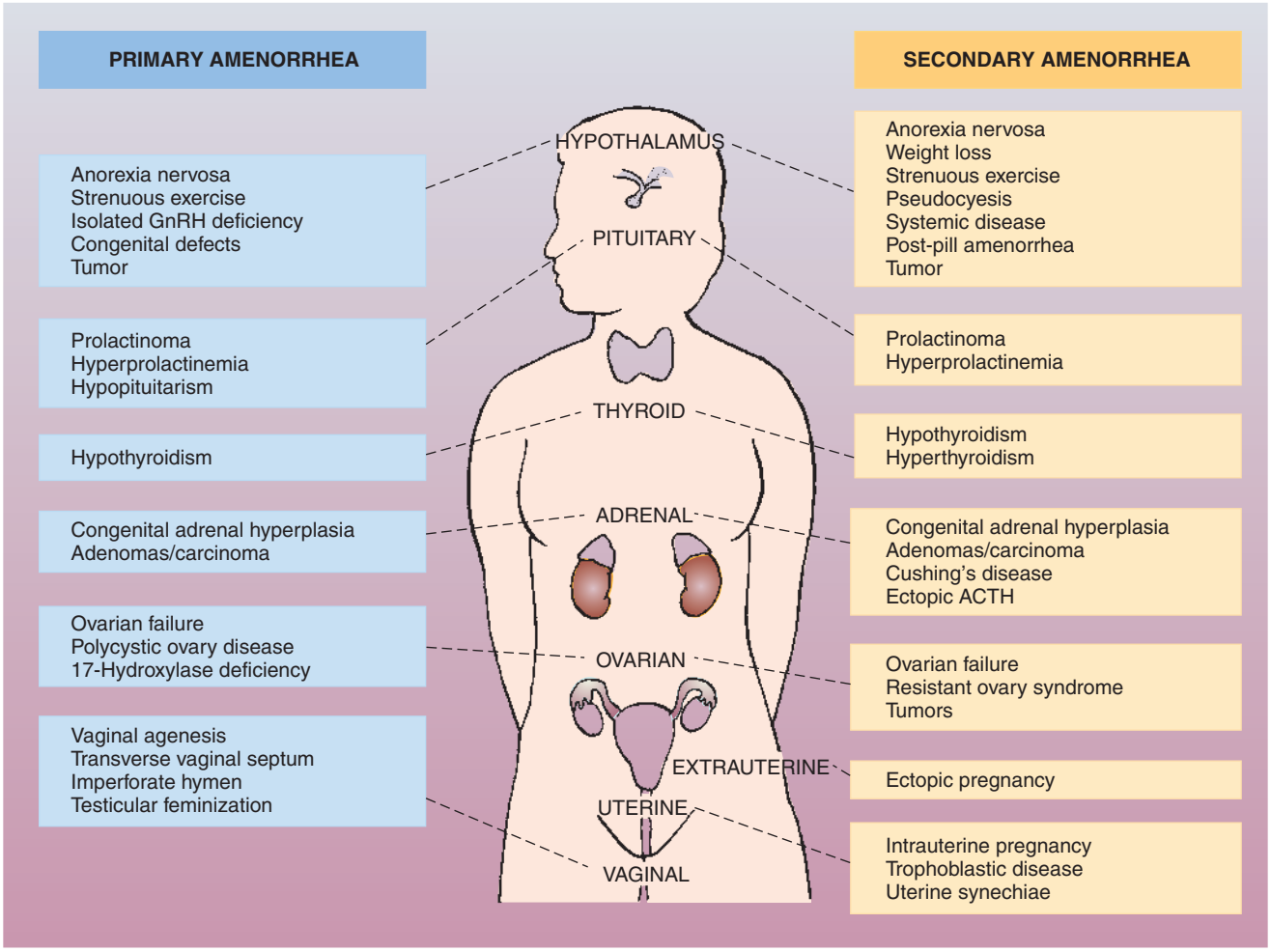


FIG 33.1 Causes of amenorrhea. *ACTH,* Adrenocorticotropic hormone; *GnRH,* gonadotropin-releasing hormone. (From Black JM, Hawks JH: *Medical-surgical nursing: clinical management for positive outcomes,* ed 8, Philadelphia, 2009, Saunders, p 915.)

the gonadotropic hormones follicle-stimulating hormone (FSH) and luteinizing hormone (LH). As described in Chapter 32, the menstrual cycle is dependent on the sequential changes in estrogen and progesterone levels. The initial rise in LH and FSH levels in the menstrual cycle occurs in response to a decline in estrogen and progesterone levels; estrogen levels then rise again in response to actions of the gonadotropic hormones, and the endometrium proliferates again in response to estrogen secretion. Thus events that prevent estrogen production interfere with the normal fluctuations in estrogen levels and those that block the action of estrogen on the endometrium will result in abnormal or absent menstrual flow. Such events may include physical or emotional stress, which can interfere with normal production of the gonadotropic hormones and alter the pattern of estrogen functioning. In addition, ovarian, adrenal, or pituitary tumors may interfere with the normal production of female sex hormones or LH and FSH. Neoplasms of the ovaries or adrenal and pituitary glands may result in excess or deficient production of these hormones, with a consequent interruption in normal menstrual flow.

Treatment. Therapeutic strategies for amenorrhea are directed to correcting the cause of the interruption in hormonal functioning and may include the use of hormonal supplementation to reinstate a normal sequence of events in the menstrual cycle. If amenorrhea is the result of a neoplastic process, surgery may be indicated for tumor removal.

Abnormal Uterine Bleeding Patterns

Irregular or excessive bleeding from the uterus is one of the most common alterations in the female reproductive system. Uterine bleeding that varies from a woman's normal pattern either in quantity or in frequency may occur at any age and for a variety of reasons.

Etiology, clinical manifestations, and treatment. The most common alterations in uterine bleeding patterns and their causes are described here. Metrorrhagia**,** or bleeding between menstrual periods, usually results from slight physiologic bleeding from the endometrium during ovulation, but may also result from other causes such as uterine malignancy, cervical erosions, and endometrial polyps or as a side effect of estrogen therapy. Hypomenorrhea**,** or a deficient amount of menstrual flow, results from endocrine or systemic disorders that may interfere with proper functioning of the hormones in the menstrual cycle, or it may be due to partial obstruction of menstrual flow by the hymen or a narrowing of the cervical os. Oligomenorrhea**,** or infrequent menstruation, usually reflects failure to ovulate because of an endocrine or systemic disorder with accompanying inappropriate hormonal function. Similarly, polymenorrhea**,** an increased frequency of menstruation, may be associated with ovulation and may be caused by endocrine or systemic factors. Menorrhagia**,** an often-debilitating increase in the amount or duration of menstrual bleeding, usually results from lesions of the female reproductive organs such as uterine leiomyomas, endometrial polyps, and adenomyosis. It is often managed with surgery, oral contraceptives, and/or antiprostaglandins. A progestin-containing intrauterine device has shown promise in reducing menorrhagia, dysmenorrhea, and anemia.

The term *dysfunctional uterine bleeding* is used to describe abnormal endometrial bleeding not associated with tumor, inflammation, pregnancy, trauma, or hormonal effects. Dysfunctional uterine bleeding is most common around the time of menarche and menopause and not as common in women before menopause. In adolescents, dysfunctional uterine bleeding is most often due to immaturity in functioning of the pituitary and ovary, which have not yet properly orchestrated their activities. Thus an imbalance may be present in the ratio of estrogen to progesterone. Absent or diminished levels of progesterone will result in a thick and extremely vascular endometrium that lacks structural support. As a result of this fragile structure, spontaneous and superficial hemorrhage occurs randomly throughout the endometrium. In addition, the blood vessels in the endometrium fail to constrict to limit the extent and duration of bleeding. Uterine bleeding that is abnormal in both quantity and frequency can therefore occur in a noncyclic pattern.

In perimenopausal women, dysfunctional uterine bleeding may be the result of progressive degeneration and failure of the ovary to produce estrogen. As the number of ovarian follicles diminishes, the production of estrogen by the ovary becomes unpredictable, and the secretion of LH and FSH may also assume an unpredictable pattern. As in adolescents with dysfunctional uterine bleeding, diminished or absent production of progesterone may result in unopposed stimulation of the endometrium by estrogen, with subsequent unpredictable bleeding from a fragile endometrium.

Dysmenorrhea

Dysmenorrhea is menstruation that is painful enough to limit normal activity or to cause a woman to seek health care. Dysmenorrhea is a widespread phenomenon that affects many women across the reproductive years, including girls of high school age through perimenopausal women. Although symptoms of dysmenorrhea tend to decrease with age, the traditional notion that childbirth permanently decreases symptoms is unfounded. In addition, the contention that women with dysmenorrhea tend to be neurotic has been refuted in well-designed psychiatric research studies. Recent research into the physiologic process of uterine contractions has enhanced our understanding of the causes of dysmenorrhea and has thus resulted in better treatment.

Etiology and clinical manifestations. Dysmenorrhea is usually classified as primary (not related to any identifiable pathologic condition) or secondary (related to an underlying pathologic condition). The cramps that occur with primary dysmenorrhea are usually located in the suprapubic region and are sharp in quality. The pain may radiate to the inner aspect of the thighs and lower sacral area and may be accompanied by nausea, diarrhea, and headache. Primary dysmenorrhea usually develops 1 or 2 years after menarche, when ovulatory cycles are established. Under the influence of progesterone, increased amounts of prostaglandins, potent hormonelike unsaturated fatty acids, are released from the endometrium. Prostaglandins have significant effects on smooth muscle and vasomotor tone; when released from the endometrium, prostaglandins promote uterine contractions and ischemia of the endometrial capillaries and thereby cause the cramping pain of dysmenorrhea.

Secondary dysmenorrhea is characterized more often by dull pain that may increase with age. It is associated with pelvic disorders such as endometriosis, leiomyomas, or pelvic adhesions.

Treatment. Recent therapeutic strategies for the management of primary dysmenorrhea have focused on the phenomenon of prostaglandin-induced enhanced uterine contractility. The use of prostaglandin synthetase inhibitors such as ibuprofen, celecoxib, and naproxen, which inhibit the formation of prostaglandins, has been effective in many women experiencing dysmenorrhea. Although evidence supporting their effectiveness is limited, other approaches that use steroid hormones, such as progestins or combined high-progestin/low-estrogen oral contraceptives, have also been advocated. The rationale is that production of the high menstrual levels of prostaglandins needed to produce dysmenorrhea requires high levels of estrogen without progesterone in the proliferative phase of the menstrual cycle. Progestin administration therefore inhibits the production of prostaglandins and relieves the symptoms of dysmenorrhea. However, the use of steroid hormones may involve significant risks, which the individual client must weigh against the benefits of such therapy.

Therapeutic strategies for secondary dysmenorrhea may involve diagnostic operative procedures such as laparoscopy, as well as medical and surgical therapy for the underlying condition.

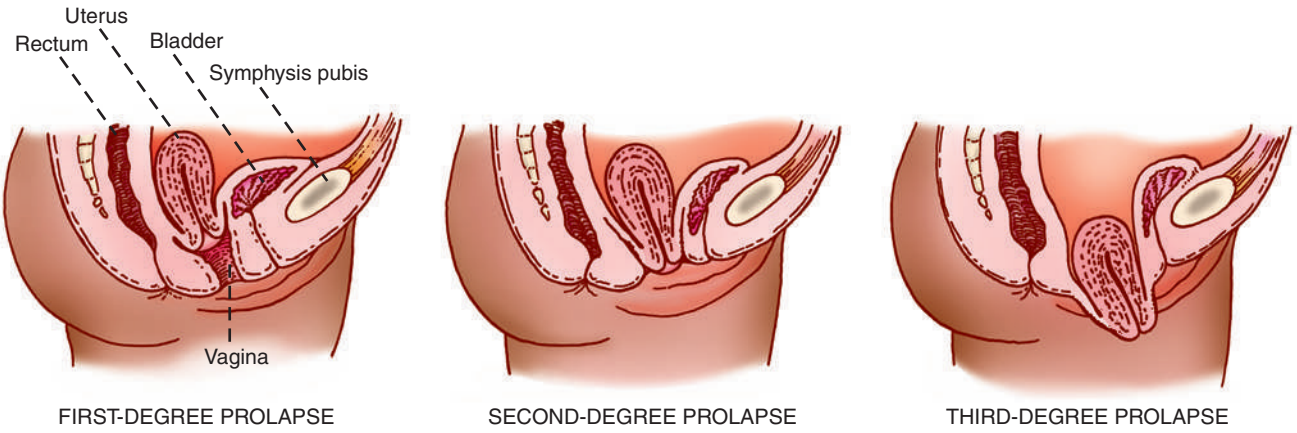


FIG 33.2 Degrees of uterine prolapse. (From Black JM, Hawks JH: *Medical-surgical nursing: clinical management for positive outcomes,* ed 8, Philadelphia, 2009, Saunders, p 931.)

KEY POINTS

- Amenorrhea, the absence of menstruation, is most commonly due to hormonal disturbances. Stress and neoplasms (ovarian, adrenal, or pituitary tumors) may interfere with the normal patterns of hormone secretion. Treatment is aimed at the underlying cause of the hormonal imbalance.
- Irregular or excessive uterine bleeding is a common problem. Metrorrhagia is bleeding between periods, hypomenorrhea is reduced menstrual flow, oligomenorrhea is infrequent menstruation, polymenorrhea is an increased frequency of menstruation, and menorrhagia is prolonged and heavy bleeding during menstruation. These disorders may be associated with hormonal imbalances or primary lesions of the reproductive tract.
- Dysfunctional uterine bleeding is common at menarche and menopause and is due to irregular secretion of reproductive hormones. Other causes of abnormal bleeding, such as tumor, trauma, inflammation, and endocrine diseases, are ruled out before a diagnosis of dysfunctional uterine bleeding is made.
- Dysmenorrhea is painful menstruation, generally described as sharp suprapubic cramping severe enough to limit activity. Dysmenorrhea may be treated with prostaglandin inhibitors. Dysmenorrhea secondary to pelvic disorders (endometriosis, adhesions) generally has a dull quality and may increase with age.

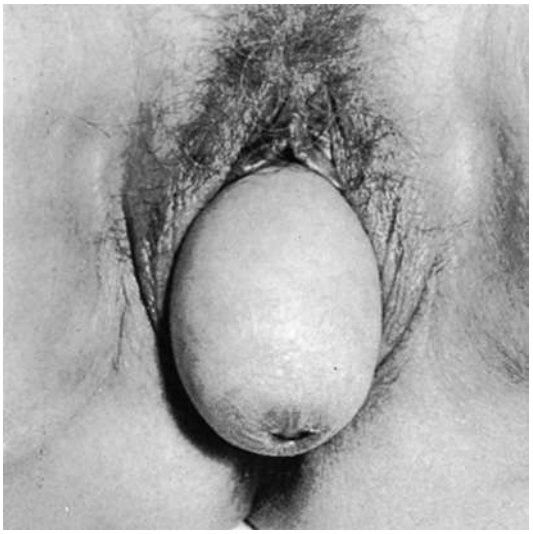


FIG 33.3 Complete uterine prolapse. (From Parsons L, Sommers SC: *Gynecology,* ed 2, Philadelphia, 1978, Saunders, p 1443.)

ALTERATIONS IN UTERINE POSITION AND PELVIC SUPPORT

Alterations in uterine position and pelvic support may occur anytime during a woman's reproductive years. The major support for the uterus and upper part of the vagina is provided by the thickenings of the endopelvic fascia known as the *cardinal ligaments.* Although tearing of the cardinal ligaments during labor and delivery is rare, they can be stretched abnormally during a difficult or prolonged delivery and subsequently fail to support the pelvic organs adequately. In addition, congenital defects in the muscles of the pelvic floor may promote alterations in position of the uterus and other pelvic structures. The two most common alterations in uterine position are uterine prolapse and retrodisplacement of the uterus. Other commonly occurring alterations resulting from a weakening of the vaginal and pelvic floor musculature are cystocele and rectocele.

Uterine Prolapse

Etiology. The axis of the uterus normally forms an acute angle with the axis of the vagina. This anatomic feature itself tends to prevent a prolapse, or sinking, of the uterus from its normal position. Descent of the uterus occurs when supporting structures, such as the uterosacral ligaments and the cardinal ligaments, relax and allow the relationship of the uterus to the vaginal axis to be altered. This relaxation permits the cervix to sag downward into the vagina. If the support of the vaginal wall is also compromised, the pressure of the abdominal organs on the uterus will gradually force it downward through the vagina into the introitus. Uterine prolapse may occur at any age. In female infants and in women who have never given birth, congenital defects in the basic integrity of the pelvic supporting structures are usually responsible. Trauma to the ligaments during childbirth is the cause of uterine prolapse in women who have given birth, particularly if multiple deliveries have occurred. Uterine prolapse is classified as first degree, second degree, or third degree according to the level to which the uterus has descended (Fig. 33.2). In first-degree prolapse, the uterus is approximately halfway between the vaginal introitus and the level of the ischial spines. In second-degree prolapse, the end of the cervix has begun to protrude through the introitus. In third-degree or complete prolapse, the body of the uterus is outside the vaginal introitus. Fig. 33.3 shows a third-degree, or complete, uterine prolapse.

Clinical manifestations. The symptoms of uterine prolapse depend on the degree of severity. The woman may become increasingly aware

of a sensation of bearing down and discomfort in the vagina. If the prolapse has advanced to the second or third degree, she may note discomfort while walking or sitting and have difficulty urinating. In addition, as the end of the cervix begins to protrude outside the body, it may be subject to trauma from friction and ulceration. Bleeding and ulceration of the cervix may be present.

Treatment. Uterine prolapse is one of the most common reasons for hysterectomy, usually from the more minimally invasive vaginal approach. In patients who are at poor risk for surgery or who choose not to have a hysterectomy, a pessary, which is a small supportive device, is inserted to hold the uterus in place.

Retrodisplacement of the Uterus

The term *retrodisplacement* refers to situations in which the body of the uterus is displaced from its usual location overlying the bladder to a position in the posterior of the pelvis. As shown in Fig. 33.4, the uterus may be in one of five positions: anteverted, midposition, anteflexed, retroflexed, or retroverted.

Etiology and clinical manifestations. Retrodisplacement can be detected in 20% to 30% of all women. It may be a normal variation and therefore be present throughout a woman's entire life, or it may develop after childbirth when the supporting structures are injured.

In many women, no symptoms occur from uterine retrodisplacement. In some women, symptoms of pelvic pain or pressure, dysmenorrhea, and dyspareunia (painful intercourse) may be present. In addition, infertility has been associated with retrodisplacement.

Treatment. If the woman has no symptoms, no treatment is indicated. The use of a pessary to support the uterus in a normal position may relieve the symptoms, but surgical correction is sometimes indicated when symptoms are severe. If surgery is indicated, less invasive, laparoscopic surgical procedures are often preferred.

Cystocele

Etiology. A *cystocele* is a protrusion of a portion of the urinary bladder into the anterior of the vagina at a weakened part of the vaginal musculature (Fig. 33.5A). The defect in the vaginal wall is usually caused by injury during childbirth or surgery, but may also result from the aging process or develop as an inherent weakness. Other predisposing factors include obesity and a history of lifting heavy objects. The pressure created by this protrusion causes the anterior vaginal wall to bulge in a downward direction.

Clinical manifestations and treatment. A wide range of symptoms may be present, depending on the degree of severity of the cystocele. A mild degree of protrusion of the bladder may result in no symptoms. In moderate to severe cases, a sensation of pressure can be felt in the vagina, along with dysuria, incontinence, and back pain. Fullness at the vaginal opening may be observed, as may a soft, reducible mucosal mass bulging into the anterior of the vaginal introitus.

Surgical repair of the vagina is done to correct the cystocele and reestablish support of the anterior vaginal wall. The bladder is restored to a normal position by reinforcement of the weakened portion of the anterior vaginal wall. Prosthetic mesh may also be inserted to further support the bladder during the repair of the cystocele.

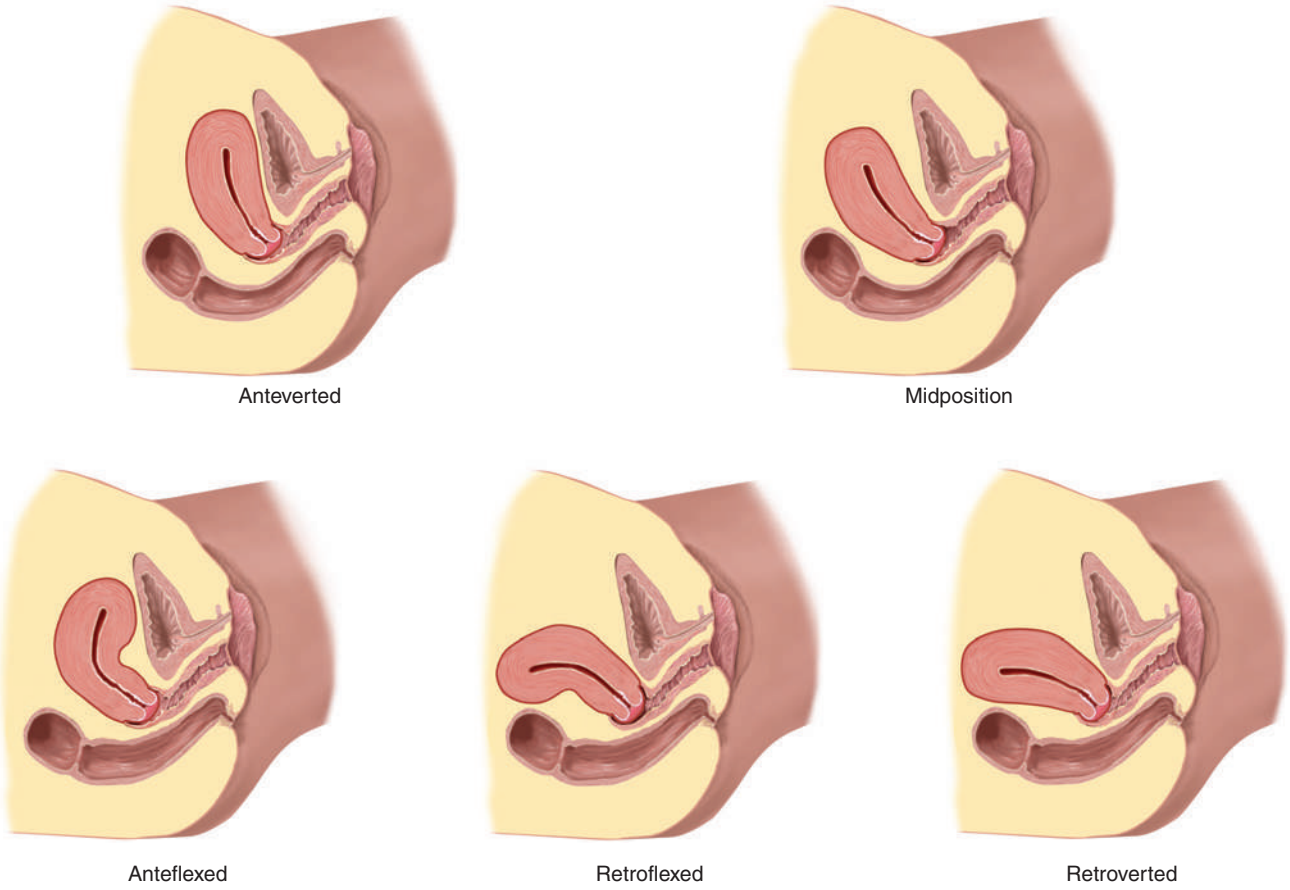


FIG 33.4 Various positions of the uterus. Note that the classifications describe the position of the long axis of the uterus with respect to the long axis of the body. (From Jarvis C: *Physical examination and health assessment,* ed 7, St Louis, 2016, Elsevier, p 756.)

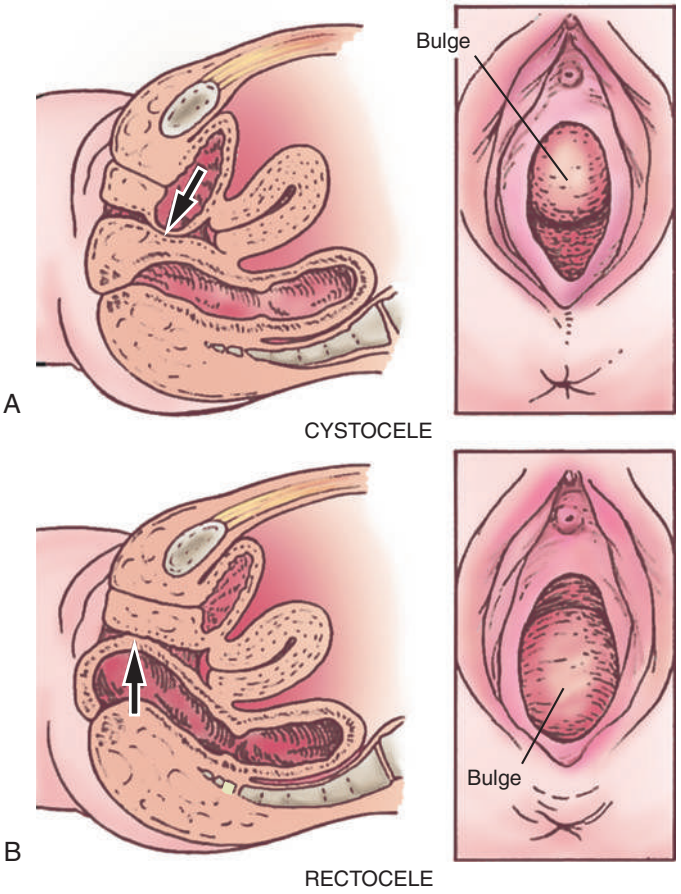


FIG 33.5 A, Cystocele. Note the bulging of the anterior vaginal wall. The urinary bladder is displaced downward. The cystocele pushes the anterior wall downward into the vagina. **B,** Rectocele. Note the bulging of the posterior vaginal wall. (From Black JM, Hawks JH: *Medical-surgical nursing: clinical management for positive outcomes,* ed 8, Philadelphia, 2009, Saunders, p 930.)

Rectocele

Etiology. A rectocele (also called *proctocele*) is a protrusion of the anterior rectal wall into the posterior of the vagina at a weakened part of the vaginal musculature (see Fig. 33.5B). As with a cystocele, the defect in the vaginal wall is usually caused by injury during childbirth or surgery, but may also occur with aging or arise as an inherent weakness. Other predisposing factors for a rectocele include multiparity, obesity, and postmenopausal status. The rectocele forms a bulging mass beneath the posterior vaginal mucosa and pushes downward into the lower vaginal canal. Gradually, the rectum may be torn from its fascial and muscular attachments to the pelvic wall. The levator ani muscles may also become stretched or torn.

Clinical manifestations and treatment. A wide range of symptoms may be present, depending on the degree of severity of the rectocele. The patient may report a history of difficulty in bowel evacuation and may have experienced chronic constipation with laxative and enema dependency. A feeling of pressure may also be reported, along with painful sexual intercourse. Physical examination reveals a mass bulging into the posterior of the vaginal introitus.

Surgical repair of the vagina is done to correct the rectocele and reestablish support of the posterior vaginal wall. The rectum is restored to its normal location, and the levator ani muscles are realigned in proper position.

KEY POINTS

- Uterine prolapse occurs when supporting pelvic structures relax and the cervix sags downward into the vagina. Congenital defects, pregnancy, and childbirth are the usual contributing factors. Prolapse may be accompanied by a sensation of pelvic fullness and vaginal discomfort.
- Retrodisplacement of the uterus is common (20% to 30% of women) and may be congenital or due to pregnancy and childbirth. The body of the uterus is flexed or rotated into the posterior of the pelvis, which sometimes leads to varied symptoms of pelvic pain or pressure, dysmenorrhea, and dyspareunia.
- A cystocele may result from weakness in the vaginal musculature that allows the urinary bladder to protrude into the anterior of the vagina. Contributing factors include childbirth, surgery, aging, obesity, and heavy lifting. Vaginal pressure, dysuria, and back pain may be present.
- A rectocele may result from weakness in the posterior vaginal musculature that allows the rectum to protrude into the vagina. Contributing factors are similar to those for cystocele. Symptoms include constipation, painful bowel evacuation, and painful intercourse.

INFLAMMATION AND INFECTION OF THE FEMALE REPRODUCTIVE TRACT

Inflammatory and infectious processes of the female reproductive tract may have effects that range from discomfort to life-threatening situations. Because the infectious agents responsible for inflammation and infection of the female genital tract may be sexually transmitted, some overlap in the discussion of these processes and sexually transmitted diseases is necessary. This chapter will describe the two principal inflammatory and infectious processes of the upper and lower female reproductive tract: pelvic inflammatory disease (PID) and vulvovaginitis. The reader may wish to refer to Chapter 34 for additional information on sexually transmitted infections.

Pelvic Inflammatory Disease

Pelvic inflammatory disease is any acute, subacute, recurrent, or chronic infection of the oviducts and ovaries with involvement of the adjacent reproductive organs (Fig. 33.6). It includes inflammation of the cervix (cervicitis), uterus (endometritis), oviducts (salpingitis), and ovaries (oophoritis). When the connective tissue underlying these structures between the broad ligaments is also involved, the condition is called *parametritis.*

Hospitalizations for PID have declined in 2014. The Centers for Disease Control and Prevention (CDC) reports an approximate 40% decrease in the number of health care provider visits for PID among women aged 15 to 44 years, from 123,000 to 88,000 visits. Significant reproductive health problems may occur as a result of PID. A substantial number of women with a history of PID eventually experience one or more long-term health problems. Among women with PID, tubal scarring can cause infertility in 8%, ectopic pregnancy in 9%, and chronic pelvic pain in 18%.

Etiology. Normally, cervical secretions provide protective and defensive functions for the reproductive organs. By providing a bacteriostatic barrier, cervical mucus prevents bacterial agents present in the cervix or vagina from ascending into the uterus. Therefore conditions or surgical procedures that alter or destroy cervical mucus may impair this bacteriostatic mechanism. PID may follow the insertion of an intrauterine device, pelvic surgery, abortion procedures, and infection during or after pregnancy. Bacteria may also enter the uterine cavity through the bloodstream or from drainage from other foci of infection

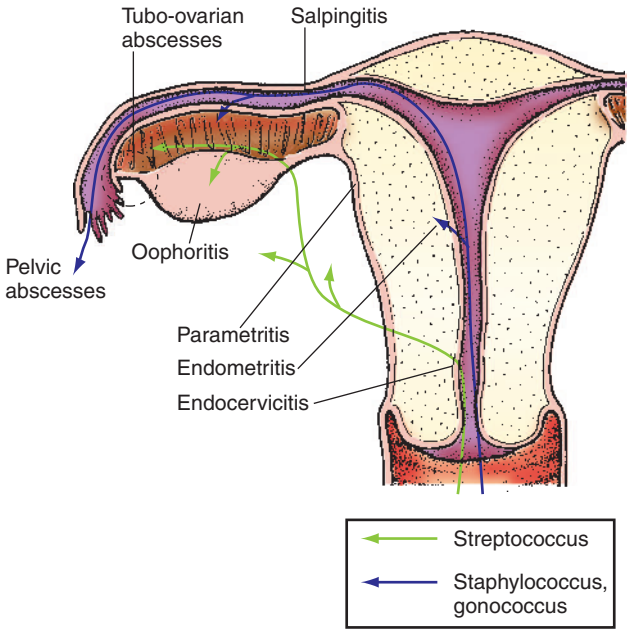


FIG 33.6 Spread of pelvic inflammatory disease. (From Ignatavicius DD, Workman ML: *Medical-surgical nursing,* ed 7, St Louis, 2013, Elsevier, p 1663.)

such as a pelvic abscess, ruptured appendix, or diverticulitis of the sigmoid colon.

PID can result from infection with aerobic and anaerobic organisms. *Neisseria gonorrheae* and *Chlamydia trachomatis* are the most common causative agents because they readily penetrate the bacteriostatic barrier of cervical mucus. However, a variety of bacterial organisms may contribute to the development of PID, including staphylococci, streptococci, diphtheroids, and coliforms such as *Pseudomonas* and *Escherichia coli.* These bacteria are commonly found in cervical mucus, and PID can result from infection by one or several of these bacteria. In addition, PID may occur after multiplication of bacteria in the endometrium that are normally nonpathogenic. During parturition, the traumatized endometrium favors the multiplication of bacteria.

Clinical manifestations. The associated signs and symptoms of PID vary with the affected part of the reproductive tract, but generally include abdominal tenderness and tenderness or pain of the cervix or adnexa on palpation. In addition, the temperature may be elevated higher than 38°C and the white blood cell count elevated greater than 10,000/mm^3. A pelvic abscess or inflammatory mass may be present on physical examination or ultrasound, and purulent vaginal discharge may be noted.

Treatment. Early and aggressive use of antibiotic agents best suited for the causative organisms is essential in preventing the progression of PID. Various oral and parenteral antibiotic regimens involving the use of multiple antimicrobial agents have been suggested by the CDC for use in PID. All regimens used to treat PID should also be effective against *N. gonorrheae* and *C. trachomatis.* Inpatient hospitalization may be indicated for patients with rapidly progressing PID and for those requiring surgical drainage of pelvic abscesses. Rupture of a pelvic abscess is a potentially life-threatening condition, and a total abdominal hysterectomy (removal of the uterus) with bilateral salpingo-oophorectomy (removal of both oviducts and ovaries) may be indicated in this situation.

Vulvovaginitis

Vulvovaginitis is an inflammation of the vulva (vulvitis) and vagina (vaginitis). Because the vulva and vagina are anatomically close to each other, inflammation of one location usually precipitates inflammation of the other. Vulvovaginitis may occur at any time during a girl's or woman's life and affects most females at some point in life.

Etiology. Infection by *Candida albicans* (formerly called *Monilia*) affects an estimated 75% of women at least once in their lifetime. Approximately 40% to 45% of women will have two or more lifetime episodes of vulvovaginitis caused by *C. albicans.* (Infection by *Candida* is referred to as *candidiasis.*) *C. albicans* is a fungus that requires glucose for growth; thus its growth may be promoted during the secretory phase of the menstrual cycle when glycogen levels increase in the vaginal environment. In addition, other conditions in which the glycogen content of the vagina is increased, such as diabetes, pregnancy, and use of oral contraceptives may favor candidiasis. Other factors predisposing to the development of vulvovaginitis from *Candida* infection include the use of estrogen supplementation and antibiotics. Women using estrogen supplementation in the perimenopausal period may be at greater risk for candidal infection of the vagina inasmuch as the glycogen content of the vagina may increase with these therapies. The mechanism by which antibiotic use promotes candidiasis is presently unclear, but it is thought that destruction of the bacteria that normally exert the protective effect of consuming *Candida* results in overgrowth of the *Candida* population with subsequent infection.

Other infectious agents that may result in vulvovaginitis include *Trichomonas vaginalis, Haemophilus vaginalis,* and *N. gonorrheae.* Viral agents that may cause vulvovaginitis include human papilloma virus (HPV) (venereal warts, condylomata acuminata) or herpesvirus type 2. These organisms can be transmitted during sexual intercourse and are discussed in detail in Chapter 34.

In addition to infectious processes, vulvovaginitis may be promoted by conditions or agents that irritate the vulva and vagina. Chemical irritation or allergic reactions to detergents, feminine hygiene products, and toilet paper may be a causative factor. Trauma to the vulva or vagina or the atrophy of the vaginal wall that occurs postmenopausally may predispose to vulvovaginitis as well.

Clinical manifestations. Vulvovaginitis from candidiasis results in a thick, white discharge and red, edematous mucous membranes with white flecks adhering to the vaginal wall. Intense itching (pruritus) usually accompanies this discharge. The vaginal pH is usually normal (less than 4.5), and fungal organisms are often seen on microscopic studies. Vulvovaginitis from other infectious agents may involve a malodorous, purulent discharge. Irritation and subsequent inflammation of the vulva and vagina may be manifested by red, swollen labia; pain on urination and intercourse; and itching.

Treatment. Appropriate medical therapy for the causative organisms is usually instituted, including local antifungal preparations for vaginal candidiasis and local and systemic antibiotic therapy for vulvovaginitis caused by bacterial agents. Cool compresses and sitz baths provide relief of itching and burning of inflamed tissues. Avoidance of factors that promote irritation of the vulva, such as drying soaps, nonabsorptive underwear, and tight clothing, is also of therapeutic benefit.

Bartholinitis

Bartholinitis is an inflammation of the Bartholin glands, which are located on either side of the vaginal orifice and lubricate the vaginal introitus with a clear, viscous secretion. The location of Bartholin glands renders them susceptible to access by bacteria such as *N. gonorrheae, C. trachomatis,* and other organisms.

Clinical manifestations and treatment. Once bacteria are established, an abscess (also referred to as a *Bartholin cyst*) may form and cause tenderness and swelling at the site. Pus may be observed exuding from the duct orifice leading to the affected gland, and symptoms of fever and malaise are present in some individuals. Laboratory culture with proper diagnosis of the causative organism is performed, and appropriate

antibiotic therapy is usually instituted. Surgical incision and drainage of the abscess may be necessary for effective management.

KEY POINTS

- Pelvic inflammatory disease (PID) refers to any infection of the oviducts, ovaries, and adjacent reproductive organs. It includes cervicitis, endometritis, salpingitis, and oophoritis. Manifestations and complications of PID include infertility, ectopic pregnancy, pelvic pain, dyspareunia, and abscesses.
- Intrauterine devices, abortion, and pelvic surgery predispose to PID. *N. gonorrheae* and *C. trachomatis* are the most common causative organisms, and treatment centers on aggressive antibiotic therapy.
- Inflammation of the vulva and vagina, or vulvovaginitis, is a common problem in women. Most cases are associated with fungal infection by *C. albicans* and are manifested as a white vaginal discharge and an irritated, itchy mucosa. Predisposing factors include chemical irritation from feminine hygiene products, trauma, allergic reactions, and antibiotic therapy that inhibits the growth of normal flora.
- Inflammation of the Bartholin glands, or bartholinitis, is typically a result of the entry and subsequent infection of the glands by *N. gonorrheae* or *C. trachomatis*. Tenderness, swelling, and pus may be present and signify the formation of an abscess within one of the Bartholin glands. Antibiotic therapy and surgical drainage are used to manage the abscess.

BENIGN GROWTHS AND ABERRANT TISSUE OF THE FEMALE REPRODUCTIVE TRACT

Benign growths and aberrant tissue in the female reproductive tract are not uncommon; for example, uterine leiomyomas develop in approximately 70% of women by the time they reach the age of 50 years. The presence of benign growths or aberrant tissue in the reproductive tract may cause no symptoms and remain entirely unnoticed, or symptoms ranging from debilitating to life threatening may be present. The diagnosis of these growths or tissue abnormalities may cause anxiety in women experiencing them; in spite of their benign classification, their presence can have devastating effects on the underlying reproductive structures. This section focuses on three of the most common forms of benign growths and aberrant tissue in the female reproductive organs: uterine leiomyomas, ovarian cysts, and endometriosis.

Uterine Leiomyomas

Uterine leiomyomas, which are also called *myomas* or *fibroids,* are the most common form of uterine growths that appear in women. Their actual incidence is difficult to establish because many myomas are either too small or inaccessibly placed to be palpated. Uterine leiomyomas occur in approximately 50% of all premenopausal women and affect black women three times more often than white women. Age appears to be a factor in their development inasmuch as myomas are not found before the onset of puberty and rarely exhibit growth activity after menopause.

Etiology. Uterine leiomyomas make their appearance and exhibit growth activity during the reproductive years. Therefore although the actual cause of myomas is presently unknown, it is thought that estrogen and human growth hormone may influence tumor formation by stimulating susceptible fibromuscular elements in the uterine wall. This theory is supported by the finding that tumor growth is enhanced with the administration of large doses of estrogen and during the later stages of pregnancy, when human growth hormone and estrogen levels are high. In addition, uterine leiomyomas usually shrink or disappear after menopause, when estrogen levels decrease.

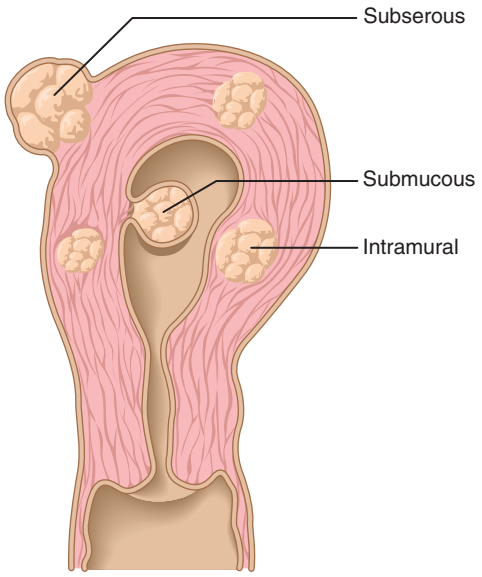


FIG 33.7 Uterine leiomyomas. (From Huether SE, McCance KL: *Understanding pathophysiology,* ed 6, St Louis, 2017, Mosby, p 817.)

Clinical manifestations. Uterine leiomyomas can grow to a large size (Fig. 33.7). Obviously, the presence of such a large mass within the uterus will cause symptoms of abdominal pain and pressure, but smaller myomas can result in such symptoms as well. Other symptoms associated with leiomyomas may include abnormal vaginal bleeding and discharge, depending on the location of the mass. If the myoma is sufficiently large to cause pressure on surrounding abdominal organs, backache, constipation, and urinary frequency or urgency may also be present. Finally, uterine leiomyomas can prevent pregnancy and make carrying a pregnancy to term difficult.

Treatment. Treatment for uterine leiomyomas depends on such factors as the severity of symptoms, the size and location of the leiomyoma, and the patient's age. Small myomas that cause no health problems are generally monitored carefully for growth patterns. Large or multiple masses that promote severe uterine bleeding or interfere with functioning of the gastrointestinal or urinary tract are surgically removed, and hysterectomy may be indicated.

Ovarian Cysts

Ovarian cysts are sacs on an ovary that contain fluid or semisolid material. Ovarian cysts can develop at any time between puberty and menopause, including during pregnancy.

Etiology. The cause of the formation of ovarian cysts is presently unknown. They can arise in several locations in the ovaries:

1. Follicular cysts result when a maturing ovarian follicle fails to release an ovum; instead, the follicle continues to enlarge and produce estrogen.
2. Corpus luteum cysts occur when the corpus luteum fails to degenerate normally; the cyst continues to grow and produce progesterone.
3. Theca-lutein cysts are commonly bilateral and filled with clear, straw-colored fluid. Often their development is associated with hydatidiform mole, choriocarcinoma, or hormone therapy.

Clinical manifestations and treatment. Normally, ovarian cysts produce no symptoms. They may be noted on periodic examination and may increase and decrease in size with the menstrual cycle. Asymptomatic simple ovarian cysts smaller than 10 cm in diameter have a low probability of malignancy and can be followed without intervention. However, when a larger ovarian cyst ruptures, an ovarian vessel may tear, with variable amounts of intraperitoneal hemorrhage and abdominal

pain. In persons with recurrent ovarian cysts, oral contraceptives may be recommended to prevent ovulation. Oral contraceptives do not promote resolution of cysts that are already formed. Occasionally, immediate surgical intervention is indicated to control the hemorrhage and repair the site of rupture.

Endometriosis

Endometriosis is the presence of endometrial tissue outside the lining of the uterine cavity. Because the only normal location for endometrial tissue is the endometrial lining of the uterus, the presence of this abnormal growth is associated with a variety of side effects ranging from mild symptoms to life-threatening consequences. These foci of abnormal endometrial tissue are called *endometriomas,* or endometrial implants, and usually occur within the pelvis. The most common sites of occurrence of endometriosis within the pelvis are the ovary, peritoneum of the cul-de-sac or pouch of Douglas, uterosacral ligaments, round ligament, oviduct, and the peritoneal surface of the uterus. Less frequently, endometrial implants occur in other body sites such as the bladder or large intestine. Although endometriosis is a benign disease, it possesses certain characteristics of malignant disease, such as the ability to grow, infiltrate, and spread. Symptoms of endometriosis may have an abrupt onset or may develop over many years.

The actual incidence of endometriosis is unknown because it can exist without any significant symptoms. Conservative population estimates show that approximately 11% of women have undiagnosed endometriosis. Active endometriosis usually occurs between 30 and 40 years of age, particularly in women who have never given birth. Endometriosis is rare in women younger than 20 years or after menopause. Although some authorities report a higher incidence of endometriosis in white women of higher socioeconomic levels, these impressions may not be accurate given the tendency of this group to delay childbearing and to have enhanced access to health care. The infertility rate for women in whom endometriosis is diagnosed is about 30%.

Etiology. At the present time, three major theories on the etiology of endometriosis have been proposed:

- ***Transportation.*** Endometrial tissue flows backward through the oviducts during a normal menstrual period. After this retrograde flow, endometrial fragments implant on the ovary, peritoneal surfaces, and other areas.
- ***Metaplasia.*** Inflammation or a hormonal change triggers metaplasia (conversion of one kind of tissue to a form that is not normal for that tissue). Thus coelomic epithelium at certain sites converts to endometrial epithelium.
- ***Induction.*** In this theory, a combination of transportation and metaplasia takes place, and regurgitated endometrium chemically induces mesenchyma to form endometrial epithelium. (At present, this theory is thought to be the most likely explanation for endometriosis.)

Once the endometrial implants arise in their abnormal locations, they continue to be under hormonal influence, just as the endometrial lining of the uterus responds to hormonal influence. Thus they periodically proliferate and bleed in response to hormonal stimulation. In some instances they may rupture, usually immediately before or after a menstrual period. Endometriomas are filled with brown blood debris; when they rupture, their contents spill onto the sensitive pelvic peritoneum. This irritative discharge establishes a local chemical peritonitis, followed by the formation of fibrous tissue in the injured location. Dense tissue adhesions in the pelvis may result as the pelvic peritoneum undergoes repeated irritation by the cyclic activities of the endometrial implants.

Clinical manifestations. The most prominent symptom of endometriosis is acquired dysmenorrhea, which produces pain in the lower part of the abdomen and in the vagina, posterior of the pelvis, and back. The pain usually begins 5 to 7 days before the peak of menses and lasts for 2 to 3 days. It differs from the pain of primary dysmenorrhea, which is more cramplike and concentrated in the abdominal midline. Pain may be extremely severe, although the degree of pain does not necessarily indicate the extent of disease. Dyspareunia and pain with defecation may also be present. Significant changes in the pattern of menstrual flow may occur, with excessive bleeding that may progress to anemia and fatigue.

Treatment. Treatment varies according to the extent of disease, and the primary goals of therapy are to relieve pain symptoms and restore or maintain fertility. Many women with endometrial implants never experience symptoms and require no treatment; others experience a rapidly progressive set of severe symptoms requiring immediate intervention. Both medical and surgical treatment modalities may be used. Therapies to reduce endometriosis-associated pain symptoms include use of nonsteroidal antiinflammatory drugs (NSAIDs) and hormonal agents, including progestins, androgenic agents, and gonadotropin-releasing hormone (GnRH). Because endometriosis responds to cyclic hormonal functioning, it is thought that the use of hormones to interrupt this cyclic pattern may result in atrophy of the endometrial implants. Surgical intervention includes removal or destruction of the endometriosis. If damage to the pelvic organs is widespread and the disease is progressing rapidly, total abdominal hysterectomy with removal of the oviducts and ovaries is performed.

KEY POINTS

- Benign fibroid tumors, or leiomyomas, are the most common uterine tumor, affecting about 20% of women older than 35 years. Depending on their size, uterine leiomyomas may be characterized by abnormal vaginal bleeding, pelvic pain, constipation, and urinary frequency.
- Ovarian cysts are usually asymptomatic and may change in size with the menstrual cycle. Rupture of an ovarian cyst may result in severe abdominal pain and hemorrhage, which occasionally necessitates immediate surgical intervention.
- Endometriosis occurs when endometrial tissue grows in areas other than the uterine lining. Endometriosis may involve the ovary, peritoneum, oviduct, outer layer of the uterus, bladder, and intestine. Although considered benign, endometriosis tends to infiltrate and spread to adjacent tissues. Endometriosis may be initiated by reflux of the uterine lining through the oviducts into the abdominal cavity during menses.
- Ectopic endometrial tissues periodically proliferate and bleed in response to fluctuations in the levels of reproductive hormones. Dysmenorrhea, with pelvic, back, and lower abdominal pain, usually begins 5 to 7 days before the peak of menses and lasts 2 to 3 days. The pain is more diffuse than that of primary dysmenorrhea. Treatment may include induction of a menopause-like state with hormone administration or the surgical excision of affected structures.

CANCER OF THE FEMALE GENITAL STRUCTURES

Malignant neoplasms occur in every part of the female reproductive system. This section describes the incidence and pathophysiologic aspects of the most common types of malignancies in female genital structures. For further information about the process of neoplasm development, the reader may wish to refer to Chapter 7 of this text.

Cancer of the Cervix

Etiology. Cancer of the uterine cervix is a neoplasm that can be detected in the early, curable stage by the Papanicolaou (Pap) test. The

main cause of cervical cancer is certain HPV types. Other factors include having intercourse at a young age or with multiple sexual partners, becoming pregnant multiple times, or being infected by sexually transmitted infections. The incidence of cervical cancer is decreasing in the United States; approximately 12,000 new cases of HPV-related cervical cancer are diagnosed each year and over 4000 women in the United States die annually from cervical cancer. Widespread screening with a yearly Pap test in women at risk has continued to decrease the mortality of cervical cancer. The American Cancer Society now recommends that all women should begin cervical cancer screening at 21 years of age. After three consecutive negative Pap tests, women older than 30 years who are not at high risk can be tested every 2 to 3 years. For women who have had a hysterectomy unrelated to cervical neoplasia, the CDC no longer recommends Pap screening unless surgery was done as treatment for cervical cancer or precancer. Low-risk women who have been screened regularly may also stop screening at age 70.

Clinical manifestations. Preinvasive cervical cancer produces no symptoms, although the Pap test can detect changes in cells of the cervical epithelium, which may be present for 10 years before invasive cancer develops. Early invasive cancer causes abnormal vaginal bleeding, persistent vaginal discharge, and pain and bleeding after intercourse. When symptoms appear, the cancer has usually progressed beyond its early stages. Squamous cell carcinoma accounts for 95% of all invasive cervical cancers diagnosed, and adenocarcinomas account for most of the rest. Invasive carcinoma of the cervix spreads by direct extension to the vaginal wall, laterally into the parametrium toward the pelvic wall, and anteroposteriorly into the bladder and rectum. Metastasis to the pelvic lymph nodes is more common than spread to distant lymph nodes.

Treatment. The treatment strategy depends on the clinical stage of the tumor at the time of diagnosis. Surgery—including cryotherapy, excision, and laser surgery for precancerous conditions and hysterectomy for invasive carcinoma—may be indicated. Chemotherapy and radiation therapy may be used in invasive disease. Radical surgery, including *pelvic exenteration,* or removal of all the pelvic organs, can now be performed with limited morbidity. Treatment works best at the early stages of cancer. The 5-year survival rate for stage I cervical cancer is 93%. The 5-year survival rate for stage IV cancer is only 15%.

Prevention. In 2006 the Food and Drug Administration licensed the first vaccine developed to prevent cervical cancer and other diseases in females caused by certain types of HPV. The quadrivalent vaccine Gardasil protects against four HPV types (6, 11, 16, 18) that are responsible for 70% of cervical cancers and 90% of genital warts. A bivalent vaccine, Cervarix, is also available and protects against HPV types 16 and 18. Most recently, a nine-valent form of the vaccine has become available (Gardasil 9). The HPV vaccine is recommended for 11- to 12-year-old girls, but can be administered to girls as young as 9 years of age. The vaccine also is recommended for 13- to 26-year-old females who have not yet received or completed the vaccine series. Ideally, the vaccine should be administered before onset of sexual activity. However, females who are sexually active also may benefit from vaccination. Experts also suggest similar vaccination recommendations for males.

Endometrial Cancer

Cancer of the endometrial lining of the uterus is less common than cervical cancer in young women, but both types of cancer occur with equal frequency in postmenopausal women. Related factors include infertility, late menopause (older than 55 years), obesity, diabetes, and hypertension. Unopposed estrogen therapy also increases the frequency.

Clinical manifestations and treatment. The most common initial symptom is bleeding between menstrual periods or postmenopausal bleeding. The diagnosis of endometrial cancer is based on histologic tissue examination. Treatment strategies for endometrial cancer include radiation therapy and total hysterectomy with possible removal of the ovaries and oviducts. The 5-year survival rate for patients in whom endometrial cancer (adenocarcinoma) is diagnosed early at a local stage is approximately 90%. The 5-year survival rate drops to 17% if the cancer has metastasized before diagnosis.

Ovarian Cancer

Ovarian cancer has replaced cervical cancer as the leading cause of death from genital cancer. The peak incidence is between 60 and 80 years of age. Because no symptoms are noted until late in the disease, the mortality rate is high, with only a 45% 5-year survival rate.

Clinical manifestations and treatment. When symptoms occur, they are related to intraabdominal metastasis and include increasing abdominal girth, weight loss, abdominal pain, dysuria or urinary frequency, and constipation. Management of ovarian cancer includes removal of the uterus, ovaries, and oviducts. Radiation therapy and chemotherapy may be used in conjunction with surgery. Increasingly, prophylactic oophorectomy or salpingo-oophorectomy is being recommended in high-risk women, particularly in women who are carriers of the *BRCA1* and *BRCA2* mutations.

Vaginal Cancer

Cancer of the vagina generally occurs in women in their early to mid-fifties, although it has an increased incidence in young women whose mothers took diethylstilbestrol during pregnancy. Because the vagina is a thin-walled structure with rich lymphatic drainage, vaginal cancer may metastasize to the bladder, rectum, vulva, pubic bone, and other surrounding structures.

Clinical manifestations and treatment. The primary signs and symptoms of vaginal cancer are vaginal spotting and discharge, pain, groin masses, and changes in urinary pattern. Early-stage therapy is designed to treat the malignant area while preserving normal parts of the vagina. Radiation therapy or surgery varies based on the size, depth, and location of the tumor. Preservation of a functional vagina is generally possible only in the early stages, although grafting from other body sites may be performed to avoid vaginal stenosis, particularly in younger women.

Cancer of the Vulva

Clinical manifestations and treatment. Cancer of the vulva is rare, with a crude mortality rate of 1.3 per 100,000 women. It can occur at any age, including infancy, but has a peak incidence in the mid-sixties. Factors that seem to predispose to the disease include sexually transmitted infections, chronic pruritus of the vulva with swelling and dryness, obesity, hypertension, diabetes, and never having been pregnant.

Leukoplakic changes (the presence of whitish plaquelike or ulcerated lesions) in the vulva may precede the development of carcinoma. Once the carcinoma develops, vulvar masses may be present, with groin masses and abnormal urination and defecation manifesting later in the disease. Management of vulvar cancer includes partial excision of the vulva to remove precancerous leukoplakic lesions and total vulvar excision for advanced disease. Local relapse is common whether conservative or radical procedures are undertaken.

KEY POINTS

- Cervical cancer may be detected by evaluation of cervical cells (Pap test). Early-stage cervical cancer may be asymptomatic. When they appear, symptoms include abnormal vaginal bleeding and discharge. Cervical cancer may spread to the vaginal wall, pelvis, bladder, rectum, and pelvic lymph

nodes. The quadrivalent vaccine Gardasil protects against four HPV types (6, 11, 16, 18), which are responsible for 70% of cervical cancers and 90% of genital warts. A nine-valent vaccine is now available and recommended. The vaccine is recommended for both young women and men.

- Other cancers of the female reproductive tract include endometrial, ovarian, vaginal, and vulvar cancers. No routine screening tests are available for these diseases. Ovarian cancer has a high mortality rate because it is usually diagnosed after it has metastasized.

DISORDERS OF PREGNANCY

Pregnancy results in a number of physiologic alterations in the mother that are usually well tolerated, particularly if adequate prenatal care is available. However, pregnancy can result in a number of conditions that may be life threatening to the mother and the developing fetus. The most common pregnancy-related disorders are described here; in addition, for information concerning diabetes in pregnancy, the reader may wish to consult Chapter 41, which covers the topic of diabetes in depth.

Pregnancy-Induced Hypertension

Pregnancy-induced hypertension (PIH) is known by other names such as *toxemia* and *preeclampsia-eclampsia.* Worldwide, 76,000 pregnant women die each year from preeclampsia and hypertension-related disorders; 500,000 babies die of these disorders annually. Hypertension complicates 5% to 8% of all births in the United States and is one of the leading causes of pregnancy-related deaths. PIH is characterized by a rapid rise in arterial blood pressure associated with the loss of large amounts of protein in the urine. Women at risk for the development of PIH include teenagers and women in their late thirties and early forties. In addition, the presence of multiple fetuses and the preexistence of hypertension, renal and cardiovascular disease, and diabetes may predispose to the development of PIH.

Etiology, clinical manifestations, and treatment. The exact causes of PIH are presently unknown, although poor nutrition and genetic and immunologic factors have been suggested. PIH is characterized by salt and water retention by the kidneys, weight gain, and edema. In addition, arterial spasm occurs in many parts of the body, most significantly in the kidneys, brain, and liver. Both renal flow and the glomerular filtration rate are decreased, a condition exactly opposite the normal changes in pregnancy. The renal effects are caused by thickening of the glomerular tufts, which contain a fibrinoid deposit in the basement membranes.

The severity of symptoms of PIH is closely related to the retention of salt and water and the degree of the increase in arterial pressure. The increasing arterial pressure seems to promote a vicious cycle in which arterial spasm and other pathologic effects give rise to further increases in arterial pressure. Milder forms of the disease are managed with bed rest. Fetal well-being is periodically assessed, and the infant is delivered if conditions deteriorate or maturity is achieved.

In its severe form, PIH is characterized by extreme vascular spasticity throughout the body, clonic convulsions followed by coma, renal failure, liver malfunction, and extreme hypertension. Usually, this severe form occurs shortly before parturition. The mortality rate in women with severe PIH who are left untreated is high. However, the immediate use of rapidly acting vasodilating drugs, seizure prophylaxis, and rapid delivery have reduced the mortality rate from PIH to less than 1%.

Hyperemesis Gravidarum

Hyperemesis gravidarum is a Latin term for excessive vomiting in pregnant women. Although transient nausea and vomiting occur in about half of women in the first trimester of pregnancy, in a few women these symptoms continue throughout the entire course of pregnancy. Intractable vomiting, or hyperemesis gravidarum, occurs in about 0.3% to 2% of pregnancies, sometimes with life-threatening consequences. Severe dehydration and electrolyte imbalance, hepatic and renal damage, encephalopathy, and ultimately death may ensue if the vomiting cannot be controlled.

Clinical manifestations and treatment. The causes of hyperemesis gravidarum are unknown, but it is thought that an abnormal response to the production of large amounts of human chorionic gonadotropin hormone by the placenta may be implicated. Intravenous therapy to correct metabolic and nutritional abnormalities, antiemetic agents, and supportive care in a hospital environment may be needed to resolve the symptoms.

Placenta Previa and Abruptio Placentae

Etiology and clinical manifestations. Placenta previa is a condition in which the placenta is implanted abnormally over the internal cervical os. Abruptio placentae is premature separation of the placenta before delivery of the fetus. Placenta previa occurs in approximately 1 in 200 deliveries and is more common in women with multiple pregnancies and previous cesarean section; its cause is unknown. Placenta previa may occur in varying degrees of severity ranging from partial to entire coverage of the internal cervical os. Abruptio placenta, or premature separation of the placenta, occurs after 20 weeks of gestation in about 1% of deliveries. The detachment may be partial or complete and may cause overt or concealed hemorrhage. Abruptio placentae can be caused by trauma, a short umbilical cord, occlusion of the inferior vena cava, PIH, or abnormal uterine anatomy.

Treatment. Therapeutic strategies for placenta previa and abruptio placentae include cesarean section for fetal distress or hemorrhage control. Medications designed to control preterm labor may also be administered.

Spontaneous Abortion

Spontaneous abortion is expulsion of the products of conception from the uterus before the period of fetal viability. It is usually called a *miscarriage* by laypersons, and it is differentiated from elective abortion. Although the precise incidence is unknown, it is estimated that 50% of all pregnancies end in spontaneous abortion. Among those women who know they are pregnant, the rate of spontaneous abortion is 15% to 20%.

Etiology. Abnormal development accounts for a large percentage of aborted pregnancies. Nearly 61% of abortuses expelled in the first trimester demonstrate chromosomal abnormalities. In addition, abnormal development may result from faulty implantation of the fertilized ovum or from an abnormality in the uterine environment. Maternal factors responsible for spontaneous abortion include both systemic and localized conditions. Infectious processes that may contribute to spontaneous abortion include cytomegalovirus, herpesvirus, and rubella infections. Abnormalities of the reproductive organs, immune disorders, endocrine malfunction, and physical and psychic trauma may all contribute to spontaneous abortion.

Clinical manifestations and treatment. Associated signs and symptoms of spontaneous abortion include vaginal bleeding and abdominal cramps. The cramps may intensify as the cervix dilates for expulsion of the uterine contents. If the entire contents are expelled, the bleeding and cramps subside. However, if any contents remain, an incomplete abortion has occurred, and intervention may be needed to control bleeding and to surgically remove the remaining uterine contents.

KEY POINTS

- Pregnancy induced hypertension (PIH) is characterized by a rapid rise in blood pressure and proteinuria. Renal blood flow and the glomerular filtration rate are reduced, and the kidneys retain salt and water. When severe, PIH may be associated with convulsions and coma. Antihypertensive therapy may be indicated.
- Excessive vomiting during pregnancy is termed *hyperemesis gravidarum.* Dehydration, electrolyte imbalance, hepatic and renal damage, and death may ensue.
- Placenta previa occurs when the placenta is implanted over the cervical os. Abruptio placentae is premature separation of the placenta. Both conditions may interrupt fetal oxygen supply and cause maternal hemorrhage. Cesarean section is indicated.
- It is estimated that 10% to 15% of known pregnancies end in spontaneous abortion. Fetal abnormalities, faulty implantation, infections, and trauma increase the risk of spontaneous abortion.

FIG 33.8 Nipple retraction in the right breast as a result of mammary duct ectasia. (From Haagensen CD: *Diseases of the breast,* ed 3, Philadelphia, 1986, Saunders, p 359.)

DISORDERS OF THE BREAST

The breast is considered an accessory organ of the female reproductive tract and is affected by many of the same factors that promote alterations in the other reproductive organs. Women's breast health has become a critical concern in the United States because the breast is the most common site of cancer in women between 25 and 75 years of age. In addition, women are playing an increasingly important role in recognizing the symptoms of breast disease and are seeking earlier intervention with improved outcomes. It is essential that health care professionals continue to encourage this enhanced role and provide accurate information about breast health to their clients. This section includes information on specific breast disorders involving reactive-inflammatory breast disorders, benign breast disorders, and carcinoma of the breast. Before reading this information, the reader may wish to review the section on the structure and function of the breast in Chapter 32 of this text and the specific information on neoplasm development in Chapter 7.

Reactive-Inflammatory Breast Disorders

Breast disorders in which an inflammatory response occurs in reaction to irritation, injury, or infection include mammary duct ectasia, breast abscess, fat necrosis, and reactions to injections or implantation of foreign materials in the breast.

Mammary Duct Ectasia

Mammary duct ectasia is a chronic inflammatory process occurring in and around the terminal subareolar ducts of the breast (it is also referred to as *periductal mastitis*). It is more prevalent in older women, primarily postmenopausal women. The Latin word ectasia means dilation, and in mammary duct ectasia the collecting ducts beneath the nipple and areola become dilated, thinned, and filled with secretions.

Pathogenesis. Over time, the ducts become distended with cellular debris, and the debris begins to have an irritating effect on the duct walls. The inflammatory response is initiated, and a zone of granulation tissue is created around a small cavity filled with thick yellowish or brownish material. This area will be palpable as a mass in the central area of the breast, beneath or near the areola. By the time the duct ectasia has grown into a palpable mass, a reactive fibrosis will also have formed in the tissue around the mass. This fibrous thickening of the surrounding breast tissue causes dimpling and distortion of the breast and nipple inversion (Fig. 33.8). However, a congenital inverted nipple is already present in some women with mammary duct ectasia, and it is thought that the presence of this nipple anomaly may in some way contribute to ductal wall irritation.

Clinical manifestations and treatment. In addition to a palpable mass and dimpling or distortion of the breast or areola, women with mammary duct ectasia may have a persistent nipple discharge. These signs must be evaluated carefully because they may also be indicative of a malignant breast mass. A biopsy is usually performed to rule out the presence of a malignancy. After confirmation of the diagnosis of mammary ductal ectasia, surgical excision of the dilated subareolar ducts is performed.

Breast Abscess

The majority of abscesses occurring in the breast are not associated with breast feeding and are referred to as *nonlactational* breast abscesses (for a complete description of abscesses or mastitis related to lactation, the reader may wish to refer to an obstetric or maternity nursing text). Nonlactational breast abscesses are most often a recurring problem and usually affect persons with conditions that predispose to infections, such as having diabetes mellitus, undergoing steroid therapy, or being afflicted with other skin lesions.

Etiology. Multiple factors may contribute to the formation of nonlactational breast abscesses. In some women, the presence of a congenital inverted nipple may predispose to abscess formation. Abscesses may also be part of the syndrome of mammary duct ectasia; in addition, women with the aforementioned preexisting conditions that predispose to infections may be at increased risk for the development of an infectious process in the breast tissue. Unlike breast abscesses occurring during breast feeding, in which *Staphylococcus aureus* is the most common causative organism, nonlactational breast abscesses usually yield multiple organisms when cultured.

Clinical manifestations and treatment. Signs and symptoms of these abscesses include an area of tenderness, redness, and induration under the periareolar skin. Unfortunately, nonlactational breast abscesses do not respond well to antibiotic therapy and often recur, and it is sometimes necessary to excise the major duct system beneath the areola to prevent further recurrence.

Fat Necrosis

Necrosis refers to the death of a portion of tissue, and fat necrosis in the breast is the death of fat tissue after trauma or injury to the breast.

The position of the breasts makes them vulnerable to trauma, particularly in larger women with pendulous breasts. This phenomenon is important for health care professionals to assess because fat necrosis may mimic or obscure carcinoma of the breast.

Clinical manifestations and diagnoses. Fat necrosis of the breast may have many of the same clinical signs as breast malignancy, including a painless mass in the breast that is firm, ill defined, and poorly mobile. Skin thickening and retraction may also be present. In addition, a mammogram may not provide a clear diagnosis. Unfortunately, many women with pendulous breasts frequently sustain injuries to the breast and may be unable to recall any specific trauma; thus a diagnosis of fat necrosis may be difficult to make. If fat necrosis cannot be reliably distinguished from carcinoma based on clinical observation or mammography, excisional biopsy must be performed.

Reactions to Foreign Material

Surgery to enlarge the female breast has become one of the most popular of all cosmetic surgical procedures in recent years. Since the early twentieth century, a variety of materials have been used for breast augmentation. Silicone implants, which consist of silicone gel encased in polyurethane or other materials, have been the most widely used devices for breast enlargement and have been implanted in more than 1 million women. At present, controversy surrounds the use of silicone breast implants because some side effects, including irritation at the implantation area and other symptoms suggestive of an immune system response, have been reported. Currently, the recommendations surrounding silicone breast implants are conflicting. The use of silicone implants for routine cosmetic breast augmentation is specifically controlled in the United States, favoring implants filled with a saline solution. Health care professionals should be aware of the reported side effects of silicone breast implants inasmuch as a substantial segment of the female population in the United States and Western Europe has undergone breast augmentation with these devices. In addition, persons with silicone implants who sustain blunt trauma to the chest are at risk for rupture of the implant, with subsequent leakage of the silicone gel into surrounding tissue. After chest trauma, the communication of information regarding the presence of silicone breast implants to other health care professionals is an important consideration in planning care and preventing further tissue exposure to silicone.

Benign Breast Disorders

The term *benign breast disorders* encompasses a group of lesions affecting the breast. These disorders are usually divided into two categories: (1) fibrocystic breast disease and (2) specific benign neoplasms of the breast such as fibroadenomas, adenomas, and papillomas. It is important for health care professionals to understand the clinical significance of these benign disorders. Although these entities are "benign" in the sense of being differentiated from malignant breast neoplasms, clients experiencing them may be at risk for experiencing a psychological crisis and may need to be educated regarding their potential risk for breast malignancy.

Fibrocystic Breast Disease

Although the term *fibrocystic breast disease* is frequently used by health care professionals, it is important to understand that it is not a distinct disease entity. Instead, it is a diagnosis classification that is applied to a condition in which the presence of palpable breast masses fluctuates with the menstrual cycle and may be associated with pain or tenderness. Laboratory examination of this breast tissue shows macroscopic and microscopic cysts, along with a variety of alterations in tissue structure such as fibrosis or overgrowth of stromal fibrous tissue. However, these alterations in breast tissue are present to some degree in all

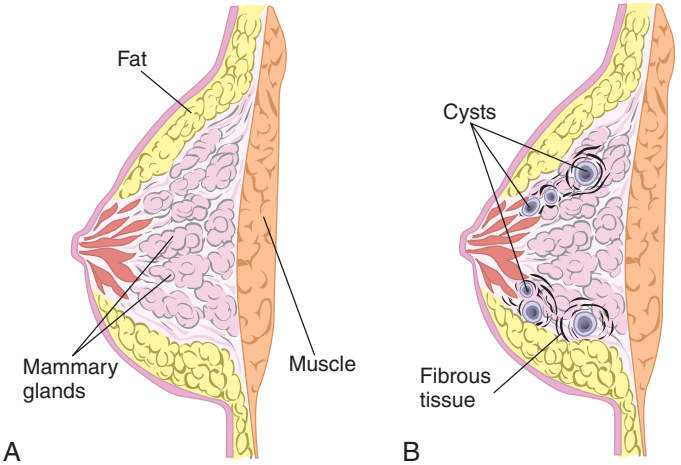


FIG 33.9 A, Normal breast. **B,** Fibrocystic breast tissue. Note image of cysts showing typical smooth margins, dark center, edge shadows, and a bright posterior wall. (From Lewis SL et al, editors: *Medical-surgical nursing: assessment and management of clinical problems*, ed 10, St Louis, 2017, Elsevier.)

female breasts, which has led some authorities to question use of the term "disease" for such a widespread condition. Until a more precise system for classifying this type of benign breast disorder is widely adopted, fibrocystic breast disease will probably continue to be used to describe this phenomenon of tender breast masses that occur on a cyclic basis. A comparison of normal and fibrocystic breast tissue is shown in Fig. 33.9.

Etiology and clinical manifestations. Hormonal imbalance in the reproductive years is thought to contribute to fibrocystic breast disease. Fibrocystic breast disease is more common in women ages 30 to 50 years. It is usually characterized by tenderness or pain in one or both breasts immediately before onset of the menstrual period. On palpation, the cysts tend to be firm, regular in shape, and mobile. They are located most often in the upper outer quadrant of the breasts, and their size may fluctuate throughout the menstrual cycle.

Although it was previously thought that all women with fibrocystic breast disease were at increased risk for breast cancer, recent research has disproved this theory. It is now known that only certain types of tissue changes may predispose a woman with fibrocystic breast disease to the development of breast malignancy. The vast majority of women with fibrocystic disease do not have these alterations in breast tissue and therefore are not at a substantially increased risk for breast cancer.

Diagnoses and treatment. Diagnostic studies can include ultrasound and needle aspiration of a cyst for histologic analysis. Oral contraceptives have been recommended to control symptoms of fibrocystic breast disease. NSAIDs may be helpful for breast discomfort. Other supportive measures include the application of local heat and use of a support bra. Nutritional therapies have shown success in some women, particularly avoidance of foods with methylxanthines, such as tea, coffee, cola, and chocolate. It is thought that methylxanthines tend to stimulate cyclic adenosine monophosphate and thus increase metabolic activity in the breast. A low-fat, high-carbohydrate diet has been shown to decrease breast swelling and tenderness.

Specific Benign Neoplasms

Specific benign neoplasms of the breast, such as fibroadenomas, adenomas, and papillomas, may occur at any time during a woman's life from childhood through old age. These neoplasms behave in a clinically "benign" fashion; that is, they do not invade the surrounding tissue or

metastasize to other sites. They generally appear as freely movable, encapsulated masses that are sharply delineated from the surrounding breast tissue. However, it is important to have any breast mass evaluated because biopsy and histologic examination may be needed to differentiate these benign neoplasms from breast carcinoma.

Malignant Disorder of the Breast

Cancer of the Breast

Carcinoma of the breast remains the most common form of cancer in women between the ages of 25 and 75 years. In the United States it is the second leading cause of cancer mortality in women. The incidence of breast carcinoma appears to be increasing in the United States, with an estimated 231,840 newly diagnosed cases per year. Although the disease is more common in white women, its incidence in blacks and Asians is rising. Breast cancer does occur in males, but is 100 times less common. Even though recent advances in early detection and treatment have afforded longer survival after diagnosis, invasive breast carcinoma remains an incurable disease that continues to take the lives of a large segment of the population.

Etiology. A substantial number of studies conducted in the past 30 years have begun to establish the risk factors and possible causes of breast cancer. Some factors that may place a woman at risk for breast cancer include hormonal influences, reproductive factors, dietary factors, family history, age, radiation exposure, history of cancer, and lack of access to health care. It should be noted that helping a client understand and interpret her personal breast cancer risk is a difficult task for a health care professional. The public media have given much attention to some of the risk factors for breast cancer, but have not provided much context in which to interpret evaluations for individual risk factors.

Risk factors are characteristics related to the probability of a certain outcome—in this case, breast cancer. These risk factors may be either causally or correlatively associated with an outcome. For example, a factor may directly *cause* an outcome (as the smallpox virus causes smallpox) or may be *correlated* with an outcome (as not wearing a seat belt is correlated with an increased degree of injury in a motor vehicle accident). The distinction between causality and correlation is an important concept to impart to clients when discussing risk factors. A client may express concern, for example, that a certain risk factor will directly cause the development of breast cancer. The ability of a health care professional to describe and discuss risk factors in a knowledgeable way will greatly enhance the client's ability to make decisions regarding such issues as hormonal replacement therapy after menopause.

Hormonal factors. Several hormonal factors have been shown to be linked to the development of breast cancer. Length of exposure to the hormones secreted by the ovary (estrogen and progesterone) has been shown to affect the risk for breast cancer in the following way: If a woman has had an early (younger than 12 years) onset of menses and a late (older than 55 years) menopause, her risk is increased. Stated another way, women with 40 or more years of menstrual activity have twice the breast cancer risk as women with fewer than 30 years of menstrual activity. Postmenopausal hormone replacement therapy may increase the risk of breast cancer; data suggest clearly that there may be risk for women taking estrogen/progesterone combination therapies. For some women, the known benefits of these medications may outweigh effects on cancer risk. Future research is needed to clarify the way in which hormonal exposure may foster breast cancer development and the many interactive factors associated with taking hormonal medications.

Reproductive factors. It has been observed in many research studies that giving birth at a young age (less than 18 years) is associated with a decreased risk of breast cancer and that giving birth for the first time at 35 years or older increases the risk. In addition, parity (the number of viable children a woman has borne) has been associated with risk, with low parity increasing risk and high parity having a protective effect.

Dietary factors. It has been suggested that the amount of fat in the diet is a risk factor for breast cancer. Researchers who favor this theory point to the relatively low rates of dietary fat ingestion in countries with low rates of breast cancer. Although the media have given a great deal of attention to this issue, scientific data have been inconclusive thus far. Countries in which low-fat diets are widespread are typically nonindustrialized countries in which other factors, such as age at first delivery or parity, differ from those in industrialized countries. No single dietary pattern or food has been shown to "cause" cancer, just as no specific food has been shown to prevent or cure cancer.

Family history. The role of heredity in contributing to breast cancer has long been recognized. Specific gene mutations such as *BRCA1* and *BRCA2* have been identified in high-risk families. Research studies have indicated that women with a mother or sister with breast cancer have an increased risk of developing breast cancer, even if specific gene mutations are not identified. Women with family risk factors and gene mutations need careful care and accurate information in order to make informed choices. Women with *BRCA1* and *BRCA2* gene mutations are at risk for both breast cancer and ovarian cancer and may benefit from prophylactic salpingo-oophorectomy to reduce their risk. Another controversial option for some high-risk women is bilateral mastectomy. Finally, chemoprevention with selective estrogen receptor modulators (SERMs) may be an option for some high-risk women. SERMs may not be recommended in low-risk to average-risk women because they have their own adverse effects such as thromboembolic events and endometrial cancer.

Age. Breast cancer is extremely rare in young women. The incidence begins to increase by 25 to 30 years of age and continues to increase with advancing age.

Other factors. Other factors, such as radiation exposure and a history of cancer, have been shown to be risk factors for the development of breast cancer. Several potential factors have been suggested, such as exposure to low-frequency electric or magnetic fields and a virus transmitted through lactation. More research is required to establish the role of these potential factors.

Clinical manifestations. Many breast cancers are discovered by the woman herself through self breast examination. She usually finds a single lump that is painless, hard, and poorly movable. Yearly clinical breast examinations by trained clinicians are also recommended after the age of 40 years. Digital mammography is an important clinical tool, and according to current screening guidelines, most women should have yearly mammograms beginning at age 40. Other technologies, such as magnetic resonance imaging (MRI), are showing promise as screening tools. Improvements in technology, such as digital mammography and MRI, have greatly increased the ability to identify breast cancers.

Half of malignant tumors occur in the upper outer quadrant of the breast. Other signs of advanced tumor development include dimpling of the skin (Fig. 33.10), retraction of the nipple, changes in breast contour, and bloody discharge from the nipple. Breast cancer is diagnosed by a number of techniques that use films (mammography, xerography) (Figs. 33.11 and 33.12), by computerized technologies (digital mammography), and by thermography, a technique in which "hot spots" indicate increased metabolic activity. A person of any age with a suspected breast mass should undergo mammography and biopsy.

Most breast carcinomas arise in the epithelium of the glandular ducts of the breast. The lesion(s) have infiltrating edges that begin to invade normal breast tissue (Fig. 33.13). After this invasion, malignant cells begin to scatter or disseminate into the lymph system of the axilla

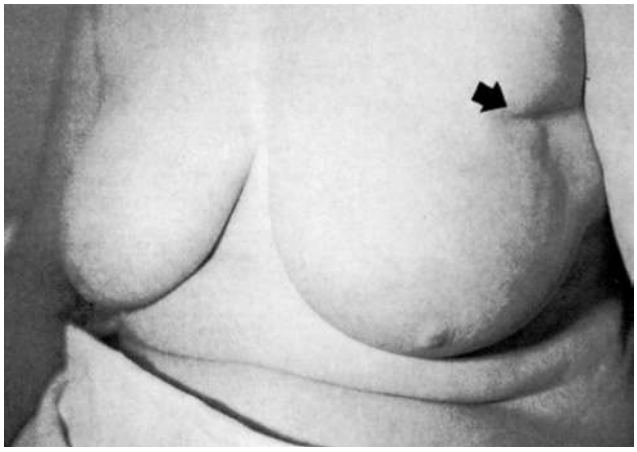


FIG 33.10 Skin dimpling caused by an underlying malignant tumor. (From Donegan WL, Spratt JS: *Cancer of the breast,* ed 5, Philadelphia, 2002, Saunders, p 321.)

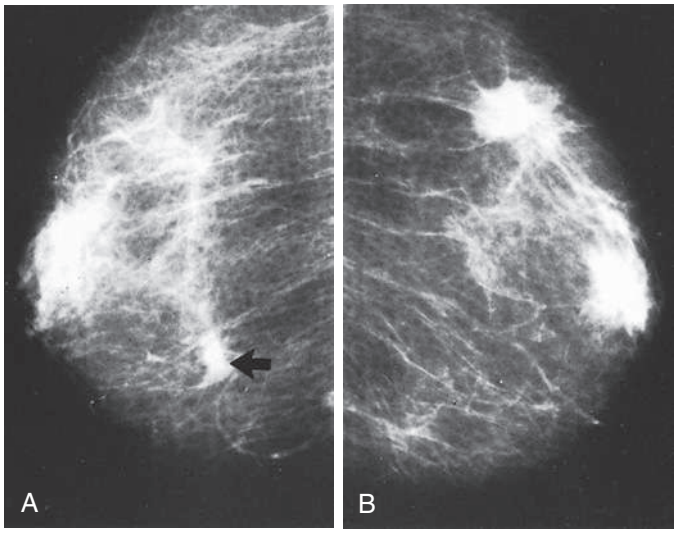


FIG 33.12 Mammogram showing bilateral invasive ductal carcinoma. **A,** Left breast. The larger mass was palpable. The smaller right mass was not palpable *(arrow).* **B,** Right breast. Multiple masses are shown. (From Powell DE, Stilling CB: *Diagnosis and detection of breast diseases,* St Louis, 1993, Mosby.)

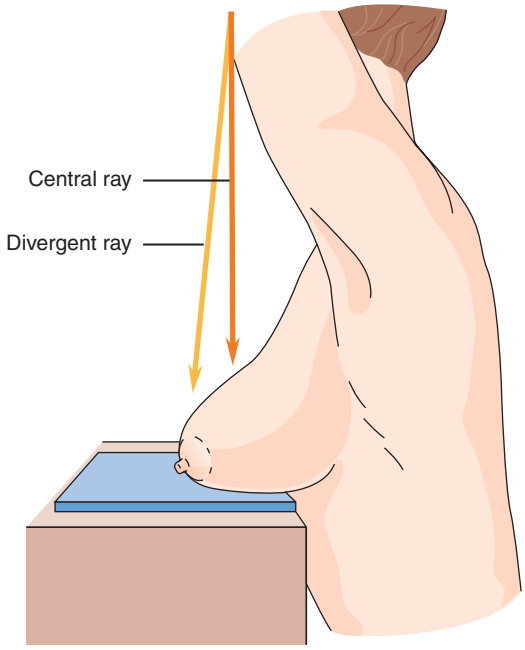


FIG 33.11 Placement of the breast for mammography, along with the direction of the x-rays.

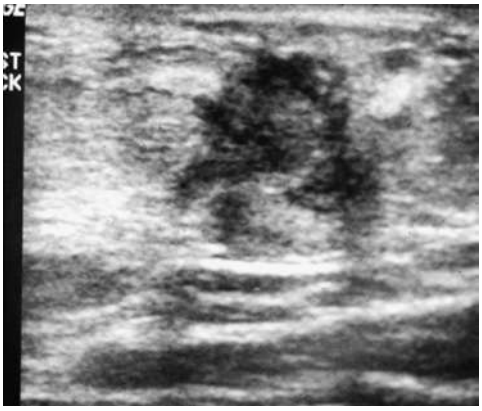


FIG 33.13 Ultrasound scan of a carcinoma. Note the ragged appearance of this invasive, malignant lesion. (From Donegan WL, Spratt JS: *Cancer of the breast,* ed 5, Philadelphia, 2002, Saunders, p 332.)

(Fig. 33.14). The breast is in close proximity to the large system of axillary lymph nodes, which makes easy dissemination of malignant cells possible. The major way by which breast carcinoma causes morbidity and death is through the dissemination of malignant cells to other body sites, most commonly lung, liver, and bone. Metastasis (or spread of carcinoma) to these other body sites signifies a poorer prognosis. The prognosis is vastly better for persons with no evidence of spread of malignant cells to the regional lymph nodes. The 5-year survival rate is 99% when no lymph node involvement is found, but averages 85% when lymph node involvement is present. For women with distant spread (metastases), the rate drops to 25%. The greater the number of positive lymph nodes (nodes with malignant cells) found at surgery, the less favorable the prognosis.

Treatment. Treatment for breast cancer includes surgery, chemotherapy, radiation therapy, and supportive measures. Surgical therapy is a controversial area, and various options are available. Breast-conserving therapy, which includes removal of only the lesion, is called a *lumpectomy.* Lumpectomy or lumpectomy in conjunction with either chemotherapy or/and radiation is becoming increasingly preferred. Removal of only the breast is called a simple mastectomy**.** Other surgical interventions include a modified radical mastectomy**,** in which the breast is removed and a portion of the axillary lymphatic system is dissected, and a radical mastectomy (rare), in which the breast, lymphatic drainage, and underlying pectoral muscles are removed.

Chemotherapy entailing a variety of hormonal and antineoplastic agents is also used. Malignant cells may have cytoplasmic hormone receptors that bind to hormone molecules and promote cellular division and growth. SERMs, estrogen antagonists, are the most common agents used. Popular examples of SERMs include tamoxifen and raloxifene. Antineoplastic agents are given to control the spread of malignant cells.

Radiation therapy may be used as an adjunct to the aforementioned therapy and to control pain by shrinking large tumor masses. Other

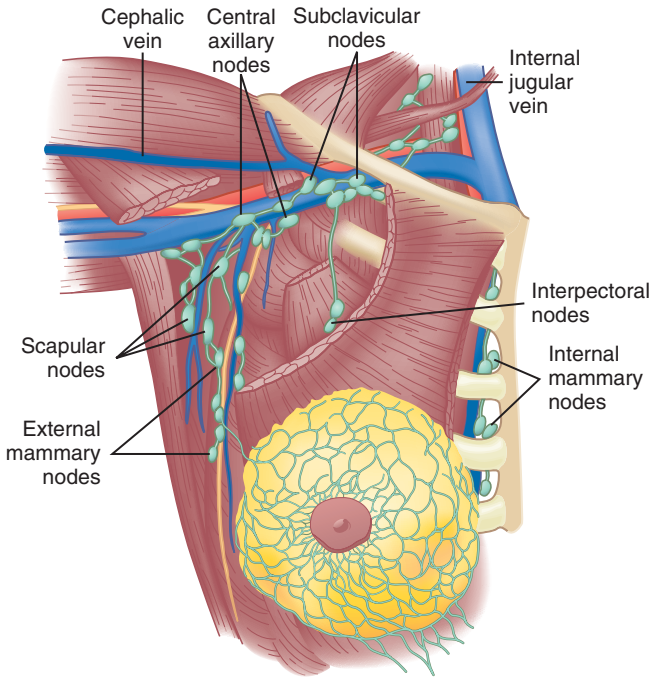


FIG 33.14 Lymphatic drainage of the breast. In general, lateral lesions in the breast metastasize to axillary and supraclavicular nodes, whereas medial tumors tend to metastasize to the internal mammary and mediastinal lymph nodes, as well as the supraclavicular nodes. (From Lewis SL et al, editors: *Medical-surgical nursing: assessment and management of clinical problems*, ed 10, St Louis, 2017, Elsevier.)

supportive measures in advanced disease include operations to reduce the bulk of tumors.

Continuum of care. Breast cancer is characterized by a wide variation in clinical course. Many patients who undergo therapy for breast carcinoma are able to achieve a satisfying quality of life. Educational and support programs for breast cancer patients and their families, both preoperatively and postoperatively, have been an important means of providing emotional support. Programs for continuing care after mastectomy have helped patients and families face the adaptive challenges of living with breast cancer. Follow-up care includes early detection of recurrent disease, with an emphasis on breast self-examination, yearly mammography, and regular examination by health care professionals.

KEY POINTS

- Chronic inflammation of the subareolar ducts may result in mammary duct ectasia. Fibrous thickening results in a palpable central mass, breast distortion and dimpling, and nipple inversion. Persistent nipple discharge may occur. These signs are similar to those of malignancy and are carefully evaluated by biopsy. Surgical excision may be performed.
- Breast abscesses in nonlactating women are commonly associated with chronic infection, diabetes, and steroid therapy. These abscesses respond poorly to antibiotics and tend to recur.
- Fibrocystic breast disease is a condition in which palpable breast masses are present and fluctuate with the menstrual cycle. Breast cysts are firm, mobile, and tender, and are usually located in the upper outer quadrant. There is no evidence that women with fibrocystic breasts are at higher risk for breast cancer. A low-fat, high-carbohydrate diet; danazol; heat therapy; and avoidance of methylxanthines may be recommended.
- Breast cancer is a common cancer in women between 25 and 75 years of age. Malignant tumors tend to be painless, hard, and fixed in place, in contrast to benign breast tumors, which are mobile and encapsulated. Risk factors for breast cancer include a first-degree relative with breast cancer, increasing age, radiation exposure, and a previous malignancy. In addition, reproductive factors such as the age at first pregnancy and the number of pregnancies may be associated with altered cancer risk.
- Breast cancer may spread to the regional lymphatics and disseminate to other sites. Localized breast cancer, without lymph node involvement, has a 99% 5-year survival rate. The survival rate falls to 85% when lymph nodes are cancerous and to 25% when there are distant metastases. Depending on the extent of tumor spread, surgery may be performed to remove the tumor only (lumpectomy); the affected breast only (simple mastectomy); the affected breast and involved lymph nodes (modified radical mastectomy); or the breast, lymphatics, and underlying muscle (radical mastectomy). In addition, radiation therapy and chemotherapy may be initiated.

SUMMARY

This chapter has described the most prevalent women's reproductive health problems at the present time. Any alteration in reproductive status may have profound implications for the individual; thus the reader should review this material carefully to acquire the ability to distinguish the differences and similarities in these alterations.

Commonly occurring alterations in reproductive health for women may have serious consequences and require immediate intervention. Menstrual disorders may have multiple manifestations, and such disorders as amenorrhea and abnormal uterine bleeding may occur at any time throughout a woman's life. Alterations in uterine position and pelvic support, including uterine prolapse, retrodisplacement of the uterus, cystocele, and rectocele, may result in severe symptoms and require surgical correction. Inflammation and infection of the female reproductive tract, including PID and vulvovaginitis, may have far-reaching effects for the individual experiencing them.

The reader should pay particular attention to the section on benign growths and aberrant tissue of the female reproductive tract given the widespread nature and potentially serious consequences of these lesions. A thorough understanding of uterine leiomyomas, ovarian cysts, and endometriosis includes the ability to define these syndromes as described in this book, as well as an ability to explain them to clients. In addition, the reader is urged to review the material regarding the efficacy of the Pap smear in detecting cervical cancer at an early stage.

The section on disorders of pregnancy highlighted the most important aspects of this topic; the reader will probably wish to use this information as a basis for a more in-depth study of this area in a specialized course in maternal-child nursing. Finally, because of the widespread threat to women's health posed by disorders of the breast, the reader should pay particular attention to the final section. Specifically, the reader must be able to compare and contrast the differences between benign breast disorders and carcinoma of the breast and discuss the meaning and importance of various risk factors for breast carcinoma in a knowledgeable way.

RESOURCES

Menstrual Disorders

American College of Obstetricians and Gynecologists: Management of acute abnormal uterine bleeding in nonpregnant reproductive-aged women. *Obstet Gynecol* 121(4):891–896, 2013.

Bahamondes L, Bahamondes MV, Monteiro I: Levonorgestrel-releasing intrauterine system: uses and controversies. *Expert Rev Med Devices* 5(4):437–445, 2008.

Bennett AR, Gray SH: What to do when she's bleeding through: the recognition, evaluation, and management of abnormal uterine bleeding in adolescents. *Curr Opin Pediatr* 26(4):413–419, 2014.

Bradley LD, Gueye N-A: The medical management of abnormal uterine bleeding in reproductive-aged women. *AJOG* 214(1):31–44, 2016.

Curtis MG, Linares ST, Antoniewicz: *Glass' office gynecology*, ed 7, Baltimore, 2014, Lippincott Williams & Wilkins.

Davidson BR, DiPiero CM, Govoni KD, et al: Abnormal uterine bleeding during the reproductive years. *J Midwifery Womens Health* 57:248–254, 2012.

Fritz M, Speroff L: *Clinical gynecologic endocrinology and infertility*, Philadelphia, 2011, Lippincott Williams & Williams.

Hacker NF, Gambone JC, Hobel CJ: *Hacker & Moore's Essentials of obstetrics and gynecology*, ed 6, Philadelphia, 2016, Elsevier.

Heikinheimo O, Gemzell-Danielsson K: Emerging indications for the levonorgestrel-releasing intrauterine system (LNG-IUS). *Acta Obstet Gynecol Scand* 91(1):3–9, 2012.

Lefebvre G, et al: Primary dysmenorrhea consensus guideline. *JOGC* 27(12):1117–1146, 2005.

Lentz GM: Primary and secondary dysmenorrhea, premenstrual syndrome, and premenstrual dysphoric disorder. In Lentz GM, Lobo RA, Gershenson DM, Katz VL, editors: *Comprehensive gynecology*, ed 6, Maryland Heights, MO, 2012, Mosby Elsevier, pp 791–804.

Marsh CA, Grimstad F: Primary amenorrhea: diagnosis and management. *Obstet Gynecol Survey* 69(10):603–612, 2014.

Marret H, et al: CNGOF College National des Gynecologues et Obstetriciens Français: Clinical practice guidelines on menorrhagia: management of abnormal uterine bleeding before menopause. *Eur J Obstet Gynecol Reprod Biol* 152(2):133–137, 2010.

Osayande AS, Mehulic S: Diagnosis and initial management of dysmenorrhea. *Am Fam Physician* 89(5):341–346, 2014.

Stoelting-Gettelfinger W: A case study and comprehensive differential diagnosis and care plan for the three Ds of women's health: primary dysmenorrhea, secondary dysmenorrhea, and dyspareunia. *J Am Acad Nurse Pract* 22:513–522, 2010.

Disorders of the Pelvis and Uterine Anatomy

Centers for Disease Control and Prevention: *Sexually transmitted disease surveillance 2014*, 2015. Atlanta: U.S. Department of Health and Human Services.

Centers for Disease Control and Prevention: Sexually transmitted diseases treatment guidelines, 2015. *MMWR Morb Mortal Wkly Rep* 64(3): 2015.

Cohen SL, Vitonis AF, Einarsson JI: Updated hysterectomy surveillance and factors associated with minimally invasive hysterectomy. *JSLS* 18(3):2014.e2014.00096.

Hall GM, Shanmugan S, Nobel T, et al: Symptomatic rectocele: what are the indications for repair? *Am J Surg* 207(3):375–379, 2014.

Lentz GM: Anatomic defects of the abdominal wall and pelvic floor. In Lentz GM, Lobo RA, Gershenson DM, Katz VL, editors: *Comprehensive gynecology*, ed 6, Maryland Heights, MO, 2012, Mosby Elsevier, pp 453–474.

Lentz GM: History, physical examination, and preventative health care. In Lentz GM, Lobo RA, Gershenson DM, Katz VL, editors: *Comprehensive gynecology*, ed 6, Maryland Heights, MO, 2012, Mosby Elsevier, pp 113–128.

O'Grady JP: *Malposition of the uterus, Medscape*, 2015. Retrieved from http://emedicine.medscape.com/article/272497-overview, Dec. 15.

Benign and Cancerous Gynecologic Tumors

British Gynaecological Cancer Society. *Guidelines for the diagnosis and management of vulval carcinoma*, May 2014. Retrieved from https://www.rcog.org.uk/globalassets/documents/guidelines/vulvalcancerguideline.pdf.

Buck Louis GM, et al: ENDO Study Working Group: incidence of endometriosis by study population and diagnostic method: the ENDO study. *Fertil Steril* 96(2):360–365, 2011.

Day Baird D, et al: High cumulative incidence of uterine leiomyomas in black and white women: ultrasonic evidence. *Am J Obstet Gynecol* 188(1):100–107, 2003.

DeCherney AH, Murphy Goodwin T, Nathan L, Laufer N: *Current diagnosis & Treatment: obstetrics & gynecology*, ed 11, Boston, 2013, McGraw Hill.

Drayer SM, Catherino WH: Prevalence, morbidity, and current medical management of uterine leiomyomas. *Int J Gynecol Obstet* 131:117–122, 2015.

Dunselman GAJ, Vermeulen N, Becker C, et al: ESHRE guideline: management of women with edometriosis. *Hum Reprod* 29(3):400–412, 2014.

Durnell Schuiling K, Likis FE: *Women's gynecologic health*, ed 2, Burlington, MA, 2011, Jones & Bartlett.

Farquhar C: Endometriosis. *BMJ* 334(7587):249–253, 2007.

Mao A, Anastasi J: Diagnosis and management of endometriosis: the role of the advanced practice nurse in primary care. *J Am Acad Nurse Pract* 22(2):109–116, 2010.

Markowitz LE, et al: Quadrivalent human papilloma virus vaccine: recommendations of the advisory committee on immunization practices. *MMWR Morb Mortal Wkly Rep* 56(RR02):1–24, 2007.

Oliver A, Overton C: Detecting ovarian disorders in primary care. *Practitioner* 258(1769):15–19, 2014.

Petrosky E, Bocchini JA, Hariri S, et al: Use of 9-valent human papillomavirus (HPV) vaccine: updated HPV vaccination recommendations of the advisory committee on immunization practices. *MMWR* 64(11):300–304, 2015.

Smith RA, Manassaram-Baptiste D, Brooks D, et al: Cancer screening in the United States, 2015: a review of current American cancer society guidelines and current issues in cancer screening. *CA Cancer J Clin* 65:30–54, 2015.

Society of Gynecologic Oncology: *SGO clinical practice statement: salpingectomy for ovarian cancer prevention*, November 2013. Retrieved from https://www.sgo.org/clinical-practice/guidelines/sgo-clinical-practice-statement-salpingectomy-for-ovarian-cancer-prevention/.

U.S. Cancer Statistics Working Group: *United States cancer statistics: 1999-2012 incidence and mortality web-based report*, 2015. Atlanta, CDC and National Cancer Institute. Available at www.cdc.gov/uscs.

Vilos GA, Allaire C, Laberge PY, et al: The management of uterine leiomyomas. *J Obstet Gynaecol Can* 37(2):157–181, 2015.

Wise LA, et al: Age-specific incidence rates for self-reported uterine leiomyomata in the Black Women's Health Study. *Obstet Gynecol* 105(3):563–568, 2005.

Disorders of Pregnancy

American College of Obstetricians and Gynecologists: *Diagnosis and management of preeclampsia and eclampsia*, 2002. Washington, DC, The College, ACOG Practice Bulletin No. 33.

Goodwin TM: Hyperemesis gravidarum. *Obstet Gynecol Clin North Am* 35(3):401–417, 2008.

Kuklina EV, et al: Hypertensive Disorders and Severe Obstetric Morbidity in the United States. *Obstet Gynecol* 113:1299–1306, 2009.

Villar J, Say L, Gulmezoglu AM, et al: Eclampsia and pre-eclampsia: a health problem for 2000 years. In Critchly H, MacLean A, Poston L, Walker J, editors: *Pre-eclampsia*, London, 2003, RCOG Press, pp 189–207.

Wang X, Chen C, Wang L, et al: Conception, early pregnancy loss, and time to clinical pregnancy: a population-based prospective study. *Fertil Steril* 79(3):577–584, 2003.

Disorders of the Breast

American Cancer Society: *Cancer facts and figures*, Atlanta, 2017, Author.

Harris JR, Lippman ME, Morrow M, Kent Osborne C: *Diseases of the breast*, ed 5, Philadelphia, 2015, Lippincott Williams & Wilkins.

Kauff ND, et al: Risk-reducing salpingo-oophorectomy in women with a BRCA1 or BRCA2 mutation. *N Engl J Med* 346:1609–1615, 2002.

Stefanick ML, Anderson GL, Margolis KL: Effects of conjugated equine estrogens on breast cancer and mammography screening in postmenopausal women with hysterectomy. *JAMA* 295:1647–1657, 2006.

Thomsen A, Kolesar JM: Chemoprevention of breast cancer. *Am J Health Syst Pharm* 65:2221–2228, 2008.

Writing Group for the Women's Health Initiative Investigators: Risks and benefits of estrogen plus progestin in healthy postmenopausal women: Principal results from the Women's Health Initiative Randomized Control Trial. *JAMA* 288(3):321–333, 2002.

34

Sexually Transmitted Infections

Rosemary A. Jadack

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- What are the characteristic clinical manifestations and lesions of gonorrhea and chlamydial infection?
- How do the pathologic changes and clinical manifestations of syphilis differ during the incubation, primary, secondary, and tertiary phases?
- How do the lesions of herpes simplex, syphilis, and lymphogranuloma venereum differ?
- Which sexually transmitted diseases remain localized, and which have systemic consequences?
- What are the causative organisms and characteristic lesions of the following localized sexually transmitted diseases: chancroid, granuloma inguinale, molluscum contagiosum, and condylomata acuminata (genital warts)?

CHAPTER OUTLINE

Urethritis, Cervicitis, Salpingitis, and Pelvic Inflammatory Disease, 690
Gonococcal Infection, 690
Nongonococcal Infection, 691
Diseases With Systemic Involvement, 691
Syphilis, 691
Lymphogranuloma Venereum, 692
Herpesvirus Infections, 692
Diseases With Localized Lesions, 694
Ulcerative Lesions, 694
Chancroid, 694
Granuloma Inguinale, 694
Nonulcerative Lesions, 694
Molluscum Contagiosum, 694
Human Papilloma Virus Infections, 694
Enteric Infections, 695

An epidemic of sexually transmitted infections (STIs) currently exists in the United States. The Centers for Disease Control and Prevention (CDC) estimate that there are 19 million new infections every year. More than 800,000 cases of gonococcal infections and 2.8 million chlamydial infections. The true incidence of these infections is likely to be significantly higher inasmuch as many STIs are unreported. The cost of STIs is extremely high. It is estimated that STIs cost the American health care system $17 billion every year. In addition, the personal costs to the individual experiencing an STI may include pain, disfigurement, psychosocial difficulties, and reproductive problems. Because of the epidemic status of these diseases and the enormous costs associated with them, it is imperative that health care providers become sufficiently knowledgeable to assess their patients' STI status and educate them about STIs in an accurate and compassionate manner.

It is important to acknowledge that STIs are also alarming global health concerns and are a major cause of serious illness worldwide. Consequences of STIs include infertility, long-term disability, and death affecting millions of men, women, and infants. According to the World Health Organization, more than 1 million STIs are acquired each day worldwide. Every year, there are an estimated 105.7 million new cases of chlamydia, 106.1 million new cases of gonorrhea, 10.6 million new cases of syphilis, and 276.4 million new cases of trichomoniasis. The prevalence of genital herpes infection is more than 500 million people; 290 million women have acquired a human papilloma virus (HPV) infection.

Control of STIs is complex. An effective global response must include consistent access to medical treatment of disease. However, careful attention to societal issues that involve disparities, social and sexual networks, risk behaviors, attitudes, and cultural meanings surrounding sexuality is also essential. According to current research, effective interventions must be multifaceted and multidisciplinary. They must include quality epidemiology and surveillance, tested individual- and community-based interventions, methods to ensure that interventions focused on STIs are included in health care policy, and advocacy to get interventions implemented.

The term *sexually transmitted infections* refers to a large group of disease syndromes that can be transmitted sexually, regardless of whether the disease has manifestations in genital structures. In older texts, STIs are referred to as *sexually transmitted diseases* and *venereal diseases.* Although STIs are more prevalent in the 15- to 25-year-old age group,

they can occur at any age. These diseases are sometimes contracted by nonsexual transmission, as when a newborn infant contracts an STI from an infected mother during passage through the birth canal.

A list of sexually transmitted organisms grouped according to type of pathogen is found in Box 34.1. A useful approach to learning the complex pathophysiologic processes of STIs is to group them according to the disease manifestations that the patient is most likely to exhibit when first seen by the health care provider. These categories of STIs and the disease manifestations associated with them are listed in Table 34.1. This chapter describes each of these categories and the pathophysiologic processes associated with each relevant STI.

Some diseases listed in Table 34.1 are discussed elsewhere in this text but have been included here for completeness. In particular, the reader may wish to refer to Chapter 33 for more detailed information on pelvic inflammatory disease and vulvovaginitis. Also, certain systemic infections are potentially transmitted by sexual contact. Cytomegalovirus infection; hepatitis A, B and C; and HIV have the potential for sexual transmission. These diseases are covered in detail in Units III and IX, along with more in-depth information concerning infectious processes and immune responses. Before studying this chapter, the reader may wish to refer to Chapters 8 and 9 for a review of basic terminology such as *incubation period* and *period of communicability.* Health care providers caring for persons at risk for STIs should be aware of the potential for acquisition of systemic diseases by sexual contact and include assessment of these diseases as part of their overall clinical evaluation.

BOX 34.1 Sexually Transmitted Organisms

Bacterial Pathogens
Calymmatobacterium granulomatis
Chlamydia trachomatis
Gardnerella vaginalis
Haemophilus ducreyi
Mycoplasma hominis
Neisseria gonorrheae
Shigella
Group B streptococci
Ureaplasma urealyticum
Treponema pallidum

Fungal Pathogens
Candida albicans
Candida glabrata

Viral Pathogens
Human immunodeficiency virus
Cytomegalovirus
Herpes simplex virus
Hepatitis virus
Human papilloma virus
Molluscum contagiosum virus

Protozoan Pathogens
Entamoeba histolytica
Giardia lamblia
Trichomonas vaginalis

TABLE 34.1 Sexually Transmitted Infections Categorized According to Disease Manifestations

Disease Manifestations	Disease
Urethritis, cervicitis, and salpingitis	Gonorrhea
	Nongonococcal urethritis
	Pelvic inflammatory disease
Ulcerative lesions with systemic involvement	Syphilis
	Lymphogranuloma venereum
	Herpes simplex virus
Ulcerative lesions only	Chancroid
	Granuloma inguinale (donovanosis)
Nonulcerative lesions	Molluscum contagiosum
	Genital warts (condylomata acuminata)
Vulvovaginitis	Trichomoniasis
	Candidiasis
	Gardnerella vaginalis vaginitis
Systemic infections	Cytomegalovirus
	Hepatitis
	AIDS
Enteric infections	Giardiasis
	Campylobacter enteritis
	Shigellosis
	Amebic dysentery

URETHRITIS, CERVICITIS, SALPINGITIS, AND PELVIC INFLAMMATORY DISEASE

Three types of STI are manifested by *urethritis* (inflammation of the urethra), *cervicitis* (inflammation of the uterine cervix), and/or *salpingitis* (inflammation of the oviduct or fallopian tube). Gonorrhea is an inflammation of epithelial tissue by the organism *Neisseria gonorrheae.* In men, nongonococcal urethritis refers to urethritis resulting from a pathogen other than the gonococcus, which is usually *Chlamydia trachomatis.* In women, mucopurulent cervicitis refers to an inflammation of the cervix, which is usually caused by either *C. trachomatis* or *N. gonorrheae. Pelvic inflammatory disease,* which was described in Chapter 33, is usually the result of acute salpingitis caused by gonococcal or chlamydial infection that has extended into nearby pelvic tissue.

Gonococcal Infection

Gonorrhea is associated with the gram-negative diplococcus *N. gonorrheae.*

Etiology and clinical manifestations. In gonorrhea, disease transmission occurs through contact with exudates from the mucous membranes of infected persons, usually by direct contact. The gonococcus then attaches to and penetrates columnar epithelium and produces a patchy inflammatory response in the submucosa. Although usually asymptomatic in women, gonorrhea may produce purulent vaginal discharge, dysuria, and abnormal vaginal bleeding. The most commonly affected areas in women are the cervix, the urethra, the Skene and Bartholin glands, and the anus. Among females, adolescents (ages 15 to 19 years) and young adults (ages 20 to 24 years) now have the highest rates of gonorrhea. In men, symptoms of urethritis, including dysuria and a purulent urethral discharge accompanied by redness and swelling at the site of infection, usually occur after a 3- to 6-day incubation period. Among males, young adults (ages 20 to 24 years) have the highest rates of gonorrhea. In both genders, infection and inflammation of the pharynx, conjunctivae, and anus may be present. Direct extension of the infection with gonococci occurs by way of the lymphatic system. In the female, extension may spread unilaterally or bilaterally to the oviducts, with subsequent salpingitis. In the male, direct extension of the infection most frequently occurs to the epididymis.

Once gonococcal infection has spread to other areas, localized infection occurs and may cause the formation of cysts and abscesses. Purulent exudate containing the organism causes damage to tissue, and fibrous tissue replaces inflamed tissue. This hardened, fibrous tissue may result in scarring and narrowing of the urethra, epididymis, or oviducts. In women, partial or complete closure of the oviducts results in sterility. Infection of the oviducts may also result in pelvic inflammatory disease if exudate is released into the peritoneal cavity. As described in Chapter 33, pelvic inflammatory disease may be an acute or chronic condition causing widespread damage to the pelvic organs in the female.

Nongonococcal Infection

Etiology. Nongonococcal urethritis and cervicitis are often caused by strains of *C. trachomatis* that act on columnar epithelium in a manner similar to that noted for the gonococcus. The symptoms of infection with *Chlamydia* are generally less severe than those of gonorrhea. As with gonorrhea, the infection may spread by extension to the oviducts, and pelvic inflammatory disease may eventually result. Upper reproductive tract infection, whether symptomatic or subclinical, is an important cause of infertility and ectopic pregnancy. Transmission of *Chlamydia* during birth may result in ophthalmia neonatorum, or infection of the eyes in the newborn.

Treatment. The resistance of *N. gonorrheae* to antimicrobial agents continues to spread and intensify, causing concern nationally and internationally. Antimicrobial agents such as ceftriaxone, cefixime, spectinomycin, and cephalosporin are used to manage uncomplicated gonococcal infections. The CDC no longer recommends fluoroquinolones for treatment of gonococcal infections because of fluoroquinolone resistance by *N. gonorrheae.* Increased resistance to cephalosporin is also currently being monitored. Unless chlamydial infection is ruled out, dual therapy for gonococcal and chlamydial infection consisting of azithromycin or doxycycline added to one of the aforementioned agents is recommended. Pelvic inflammatory disease is also generally managed with two agents to cover potential chlamydial and gonorrheal infection. A number of organizations worldwide now recommend *Chlamydia* (as well as gonorrhea) screening for sexually active adolescents and women through age 25 who have no symptoms in order to reduce the sequelae of infection.

KEY POINTS

- Urethritis, cervicitis, salpingitis, and pelvic inflammatory disease commonly result from gonorrheal or chlamydial infection. Transmission is usually by direct contact with infected mucous membranes. The symptoms of chlamydial infection are similar to but usually less severe than those of gonorrhea.
- Gonorrhea may produce purulent discharge, dysuria, and abnormal vaginal bleeding. Cysts and abscesses may form in localized areas of infection, followed by scarring and fibrosis. Inflammation of the pharynx, conjunctivae, and anus may be present. Antibiotic therapy is indicated.

DISEASES WITH SYSTEMIC INVOLVEMENT

Several STIs cause a distinctive ulcerative lesion and disseminate throughout the body to affect multiple organ systems. Most prominent of this type of STI are syphilis, herpesvirus infections, and lymphogranuloma venereum.

Syphilis

Syphilis is a systemic infection of the vascular system consisting of five distinct stages: incubation, primary and secondary stages, latency, and late syphilis. Syphilis is communicable by persons with primary, secondary, or early latent syphilis. The incidence has varied in the United States since reporting began in 1941. The rate of primary and secondary syphilis decreased during the 1990s to its lowest levels since 1941. However, overall rates increased again between 2001 and 2014. Until 2013, increased incidence rates were attributed primarily to men who have sex with men (MSM). However, rates increased 22.7% for women between 2013 and 2014. This increase is concerning because congenital syphilis rates tend to increase with increased incidence rates in women. Current estimates of primary and secondary syphilis are 6.3 cases per 100,000 population in the United States (2013–2014).

Etiology. Syphilis is caused by *Treponema pallidum,* an anaerobic spirochete. The disease is acquired when *T. pallidum* penetrates intact mucous membranes or abraded skin during sexual contact. (The process of transmission of congenital syphilis is described later.) Some of the *T. pallidum* pathogens remain at the original invasion site, whereas others migrate to regional lymph nodes within hours. During this incubation phase, *T. pallidum* is disseminated throughout the body and can invade and multiply in any organ system.

Pathogenesis. During all stages of syphilis, invasion of tissue by *T. pallidum* results in pathologic changes in the vascular system. The inflammatory response in endothelial tissue causes the infiltration of lymphocytes and plasma cells, with subsequent endothelial swelling. The terminal arterioles and small arteries may become obliterated and no longer functional. Finally, long-term inflammation of vascular tissue results in the formation of hardened, fibrous thickening in the blood vessels and eventually tissue necrosis.

After the initial incubation period of 10 to 90 days, the primary phase begins with the formation of a chancre, a painless, ulcerative lesion that arises at the original spirochete portal of entry (Fig. 34.1). The chancre may remain unnoticed in a female if it occurs on the cervix or in the vagina; in fact, most cases of syphilis in women are undiagnosed until recognized by positive testing of the blood in the latent phase. In males, the chancre may form on the genitalia; in both genders, chancres may erupt on the anus, fingers, lips, tongue, nipples, tonsils, or eyelids.

Untreated chancres will resolve spontaneously within 3 to 6 weeks and are followed by the secondary stage of syphilis, which is characterized by a low-grade fever, malaise, sore throat, headache, lymphadenopathy, and mucosal or cutaneous rash (Fig. 34.2). This secondary stage occurs as *T. pallidum* is spread throughout the bloodstream and lymphatic system. The secondary stage is also self-limiting and is followed by a latent phase in which no symptoms are present. During the latent stage, the affected person will test positive for syphilis on serologic assays and may still experience infectious mucocutaneous lesions during the early latent stage. Thus the early latent stage is considered contagious. The latent stage is of variable length and may last more than 40 years. In approximately two-thirds of patients, the infection remains asymptomatic and never causes a recurrence of symptoms. If syphilis remains untreated, then late syphilis—the final, destructive phase of the disease—will

FIG 34.1 Typical syphilitic chancre, a painless, ulcerative lesion that arises at the original spirochete portal of entry. (From Morse SA et al: *Atlas of sexually transmitted diseases and AIDS,* ed 4, St Louis, 2010, Elsevier.)

FIG 34.2 Typical generalized skin rash of secondary syphilis. (From Morse SA et al: *Atlas of sexually transmitted diseases and AIDS,* ed 4, St Louis, 2010, Elsevier.)

FIG 34.3 Lymphogranuloma venereum is characterized by a small, transient genital ulcer with swollen, extremely painful inguinal lymph nodes. (From Lewis SM et al: *Medical-surgical nursing,* ed 6, St Louis, 2004, Mosby, p 1392.)

eventually develop in approximately one-third of affected people. The manifestations of late syphilis depend on the area of arterial lesions and the extent of circulatory insufficiency. Body systems particularly at risk are the cardiovascular and central nervous systems. Damage to the cardiovascular system may include aortic necrosis and subsequent aortic insufficiency; damage to the central nervous system may be progressively widespread, with degeneration of the cortical neurons and, eventually, paresis, blindness, and mental deterioration.

Transmission of *T. pallidum* from the mother to the fetus may occur transplacentally at any point during pregnancy, but an inflammatory response to the pathogen does not develop in the fetus until around the fifteenth week of gestation. Therefore treatment of infected women before the fifteenth week may prevent damage to the fetus. Infection with syphilis before birth may result in physical deformities and developmental disabilities in the infant. Infants born to untreated or inadequately treated mothers will have active infection and must be treated. A presumptive diagnosis of syphilis is generally based on a positive result of a serologic screening test, such as the Venereal Disease Research Laboratories or rapid plasma reagin test, followed by a positive result of a treponemal serologic test. Dark-field examination of tissue and exudates or direct fluorescent antibody tests are also useful.

Treatment. Benzathine penicillin G is the first choice for the management of syphilis. If the affected person is allergic to penicillin, tetracycline or doxycycline is given. Treatment is administered to all individuals with positive evidence of syphilis on laboratory testing and to people who have had sexual contact with infected individuals. Response to antibiotic treatment is monitored by repeating laboratory testing for evidence of syphilis at regular intervals up to 24 months after therapy. Management during pregnancy is complex, but it focuses on maternal cure and prevention of congenital syphilis.

Lymphogranuloma Venereum

Lymphogranuloma venereum (LGV) is a highly contagious systemic infection caused by a number of closely related strains of *Chlamydia.* The disease occurs more commonly in the tropics, but rates have been increasing in industrialized countries. LGV develops more often in males than in females and has a higher incidence among sexually active young adults. In the United States increasing rates are in part attributed to MSM.

Etiology and pathogenesis. Like syphilis, LGV has stages of development in which an initial lesion forms and systemic disease occurs after dissemination via the lymphatic system. After invasion of the mucosa by *Chlamydia* during sexual contact, a painless lesion appears on the genitalia after a 1- to 3-week incubation period. The lesion may range from a slight erosion to a small papule and often remains undetected (Fig. 34.3). This lesion heals spontaneously in a few days. During this period, the pathogens are disseminated to regional lymph nodes, primarily the inguinal lymph nodes.

About 2 weeks after appearance of the primary lesion, the inguinal lymph nodes begin to swell, and the systemic symptoms of fever and malaise develop. The nodal swelling is a manifestation of inflammation of the lymphatic system in which lesions filled with polymorphonuclear leukocytes are forming in the lymph nodes. Spread of the inflammation throughout adjacent lymph nodes causes multiple nodes to become matted together and form a large abscess. These abscesses are said to be regional because they develop in one or more areas along the lymphatic system. If a person with LGV remains untreated, the abscesses rupture through the skin and other body cavities to create chronic fistulas. Thus, as the regional lymphadenitis progresses, complications such as perianal and rectovaginal fistulas develop, along with strictures of the rectum. Other complications include extreme swelling of the genitalia; this occurs because the normal lymph drainage of this area is impeded. The diagnosis is usually made by serologic testing (antibody titers).

Treatment. Doxycycline is the recommended antibiotic, with erythromycin being the alternative. Surgical treatment may include aspiration of lymph nodes as needed; rectal strictures and fistulas may require surgical correction.

Herpesvirus Infections

Herpesviruses are an important group of viral agents that produce infection in humans. Two types of herpes simplex virus (HSV)—type 1 and type 2—may be sexually transmitted and are discussed in this section. HSV type 1 is most often associated with herpetic infections above the waist, typically in the oral cavity and on the lips, but also in the eyes or on the epidermis. HSV type 1 can be transmitted sexually and can cause genital herpetic infections. It is present in saliva, stool, and urine. The vesicles resulting from type 1 infection in the oral cavity are commonly referred to as *cold sores* or *chancres* and often affect children younger than 5 years. HSV type 2 is implicated in most genital,

anal, and perianal herpes and is sometimes referred to as *genital herpes* for this reason. Type 2 HSV can also result in oral lesions after sexual contact. Serologic studies indicate that one in five individuals in the United States have been infected with type 2 HSV. Although type 2 is primarily transmitted through sexual contact, pregnant mothers can transmit the infection to newborns during vaginal delivery.

Etiology. HSV types 1 and 2 have certain characteristics in common. Both produce an initial infection that is self-limiting. The lesions produced by this infection heal, but HSV continues to be present in the body. Recurrence of the lesions, usually in the area of the initial infection, may take place as the virus is reactivated. Recurrence of either type may be triggered by an infectious disease, emotional stress, or immunosuppression. The exact mechanism for reactivation of the virus is presently unknown, but it is thought that ganglion neurons may contain latent forms of the virus and then receive a trigger to stimulate replication of the virus under certain conditions.

Clinical manifestations. Genital infection with HSV type 2 is manifested by the appearance of fluid-filled vesicles after a 3- to 7-day incubation period. In the female, the cervix is usually the primary infection site, although the labia, perianal skin, vulva, or vagina may also be involved (Fig. 34.4). In the male, the vesicles are located on the glans penis, foreskin, or penile shaft (Fig. 34.5). Extragenital lesions may appear on the mouth or anus. In both males and females, the vesicles, which are usually painless at first, may rupture and develop into extensive shallow, painful ulcers. The virus may enter the lymphatic system and create localized lesions there; thus the inguinal lymph nodes may be edematous and tender. Rarely, the virus spreads to visceral organs and can produce areas of necrosis in the liver, adrenal glands, lungs, and central nervous system. In newborns and people with weak immune defenses (particularly people with AIDS), HSV type 2 may result in severe damage to these organ systems, with high related mortality. Most HIV-infected individuals are HSV-2 seropositive, and perirectal involvement is common.

HSV type 1 infections may appear as single or multiple fluid-filled, tender vesicles in the oral cavity or on the lips. Usually, the appearance of the lesions is preceded by 1 or 2 days of paresthesia before the chancre, or "cold sore," erupts. These lesions will generally crust and heal within 3 to 10 days. HSV infections are usually diagnosed by cell culture. Exudate from early lesions gives the most positive results.

Treatment. Normally, HSV type 1 lesions are self-limiting and respond to measures that promote good oral hygiene. HSV type 2 genital lesions are usually self-limiting but may be extremely painful. The use of antiviral agents such as acyclovir, famciclovir, and valacyclovir has been shown to accelerate healing time and reduce the duration and severity of symptoms in initial episodes of HSV type 2 infection. Long-term or episodic suppression may also be used to reduce the number or severity of recurrent episodes. Consistent use of condoms reduces transmission from men to women and is a major focus of STI prevention worldwide. Prevention teaching should include the possibility of asymptomatic shedding of the virus. If HSV type 2 lesions are active in a pregnant mother at term, a cesarean section may be recommended to reduce the risk of transmission to the newborn.

KEY POINTS

- Syphilis is caused by an anaerobic spirochete that is transmitted sexually but disseminates throughout the body during incubation. Manifestations of early syphilis include chancre formation at the portal of entry, which spontaneously resolves in 3 to 6 weeks if untreated. General malaise, fever, sore throat, and rash may then occur, followed by an asymptomatic latent phase. The latent phase may last more than 40 years. Late syphilis is characterized by central nervous system degeneration, blindness, and paresis.
- Lymphogranuloma venereum (LGV) is a highly contagious systemic infection caused by strains of *Chlamydia*. An initial painless genital lesion appears after 1 to 3 weeks of incubation. The infection spreads to regional lymph nodes and is accompanied by fever and malaise. Infected lymph nodes become abscessed and may rupture through the skin and body cavities, causing fistula formation.
- Herpes simplex virus (HSV) types 1 and 2 are implicated in cases of genital herpes. Herpes lesions are fluid-filled vesicles that appear 3 to 7 days after infection. The virus may enter the lymphatics and cause inguinal lymph node tenderness. Although the lesions may disappear, the virus remains in the body, thus predisposing to recurrence. Herpes may be transmitted from mother to newborn during the birth process.

FIG 34.4 Primary genital herpes in the female showing herpetic vesicles and tender ulcerations. (From Morse SA et al: *Atlas of sexually transmitted diseases and AIDS,* ed 4, St Louis, 2010, Elsevier.)

FIG 34.5 Recurrent genital herpes in the male showing erythema, groups of vesicles, erosions, and edema on the shaft of the penis. (From Morse SA et al: *Atlas of sexually transmitted diseases and AIDS,* ed 4, St Louis, 2010, Elsevier.)

DISEASES WITH LOCALIZED LESIONS

Ulcerative Lesions

Two types of STIs result in the formation of ulcerative lesions but do not progress to systemic involvement. Chancroid (also called *soft chancre*) and granuloma inguinale are both manifested by ulcerative lesions, although their pathophysiologic courses differ.

Chancroid

Etiology. Chancroid is an ulcerative, infectious disease of the genital tract caused by the sexually transmitted anaerobic bacillus *Haemophilus ducreyi.* Chancroid is relatively rare in the United States. The disease is a cofactor for HIV infection.

Pathogenesis and clinical manifestations. *H. ducreyi* initially invades the genital skin or mucous membranes at sites traumatized by sexual contact. The patient generally has one or more painful genital ulcers, unlike the chancre in syphilis, which is generally solitary and painless. Fresh lesions may occur from autoinoculation (self-infection). The ulcerated lesions may enlarge, continue to erode (Fig. 34.6), and produce destruction of surrounding tissue. In addition, inguinal lymph nodes may become tender and painful as the infection is disseminated to this region. If the infection goes untreated, the enlarged lymph gland (called a *bubo*) may rupture, draining pus and leaving a large inguinal ulcer. The infection is communicable until the lesions heal, which may be a period of weeks. Scarring may occur in advanced cases. Diagnosis is usually made by culture for *H. ducreyi.*

Treatment. Antiinfective agents recommended for management of chancroid include azithromycin, erythromycin, ceftriaxone, and ciprofloxacin. Large ulcers may not heal for more than 2 weeks. As with all STIs, sexual partners should be treated simultaneously and reexposure avoided until therapy is completed.

Granuloma Inguinale

Etiology. *Calymmatobacterium granulomatis* is the causative agent of granuloma inguinale. This intracellular bacterium is also referred to as a *Donovan body* and the disease as *donovanosis.* Granuloma inguinale is rare in the United States.

FIG 34.6 Eroded, purulent ulcer of chancroid. (From Morse SA et al: *Atlas of sexually transmitted diseases and AIDS,* ed 4, St Louis, 2010, Elsevier.)

Pathogenesis and clinical manifestations. Transmission of granuloma inguinale is not clearly understood. It is generally thought to be an STI, but the disease is also seen in adults who are not sexually active and in young children, possibly as a result of autoinoculation. The causative bacterium is found in the rectum of nondiseased persons, which suggests that the organism may be part of the normal gastrointestinal flora in some persons.

The incubation period is variable and ranges from a few days to months. The initial sign of the disease may be a painless papule or nodule that subsequently ulcerates into an enlarging, granulomatous, red velvety ulcer. The raised mass of granulation tissue may look more like a tumor than an ulcer. The lesions are highly vascular and bleed easily with minor contact. Single or multiple lesions may coalesce, or lesions may spread to nearby tissue. Secondary infection of the ulcers and expanding tissue necrosis in lesions may lead to erosion of the genitals. Diagnosis is by identification of the dark-staining Donovan bodies on biopsy or tissue crush preparations.

Treatment. Doxycycline is given for at least 3 weeks or until all lesions are healed. Alternative agents include trimethoprim-sulfamethoxazole, ciprofloxacin, erythromycin, and azithromycin.

Nonulcerative Lesions

Molluscum contagiosum and infections caused by HPV (also called *genital warts*) are two prevalent types of STIs that produce nonulcerative lesions. Both are caused by viral agents that invade superficial layers of the epidermis during sexual contact.

Molluscum Contagiosum

Etiology. Molluscum contagiosum is a viral skin disease caused by a member of the poxvirus family. (The term *poxvirus* refers to a viral agent that causes an eruption, or "pox," on the skin.) The manifestations are much milder than those of smallpox or chickenpox. Two forms of the disease exist. One affects children and is transmitted by skin-to-skin contact and indirect contact; the other affects young adults and is transmitted during sexual contact.

Pathogenesis and clinical manifestations. After invasion of the epidermis by the virus, pink to white lesions with an exudative core appear on the genitalia. The lesions are multiple, are slow to develop, and remain stable for long periods. The disease is usually asymptomatic.

Treatment. The goal of treatment is primarily to prevent spread of the infection for cosmetic reasons. The lesions can be removed by minor surgery or frozen with liquid nitrogen. Sexual contacts of affected persons should be examined to prevent further spread.

Human Papilloma Virus Infections

Etiology. Human papilloma virus (HPV) infections cause epithelial lesions of the anogenital region. Also called *genital warts* or *condylomata acuminata,* HPV is predominantly transmitted sexually in young adults, with the highest prevalence in the 16- to 25-year-old age group. The risk of contracting the disease by sexual contact with an infected person is high; lesions will develop in up to two thirds of the sexual contacts of affected persons. Nonsexual transmission has also been documented, and lesions have been found in infants. The period of communicability is unknown, but is thought to last as long as the lesions persist, and perhaps even after they are clinically removed.

Pathogenesis and clinical manifestations. After invasion of the epidermis by HPV, an incubation period of 1 to 20 months (usually about 4 months) precedes the appearance of lesions. It is thought that the virus infects single epithelial cells and stimulates the cells to divide and proliferate into the wartlike lesions. The lesions can be single or multiple and may have a soft pink to brown coloring. They can be small or large and raised or flat (Fig. 34.7). The lesions are generally

FIG 34.7 Human papilloma virus of the vulva. (From Morse SA et al: *Atlas of sexually transmitted diseases and AIDS,* ed 4, St Louis, 2010, Elsevier.)

asymptomatic but may be pruritic (itchy), painful, or friable (bleed easily). In females, HPV may be found in the vagina and cervix, as well as in the anogenital area. In males, HPV lesions may occur in the anterior urethra and anogenital area.

Treatment. External genital warts may be treated with patient-applied podofilox or imiquimod topical preparations. Providers can perform cryotherapy; administer podophyllin, trichloroacetic acid, or bichloroacetic acid; or carry out surgical excision of the warts. Patient-applied preparations are also available and include podofilox (0.5%) and imiquimod (5%) cream. Alternative regimens include intralesional interferon and laser surgery. Malignant transformation to invasive carcinoma has been observed with some types of genital warts. Persistent infections of HPV types 6, 11, 16, and 18 are thought to be responsible for 70% of cervical cancers and 90% of genital warts. Therefore it is generally agreed that affected persons should be treated or monitored carefully. Because HPV has been associated with cervical cancer, there has been an increasing use of HPV testing in evaluating abnormal Papanicolaou smears.

In 2006 the Food and Drug Administration licensed the first vaccine developed to prevent cervical cancer and other diseases in females caused by certain types of HPV (see Chapter 33). Bivalent, quadravalent, and nine-valent forms of the vaccine are available. The bivalent version of the vaccine protects against HPV types 16 and 18. The quadravalent form protects against HPV types 6, 11, 16, and 18. The nine-valent form of the vaccine is recommended for 11- to 12-year-old girls, but can be administered to girls as young as 9 years of age. The vaccine is recommended for young men if it is readily available. Ideally, the vaccine should be administered before the onset of sexual activity. However, females and males who are sexually active also may benefit from vaccination.

KEY POINTS

- Chancroid is caused by infection with an anaerobic bacillus. Initially the lesion is a small erythematous papule, and after 2 to 3 days the painful lesion ulcerates. Lesions resemble those of syphilis; however, the lesions of syphilis are painless.
- Granuloma inguinale is caused by an intracellular bacterium. The initial papule is painless and subsequently ulcerates into a growing granulomatous ulcer resembling a tumor.
- Molluscum contagiosum is associated with infection by a poxvirus. Genital lesions are pink to white with an exudative core. The disease is usually asymptomatic and self-limiting.
- Condylomata acuminata, or genital warts, is associated with infection by human papilloma virus (HPV). Warts are pink to brown and painless and may occur in clusters. Persistent HPV infection is an important risk factor for cervical cancer.

ENTERIC INFECTIONS

Until recently, information regarding the transmission of enteric infections of the gastrointestinal tract through sexual contact was limited. Enteric pathogens may be transmitted sexually among any individuals who engage in direct or indirect fecal-oral contact. Enteric organisms that may be transmitted through sexual contact include *Giardia, Campylobacter, Shigella,* and the agents causing amebic dysentery.

The pathophysiologic process of enteric infections of the gastrointestinal tract is described in Chapter 36, and the reader may wish to refer to this material. In general, persons who have acquired enteric infections by sexual contact will have variable manifestations. Some individuals may experience no symptoms, whereas others will have marked symptoms of enteritis or proctitis. All individuals who engage in oral-anal sexual practices should be monitored for the presence of enteric infections with laboratory studies and diagnostic examinations. Education for persons at risk for sexually transmitted enteric infections includes an emphasis on protective hygienic practices. Infected persons should avoid all sexual contact until all partners are examined and treated if necessary. After completion of appropriate therapy for enteric infections, affected individuals should be retested for assessment of therapeutic effectiveness.

SUMMARY

Because of the epidemic nature of STIs worldwide, it is essential for readers preparing for careers in the health sciences to have a complete grasp of the material in this chapter. The STIs considered in the chapter are grouped according to the disease manifestations that the patient is most likely to exhibit. Gonorrhea, most chlamydial infections, and pelvic inflammatory disease are manifested by urethritis, cervicitis, or salpingitis. A second group of STIs cause ulcerative lesions with systemic involvement. Syphilis, herpes, and LGV all cause a distinctive ulcerative lesion and may disseminate throughout the body to affect multiple organ systems.

In reviewing the material on STIs related to ulcerative and nonulcerative lesions, the reader should compare and contrast the appearance of these lesions and consider the differing pathophysiologic characteristics of each type. Finally, the reader should consider how he or she would incorporate this material into an overall assessment process, as well as interventions focused on prevention. Nurses and other health care providers caring for persons at risk for STIs should be aware of the potential for acquisition of these diseases as well. The overall goal in learning the material in this chapter is to be able to assess and educate clients with STIs in a comfortable and accurate manner.

RESOURCES

Centers for Disease Control and Prevention: Sexually transmitted diseases treatment guidelines. *MMWR Morb Mortal Wkly Rep* 64(3):2015, 2015.

Centers for Disease Control and Prevention: *Sexually transmitted disease surveillance 2014*, Atlanta, GA, 2015, U.S. Department of Health and Human Services.

Centers for Disease Control and Prevention: CDC grand rounds: the growing threat of multidrug-resistant gonorrhea. *MMWR Morb Mortal Wkly* 62(6):103–106, 2013.

Centers for Disease Control and Prevention: Cephalosporin susceptibility among Neisseria gonorrhoeae isolates—United States, 2000-2010. *MMWR Morb Mortal Wkly Rep* 60:873–877, 2011.

Centers for Disease Control and Prevention: Lymphogranuloma venereum among men who have sex with men: Netherlands, 2003-2004. *MMWR Morb Mortal Wkly Rep* 53(42):985–988, 2004.

Giuliano AR, et al: Efficacy of quadrivalent HPV vaccine against HPV infection and disease in males. *N Engl J Med* 363(5):401–411, 2011.

Holmes KK, et al: *Sexually transmitted diseases*, ed 4, New York, 2007, McGraw-Hill.

Kirkcaldy RD, Ballard RC, Dowell D: Gonococcal resistance: are cephalosporins next? *Curr Infect Dis Rep* 13:196–204, 2011.

LeFerve ML: Screening for chlamydia and gonorrhea: US Preventive Services Task Force recommendation statement. *Ann Intern Med* 161(2):902–910, 2014.

Low N, Broutet N, Adu-Sarkodie Y, et al: Global control of sexually transmitted infections. *Lancet* 368(9551):2001–2016, 2006.

Menon S, Timms P, Allan JA, et al: Human and pathogen factors associated with Chlamydia trachomatis-related infertility in women. *Clin Microbiol Rev* 28(4):969–985, 2015.

Moore MS, Golden MR, Scholes D, Kerani R: Assessing trends in chlamydia positivity and gonorrhea incidence and their associations with the incidence of pelvic inflammatory disease and ectopic pregnancy in Washington state, 1988-2010. *Sex Transm Dis* 43(1):2–8, 2016.

Moreno R, Nababan HY, Ota E, et al: Structural and community-level interventions for increasing condom use to prevent the transmission of HIV and other sexually transmitted infections. *Cochrane Database Syst Rev* (7):CD003363, 2014.

Morris GC, Stewart CM, Schoeman SA, Wilson JD: A cross-sectional study showing differences in the clinical diagnosis of pelvic inflammatory disease according to the experience of clinicians: implications for training and audit. *Sex Transm Infect* 90(6):445–451, 2014.

Panel on Opportunistic Infections in HIV-Infected Adults and Adolescents: *Guidelines for the prevention and treatment of opportunistic infections in HIV-infected adults and adolescents: recommendations from the Centers for Disease Control and Prevention, the National Institutes of Health, and the HIV Medicine Association of the Infectious Diseases Society of America.* Available at http://aidsinfo.nih.gov/contentfiles/lvguidelines/adult_oi.pdf.

Petrosky E, Bocchini JA, Hariri S, et al: Use of 9-valent human papillomavirus (HPV) vaccine: updated HPV vaccination recommendations of the advisory committee on immunization practices. *MMWR* 64(11):300–304, 2015.

Schwarz TF, et al: Four-year follow-up of the immunogenicity and safety of the HPV-16/18 AS04-adjuvanted vaccine when administered to adolescent girls aged 10-14 years. *J Adolesc Health* 50(2):187–194, 2012.

Van der Bij AK, Spaargaren J, Morre SA, et al: Diagnostic and clinical implications of anorectal lymphogranuloma venereum in men who have sex with men: a retrospective case-control study. *Clin Infect Dis* 42(2):186–194, 2006.

Van Velzen M, Ouwendijk W, Selke S, et al: Longitudinal study on oral shedding of herpes simplex virus 1 and varicella-zoster virus in individuals infected with HIV. *J Med Virol* 85:1669–1677, 2013.

World Health Organization: *Global incidence and prevalence of selected curable sexually transmitted infections – 2008*, Geneva, 2012, WHO.

World Health Organization: *Sexually transmitted infections: Fact sheet No. 110*, Dec 2015. Retrieved from http://www.who.int/mediacentre/factsheets/fs110/en/.

Zenilman JM, Shahmanesh M: *Sexually transmitted infections: diagnosis, management, and treatment*, Sudbury, MA, 2011, Jones and Bartlett Learning.

35

Gastrointestinal Function

Jeffrey S. Sartin

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- What are the major structures of the gastrointestinal tract and their corresponding functions?
- How does the autonomic nervous system influence gastrointestinal motility?
- How do segmental and propulsive movements influence the digestive and absorptive functions of the small intestine?
- What are the major secretions of each of the following secretory cells and glands: salivary, gastric, intestinal epithelium, pancreas, and gallbladder?
- How and where are complex carbohydrates, proteins, and lipids digested and absorbed?
- How and where are water and electrolytes absorbed?
- What alterations in gastrointestinal function occur in association with very young or very old age?

CHAPTER OUTLINE

Structure and Organization of the Gastrointestinal Tract, 698
Embryology, 698
Functional Anatomy, 699
Oral Cavity and Pharynx, 699
Esophagus, 700
Stomach, 700
Small Intestine, 701
Large Intestine, 703
Gastrointestinal Motility, 704
Characteristics of the Intestinal Wall, 704
Neural Control, 704
Parasympathetic Innervation, 705
Sympathetic Innervation, 705
Afferent Nerve Fibers, 705
Electrical Activity of Gastrointestinal Smooth Muscle, 705
Hormonal Control, 706
Movement in the Gastrointestinal Tract, 706
Contraction of Gastrointestinal Smooth Muscle, 706
Propulsive Movements, 706
Mixing Movements, 707
Movement of Nutrients, 707
Chewing, 707
Swallowing, 707
Motor Functions of the Stomach, 708
Motility of the Small Intestine, 709
Ileocecal Sphincter, 710
Motility of the Colon, 710
Secretory Function, 712
Secretion of Gastrointestinal Juices, 712
Gastrointestinal Hormones, 712
Digestion and Absorption, 712
Digestion of Carbohydrates, 713
Digestion of Lipids, 713
Digestion of Proteins, 714
Absorption, 715
Carbohydrates, 715
Lipids, 715
Proteins, 715
Water and Electrolytes, 715
Gastrointestinal Function Across the Life Span, 717
Maturation, 717
Age-Related Changes, 717

The gastrointestinal (GI) system represents a remarkable interface between the human organism and the external environment for the purpose of providing nutrients for the body. The components of the GI tract can be thought of as a continuous tube about 7 m in length extending from the mouth to the anus. Beginning with the mouth and pharynx, the GI tract includes the esophagus, stomach, and small and large intestines (Fig. 35.1). Other parts of the GI system located outside the GI tract include the salivary glands, the pancreas, and the biliary system (liver, gallbladder, and bile ducts).

The process of ingesting nutrients, propelling them through the GI tract, and transforming them into a form capable of absorption into the body's internal milieu is remarkably complex. The general functions of the GI tract can be divided into (1) movement of nutrients, including propulsive and mixing movements; (2) secretion of digestive juices; (3) digestion of nutrients; and (4) absorption of nutrients. This chapter describes each of these functions in detail and provides an overview of the structure and organization of the GI tract and its growth and alteration across the life span.

STRUCTURE AND ORGANIZATION OF THE GASTROINTESTINAL TRACT

Embryology

As early as the third week of gestation, the embryonic structure begins to fold inward to create the primitive gut. By the fourth week structures of the foregut, midgut, and hindgut are in place.

The foregut develops into the pharynx, the esophagus, the stomach, the duodenum proximal to the opening of the common bile duct, the hepatobiliary system, and the pancreas. The midgut forms the small intestine (below the opening of the common bile duct), the cecum, the

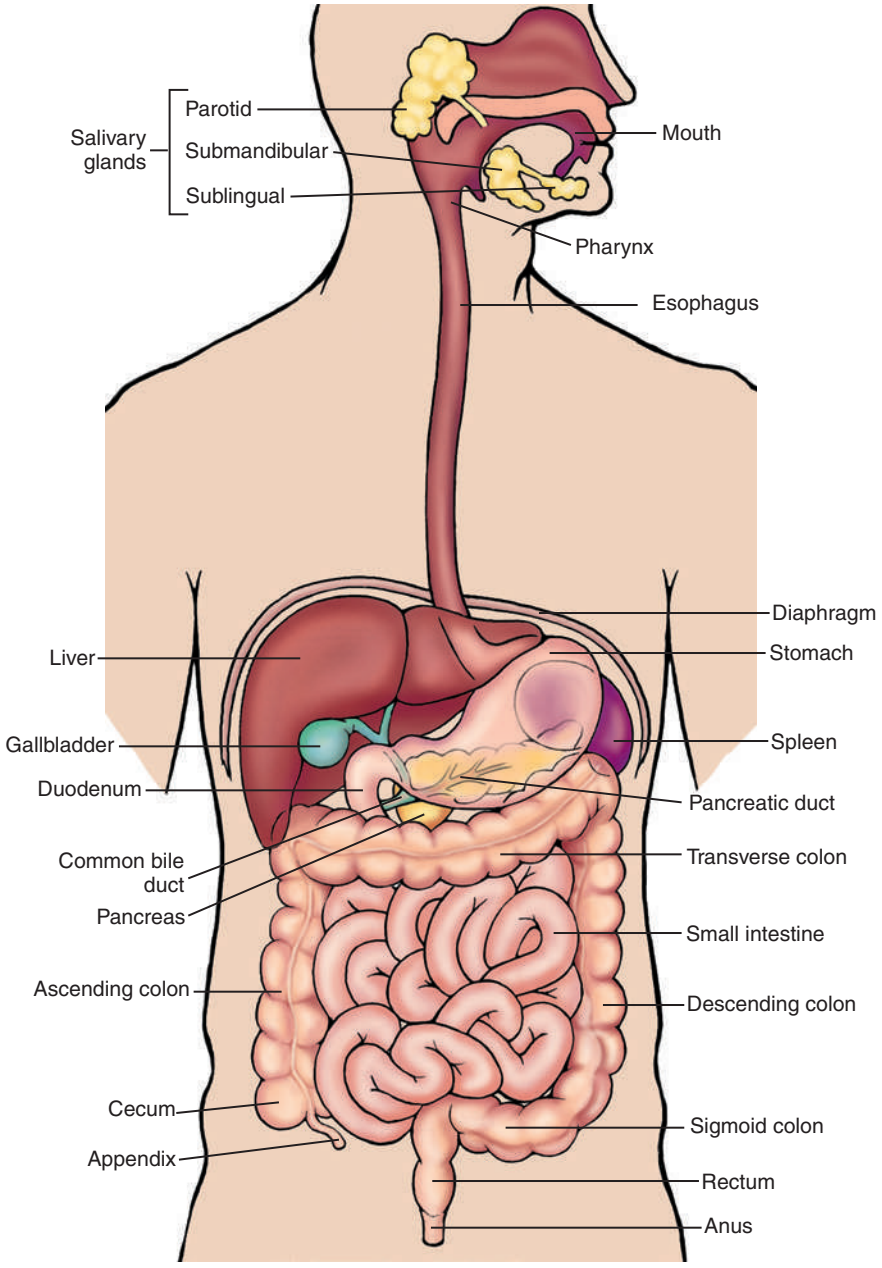


FIG 35.1 The gastrointestinal system. (From Monahan FD, Neighbors M: *Medical surgical nursing: foundations for clinical practice,* ed 2, Philadelphia, 1998, Saunders, p 950.)

appendix, the ascending colon, and the proximal portion of the transverse colon. The hindgut develops into the distal part of the transverse colon, the descending and sigmoid colon, the rectum, and the superior portion of the anal canal.

Developmental abnormalities are relatively common as a result of incomplete partitioning during gut differentiation. The trachea and esophagus share a common developmental origin, and incomplete development of these structures may lead to tracheoesophageal fistula (TEF), an anomaly characterized by an abnormal connection between the trachea and esophagus (often accompanied by esophageal atresia (EA), where the esophagus is closed off in a blind pouch). The fistula tract is thought to occur when a branch of the embryonic lung bud fails to undergo branching because of defective epithelial-mesenchymal interactions. These disorders occur in about 1 in 3500 live births, with about one-third of affected infants born prematurely, and there are associated anomalies in about half of the cases of TEF and EA. One of the more serious surgical emergencies in newborns, TEF requires immediate diagnosis and surgical correction. The current prognosis for such infants is very good, with about 90% survival beyond the neonatal period.

Failure of normal partitioning between the foregut and midgut can lead to intestinal atresia (IA), a condition in which the lumen of the small intestine is obliterated. Congenital IA may also occur as the result of an ischemic event, in which the necrotic area of gut is resorbed by the developing fetus, leading to a blind pouch and area of discontinuity. One potential etiology for such events is maternal use of vasoconstrictive drugs such as methamphetamine and pseudoephedrine, as well as maternal smoking. IA is an important cause of intestinal obstruction in the newborn and necessitates prompt surgical correction. Failure of the midgut to develop or rotate properly with respect to the umbilical cord can result in omphalocele, a congenital herniation of viscera into the base of the umbilical cord, which may require surgical correction as well.

Congenital malformations resulting from inappropriate development of the anorectal portion of the GI tract include colonic or anal agenesis, a condition in which the rectal pouch ends blindly. Other developmental anomalies of this portion of the GI tract include anal stenosis, in which the anal aperture is small, and anal atresia (or imperforate anus), in which the anal membrane persists and covers the aperture to create an obstruction.

After its initial embryologic development, the GI tract continues to grow in length and diameter until somatic growth ends with puberty. General factors such as adequate nutrition and appropriate levels of insulin, growth hormone, thyroid hormone, cortisol, androgens, and estrogens play a role in GI development, as well as local factors such as the direct effect of ingested nutrients, GI hormones, and secretions.

Functional Anatomy

Each part of the GI tract is uniquely adapted for a specific function in providing nutrients for the body. The role of each major component of the GI tract will be described in some detail.

Oral Cavity and Pharynx

The mouth, or oral cavity, is the usual point of entry for nutrients and is the site of the initial breakdown of nutrient substances into a form usable by the body. Food is pushed toward the side of the mouth by the tongue to facilitate chewing and grinding on the surfaces of the molar and premolar teeth. As the food is manipulated and broken down, it is moistened by saliva secreted by three major pairs of salivary glands: the parotid, submandibular, and sublingual glands (Fig. 35.2). Saliva serves three major functions: (1) through its moistening action, saliva allows the tongue to convert a mouthful of food into a bolus, or semisolid mass, that can be swallowed easily; (2) it allows for taste perception by the papillae on the surface of the tongue, which are sensitive to chemical differences among dissolved food molecules; and (3) the digestive enzyme contained in saliva, salivary amylase (also called *ptyalin*), initiates carbohydrate digestion by breaking down polysaccharides (also called *starch*) into the simpler molecular structures of dextrin and maltose. Important changes in oral structure and function that occur with aging are detailed in Geriatric Considerations: Changes in the Mouth.

The pharynx, or throat, is about 12 cm long and serves as the entryway for both the respiratory and the GI systems. The oropharynx, the portion of the pharynx posterior to the mouth, is separated from the nasopharynx, the portion of the pharynx posterior to the nose, by the soft palate. The laryngopharynx is the portion of the pharynx that opens into the larynx and the esophagus. During swallowing, the soft palate is pulled upward to close off the nasopharynx. The bolus of food being swallowed is propelled by reflex movements of muscles in the pharynx through the laryngopharynx and into the esophagus. Simultaneously, the opening to the larynx is closed by the epiglottis. This coordinated set of actions prevents food substances and liquids from inadvertently entering the respiratory system, a potentially life-threatening occurrence referred to as *aspiration*.

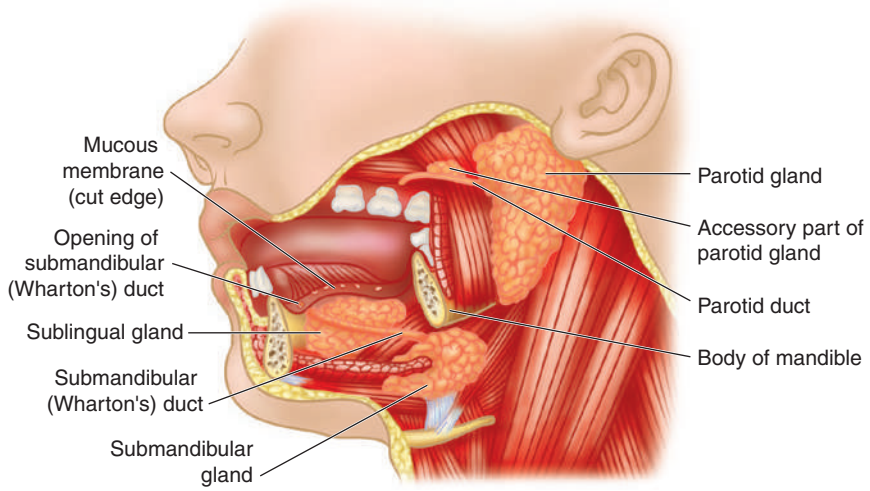


FIG 35.2 Oral cavity. Structures of the mouth and location of the salivary glands. (From O'Toole M, editor: *Miller-Keane encyclopedia and dictionary of medicine, nursing, and allied health,* ed 7, Philadelphia, 2003, Saunders.)

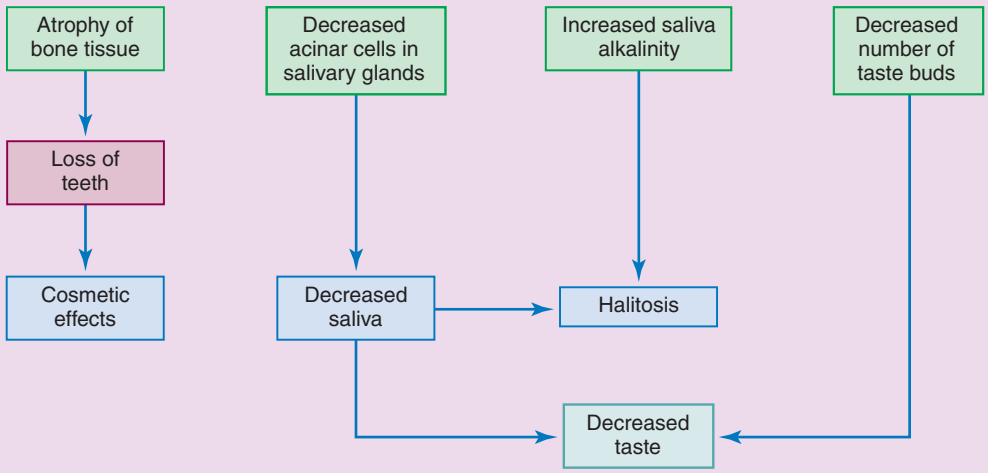
GERIATRIC CONSIDERATIONS

Changes in the Mouth

Elderly people experience a decline in taste. This decline is due to both an increase in the sensation threshold for all four tastes and a decrease in the number of papillae. For example, children have more than 200 taste buds, whereas the elderly have fewer than 100. Of the four basic tastes, elderly people experience a particular decrease in salt and sugar tastes.

Older individuals also experience a decrease in the number of acinar cells in the salivary glands, leading to a reduction in salivary secretion. These changes can contribute to halitosis (bad breath).

Loss of teeth in the elderly is due to atrophy of gum and bone tissue as well as actual tooth deterioration. As tooth enamel is destroyed, dentin is exposed, allowing development of caries (cavities). Gingival epithelial loss may occur as a result of pathologic processes as well as normal aging.



Esophagus

The esophagus is a muscular tube approximately 25 cm in length that initiates the progress of food through the gut after ingestion. Passage of food through the esophagus is greatly facilitated by mucus secreted by cells in the epithelial lining. Extremely rough or fibrous foods may potentially penetrate the mucous lining of the esophagus and cause damage. The stratified squamous epithelium lining the esophagus normally repairs itself quickly after injury and is constantly renewed by cells moving to the surface from below, though bones in foods such as fish or chicken may cause serious penetrating injuries.

The esophagus propels nutrients to the stomach by means of strong muscular contractions. (Presbyesophagus, or abnormal esophageal motility occurring with advanced age, is described in detail in the Age-Related Changes section.) When the body is in an upright position, gravity assists in the downward movement of food to the stomach. However, the muscular contractions of the esophagus are strong enough to transport nutrients to the stomach even in the absence of gravity, as persons living (and eating) in the weightless conditions of space have demonstrated.

At the lower end of the esophagus, about 2 to 5 cm above its juncture with the stomach, the circular muscle of the esophagus forms the lower esophageal sphincter (LES). Although anatomically this sphincter is no different from the remainder of the esophagus, it remains tonically constricted, in contrast to the middle and upper portions of the esophagus, which are completely relaxed under normal conditions. Thus the LES serves to prevent the highly acidic gastric contents from moving in a retrograde motion (refluxing) back into the esophagus. Under certain conditions the LES does not function properly, and reflux of gastric contents into the esophagus—gastroesophageal reflux disease (GERD)—may occur. The resulting subjective sensation of irritation and spasms of the distal portion of the esophagus is often referred to as *heartburn* or *dyspepsia*.

Stomach

The stomach (Fig. 35.3) is essentially an elastic food reservoir. Under normal circumstances, its capacity is 1000 to 1500 mL, although a capacity of as much as 6000 mL is possible. The portion of the stomach immediately below the LES is called the *cardia*. The fundus is the part of the stomach that continues lateral to and above the cardia; the body of the stomach extends from the cardia to the antrum, which stretches from the angulus to the pylorus. The antrum differs markedly from the rest of the stomach in function and is distinguished by the absence of rugae, the folds present in the mucous membrane of the other areas of the stomach. The pylorus is a muscular sphincter between the stomach and duodenum that serves to control gastric emptying and limits the reflux of bile from the small intestine.

The stomach is lined with simple columnar epithelium containing millions of gastric glands that extend down to the mucosa. A typical gastric gland is shown in Fig. 35.4. As shown in this illustration, gastric glands are lined by several types of specialized cells. Chief cells produce pepsinogen, the inactive form of the enzyme pepsin; parietal cells produce hydrochloric acid and also a substance called *intrinsic factor (IF)* that is needed for intestinal absorption of vitamin B12. Mucous cells produce a layer of alkaline mucus more than 1 mm thick that serves to shield the stomach wall and neutralize the acidity in the immediate area of the lining. In addition to these cells, gastrin cells located in the antral epithelium have surface microvilli that monitor

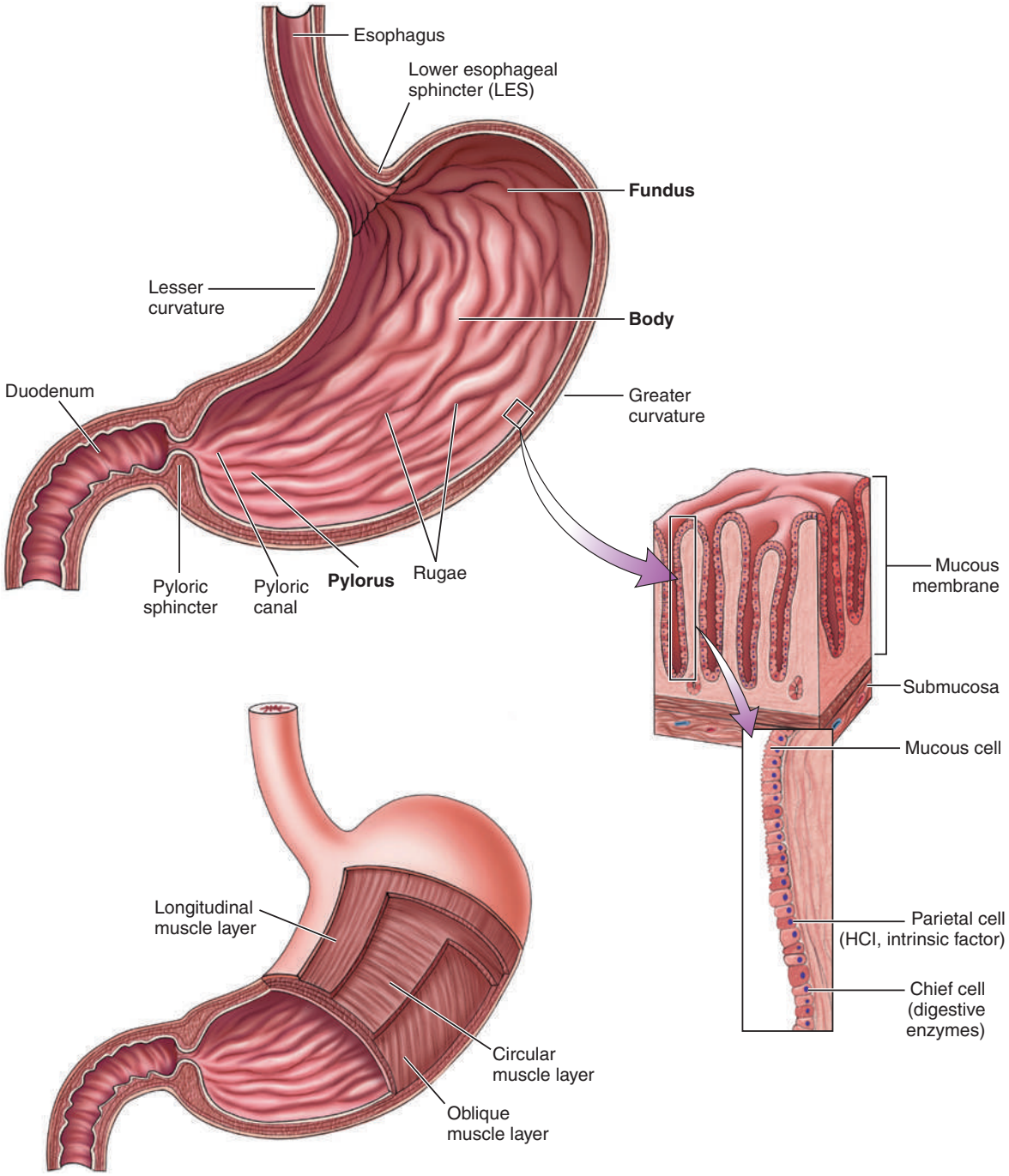


FIG 35.3 Physiologic anatomy of the stomach. *HCl,* Hydrochloric acid. (From Herlihy B: *The human body in health and illness,* ed 4, Philadelphia, 2011, Saunders.)

intragastric pH. The role of these cells and the substances they secrete in the digestion of nutrients are described in detail in the Secretory Function section.

Small Intestine

The small intestine of an adult is approximately 5 to 6 m long (the longest portion of the GI tract). The first 22 cm of the small intestine is called the *duodenum;* the jejunum constitutes the next 2 m, and the ileum forms the remainder. The entire inner wall of the small intestine is marked by circular folds of a mucous membrane called the *plicae circulares,* permanent ridges that do not lose their elasticity when the intestine is distended.

On microscopic examination, the lining of the small intestine contains millions of fingerlike projections called *intestinal villi* (Fig. 35.5). Like the circular folds just described, these villi serve to increase the surface area of the intestine for digestion and absorption of nutrients. Each villus has its own microscopic projections called *microvilli,* which in turn are covered by a fuzzy coat (called the *brush border* because of its brushlike appearance when viewed with an electron microscope) containing many digestive enzymes. The combined effect of the circular folds,

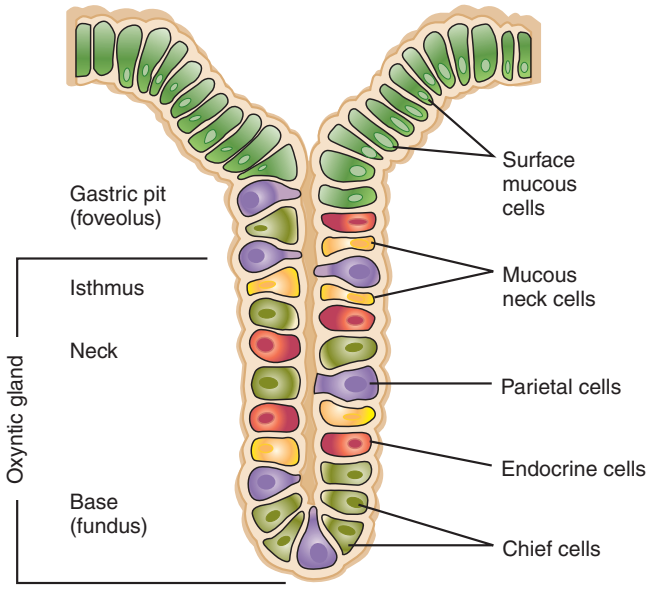


FIG 35.4 Gastric mucosa and gastric glands.

villi, and microvilli is to increase the surface area of the small intestine by about 600 times, thereby creating a remarkably efficient interface for nutrient digestion and absorption. Fig. 35.6 shows a microscopic section of the small intestine.

Between the villi are situated the intestinal glands, or crypts of Lieberkühn. The intestinal glands secrete about 2 L of fluid daily into the lumen of the intestine, most of which is quickly reabsorbed by the villi. Goblet cells throughout the intestinal mucosa secrete large amounts of mucus. In addition, specialized mucous glands located in the first few centimeters of the duodenum, called *Brunner glands,* release a thick coating of mucus to protect the mucosa of this intestinal region from the potentially damaging effects of acidic gastric juice entering through the pylorus.

Although the details of the process of digestion and absorption of nutrients in the intestinal mucosa will be covered in greater detail in subsequent sections, a unique and salient feature of the villus epithelial cells is described here. Villus epithelial cells have both digestive and absorptive functions, apparently dependent on their current stage of maturation. The rapidly dividing cells at the base of the intestinal glands are responsible for secretion, but as they migrate to the villus, they mature into absorptive cells and are eventually pushed out of the villus tip. Turnover of cells in the small intestine occurs in 48 to 72 hours, one of the fastest cell turnover rates in the body. Therefore conditions

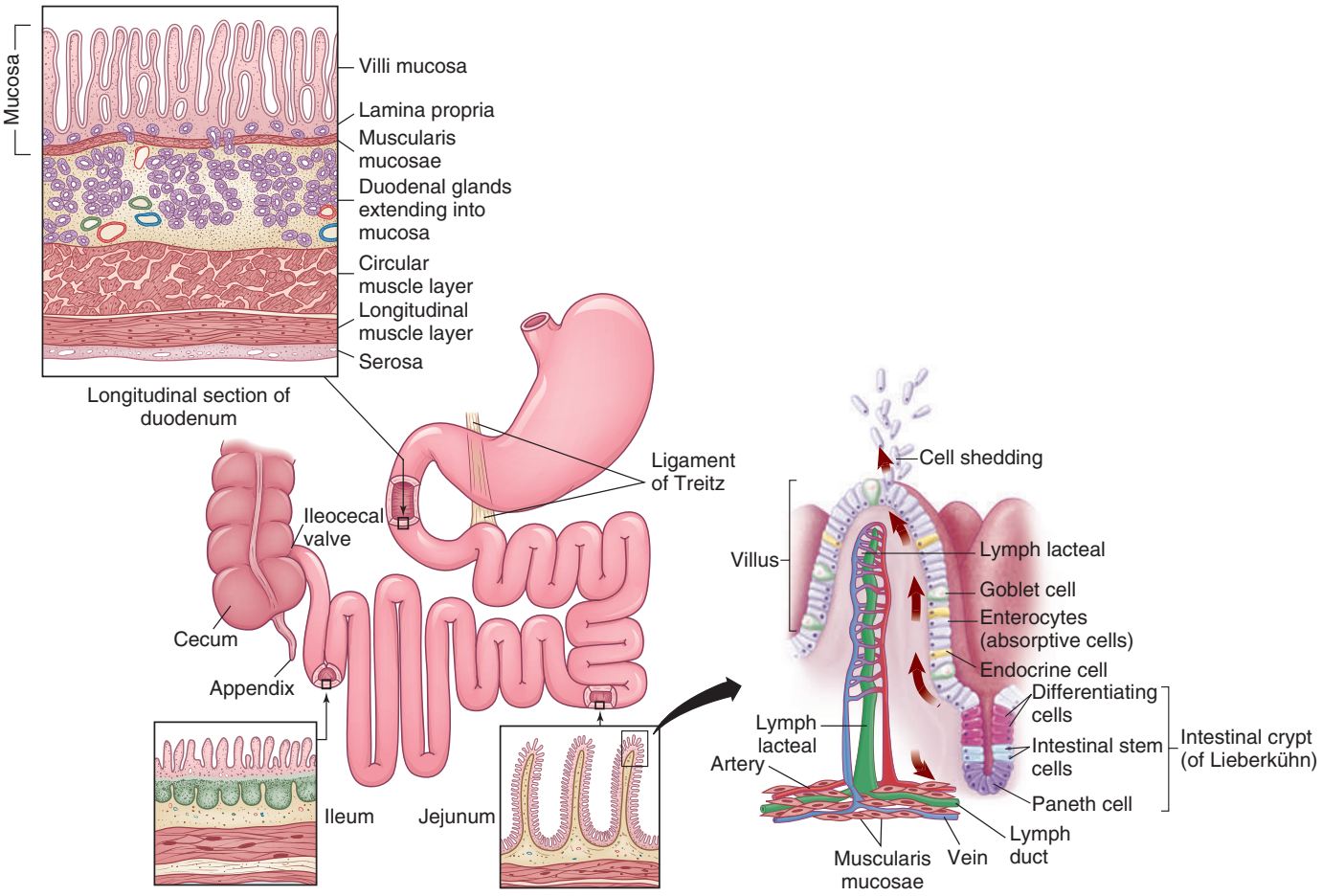


FIG 35.5 The small intestine. (From Huether S, McCance K: *Understanding pathophysiology,* ed 6, St Louis, 2016, Mosby.)

such as malnutrition or substances that interfere with cell replication or protein synthesis, such as chemotherapeutic agents, may severely compromise intestinal function.

The ileocecal valve, a sphincter between the small and large intestines, is normally closed so that the contents of the large intestine cannot reflux back into the small intestine. As peristaltic contractions move intestinal contents toward the large intestine, the ileocecal valve opens.

FIG 35.6 Microscopic section of the small intestine. (From Klatt EC: *Robbins and Cotran atlas of pathology,* ed 3, Philadelphia, 2015, Saunders.)

Large Intestine

The large intestine (Fig. 35.7) is a muscular tube 1.5 m long and 6.5 cm in diameter that forms a frame around the small intestine. The portion of the large intestine from the cecum to the rectum is known as the *colon*. The ascending colon extends from the cecum straight up to the lower border of the liver; the transverse colon then extends across the abdomen, anterior to the small intestine. The descending colon turns downward on the left side of the abdomen, finally becoming the S-shaped sigmoid colon, which empties into the rectum. The rectum has its outlet at the anus, the opening for elimination of feces (see Fig. 35.7).

The vermiform appendix, attached to the cecum, is a worm-shaped blind tube containing specialized lymphatic structures. It contains T and B lymphocytes, secretes immunoglobulin A, and contributes to gut-associated lymphoid function. Inflammation of the appendix, or appendicitis, is one of the most common catastrophic events involving the gut, which can quickly lead to peritonitis and death if not diagnosed and managed promptly.

The mucosa of the large intestine has no villi and does not produce digestive enzymes (Fig. 35.8). The epithelial surface of the colon consists of absorptive cells that predominantly absorb water and electrolytes. Mucus-producing goblet cells line the glandular crypts present in the surface epithelium. Endocrine cells are also present, perhaps helping coordinate colon neurologic activity, but at present the function of hormones in the large intestine is poorly understood. The turnover time of cells in the colonic mucosa is 3 to 8 days, comparatively longer than that of cells in the small intestine.

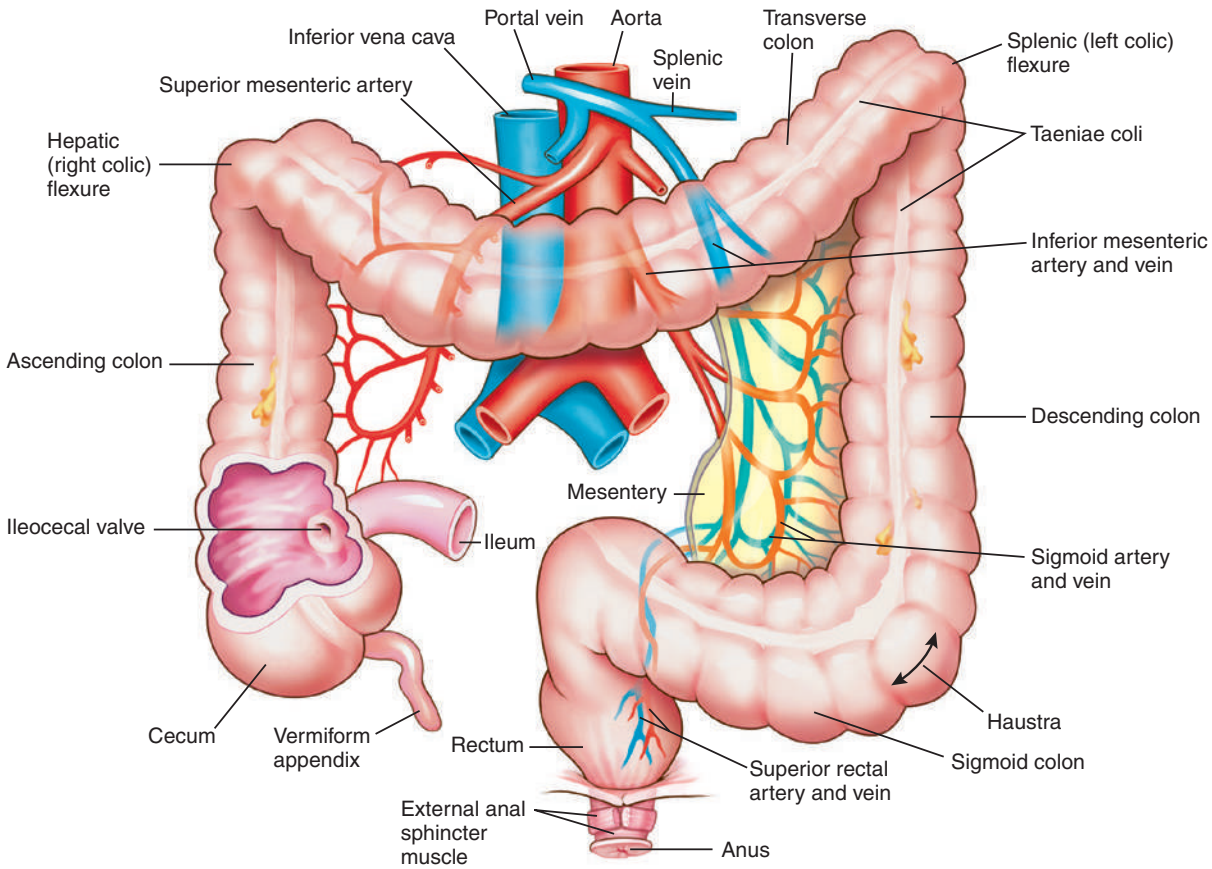


FIG 35.7 Divisions of the large intestine. (From Patton KT, Thibodeau GA: *Essentials of anatomy and physiology,* St Louis, 2012, Mosby.)

KEY POINTS

- Tracheoesophageal fistula, esophageal and duodenal atresia, and anal agenesis are congenital disorders that occur with abnormal development of the GI tract. These disorders are usually manifested as obstructions in the neonatal period.
- The major structures and corresponding functions of the GI tract can be summarized as follows:
- Mouth and salivary glands: Mastication, moistening, and the beginning of starch digestion (by the enzyme salivary amylase) of foodstuff.
- Pharynx: Transport of food to the esophagus and protection of the airway from aspiration of food particles.
- Esophagus: Movement of food to the stomach by peristaltic waves. The lower esophageal sphincter (LES) prevents reflux of stomach contents.
- Stomach: Reservoir for food, mixing, and initial digestion of proteins (by the enzyme pepsin); secretion of hydrochloric acid, intrinsic factor, and gastrin. The pyloric sphincter prevents reflux of intestinal contents.
- Small intestine: Digestion and absorption of nearly all nutrients in the duodenum and jejunum; absorption of bile salts in the terminal ileum. The brush border contains numerous digestive enzymes. The enzymes secretin and cholecystokinin are secreted by intestinal mucosa.
- Pancreas and gallbladder: The pancreas delivers digestive enzymes and bicarbonate to the duodenum. The gallbladder delivers bile salts to the duodenum.
- Large intestine: Reabsorption of water and storage of feces. Feces are delivered to the rectum for defecation.

GASTROINTESTINAL MOTILITY

The way in which nutrients and their eventual waste products are propelled through the GI tract is a complex and fascinating process, involving an exquisitely timed set of autoregulatory actions and responses. A summary of the characteristics of the intestinal wall, innervation of the gut, and hormonal control of GI motility will provide a basis for understanding how nutrients move through the GI tract.

Characteristics of the Intestinal Wall

A typical cross-section of the intestinal wall is depicted in Fig. 35.9. From the outer surface inward are five main layers: the serosa, a longitudinal muscle layer, a circular muscle layer, the submucosa, and the mucosa. A small layer, the muscularis mucosa, is located between the mucosa and submucosa. The muscular movements of the GI tract are performed mostly by the different layers of smooth muscle, which extend from the distal end of the esophagus through most of the large intestine. However, skeletal muscle has a key role in motility at both ends of the GI tract; motility from the mouth through the proximal portion of the esophagus at the upper end and through the external sphincter of the anus at the lower end is mediated by the action of skeletal muscle.

The general characteristics of smooth muscle are covered in Chapter 5. Two specific characteristics of smooth muscle in the gut that enable its unique function are the close proximity of these fibers to each other and the ability of these cells to work in an integrated fashion. In most areas of the GI tract, smooth muscle fibers are extremely close; about 12% of their membrane surfaces are actually fused with the membranes of other adjacent muscle fibers to form a nexus, or junction. This allows intracellular current to travel very easily from one muscle fiber to another. Moreover, action potentials originating in one smooth muscle fiber in the GI tract are generally propagated from one fiber to another; therefore the GI tract acts as a functional syncytium, where separate cells have the ability to function in a unified manner.

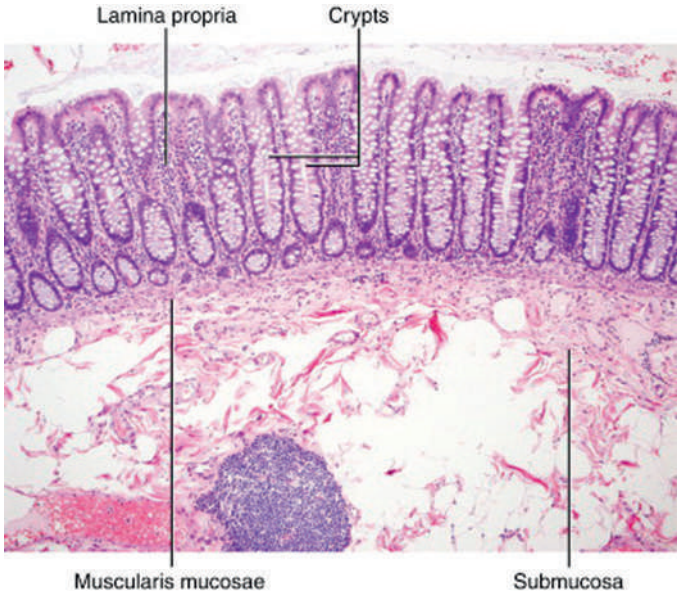


FIG 35.8 Normal colon histology. Mucosal crypts are aligned parallel to one another "like a row of test tubes." Epithelium on the surface and lining the crypts consists of absorptive tall columnar cells and goblet cells. Lamina propria invests the crypts and contains fibroblasts, macrophages, neuroendocrine cells, plasma cells, lymphocytes, eosinophils, and mast cells. A thin but distinct layer of smooth muscle (muscularis mucosae) separates mucosal elements from the submucosal space. The submucosa contains neural plexuses, fat, blood vessels, and lymphatic vessels. The muscularis externa is composed of an inner circular and an outer longitudinal layer of smooth muscle. (From MacLennan GT: *Hinman's atlas of urosurgical anatomy,* ed 2, Philadelphia, 2012, Saunders.)

Neural Control

Movement of nutrients through the GI tract is controlled by the central nervous system through its autonomic division and is modulated by numerous hormonal interactions. In addition, the GI system has an intrinsic nervous system of its own affecting most GI functions. The intrinsic nervous system is composed of two layers: (1) the myenteric, or Auerbach, plexus, which lies between the longitudinal and circular muscular layers; and (2) the submucosal, or Meissner, plexus, which lies in the submucosa. The myenteric plexus is largely responsible for control of GI movements; the submucosal plexus serves to control secretion and is also involved in many sensory functions, with information being received from the gut epithelium and stretch receptors in the intestinal wall. The entire intrinsic nervous system, including both the myenteric plexus and the submucosal plexus, is responsible for many reflexes that occur locally in the GI tract, such as the localized secretion of digestive juices by the submucosal glands or an increase in gut smooth muscle activity.

In general, when the myenteric plexus is stimulated, activity in the GI tract increases. This stimulation has four principal effects: (1) tonic contraction of the intestinal wall increases; (2) rhythmic contractions increase in intensity; (3) rhythmic contractions increase in rate; and (4) the velocity of conduction of excitatory waves along the intestinal wall increases. As part of the parasympathetic nervous system, these excitatory fibers of the myenteric plexus are primarily cholinergic (i.e., secrete acetylcholine), in addition to one or more other excitatory transmitter substances. However, some myenteric plexus fibers have an inhibitory effect and may secrete purine-based transmitter substances such as adenosine triphosphate (ATP).

Input from the sympathetic and parasympathetic nervous systems can strongly affect the activity of the intrinsic nervous system. In general,

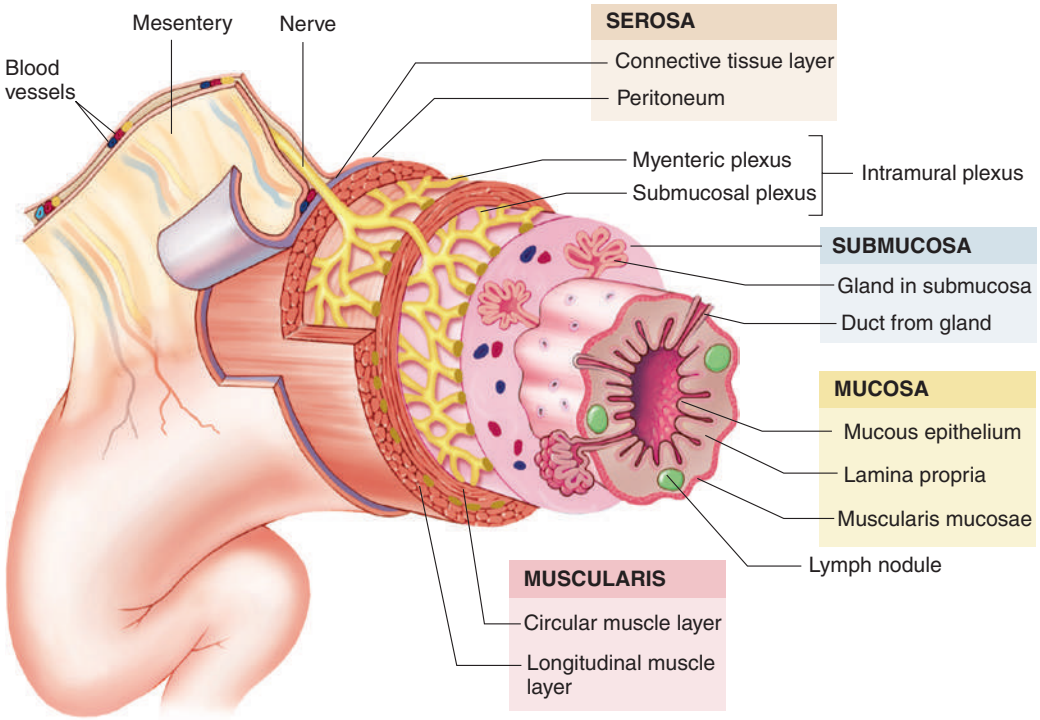


FIG 35.9 Cross-section of a typical segment of the intestinal wall showing the four principal layers and associated structures: mucosa, submucosa, muscularis, and serosa. Although different areas of the GI tract specialize in function, the anatomy of the wall is similar in structure. (From Patton KT, Thibodeau GA: *Essentials of anatomy and physiology,* St Louis, 2012, Mosby.)

sympathetic stimulation decreases the activity of the intrinsic nervous system, whereas parasympathetic stimulation increases its activity.

Parasympathetic Innervation

The parasympathetic supply to the GI tract is divided into cranial and sacral divisions. Cranial parasympathetic stimulation is transmitted almost entirely by the vagus nerves, which provide extensive innervation to the esophagus, stomach, pancreas, and the first half of the large intestine (with little innervation of the small intestine). The sacral parasympathetic division originates in the second, third, and fourth sacral segments of the spinal cord and innervates the distal half of the large intestine. The sigmoid, rectal, and anal regions of the large intestine are especially well supplied with parasympathetic fibers; these fibers have a key role in the defecation reflex.

Sympathetic Innervation

The sympathetic fibers that innervate the GI tract have their origin in the spinal cord between T8 and L3. After exiting the cord, the preganglionic fibers enter the sympathetic chains and then pass through these chains to various ganglia adjacent to the GI tract, such as the celiac ganglion and the mesenteric ganglia. From these locations, postganglionic fibers radiate out to all parts of the gut. These sympathetic fibers supply essentially all parts of the GI tract (in contrast to the concentration of parasympathetic innervation at locations close to the entry and exit points of the gut). The sympathetic nerve endings in the GI tract secrete norepinephrine, which has an inhibitory effect on the parasympathetic nervous system. Norepinephrine acts directly on smooth muscle in the GI tract to inhibit activity; in addition, norepinephrine has an inhibitory effect on the neurons of the intrinsic nervous system of the GI tract. Strong stimulation of the sympathetic nervous system can effectively stop motility in the gut and therefore block the movement of nutrients through the GI tract. GI sympathetic activity can also initiate vomiting through a complex sequence of events mediated by various neurotransmitters.

Afferent Nerve Fibers

The GI tract is richly supplied with afferent nerve fibers arising from the gut that transmit important information about the status of the GI tract. Afferent fibers that have their cell bodies in the submucosal plexus and terminate in the myenteric plexus transmit signals in response to irritation of the gut mucosa, excessive distention, or the presence of specific chemical substances. These signals can result in excitation or, in some circumstances, inhibition of intestinal motility or secretion. Other afferent fibers with cell bodies in the dorsal root ganglia of the spinal cord or cranial nerve ganglia can transmit signals to higher levels of the central nervous system by traveling along sympathetic or parasympathetic pathways. For example, the vagus nerves contain many afferent fibers that transmit signals to the medulla; this information is then used to initiate and modulate vagal signals that control many important functions of the GI tract.

Electrical Activity of Gastrointestinal Smooth Muscle

Electrical activity is almost constantly present in the smooth muscle layers of the GI tract. Two basic types of electrical wave activity have been identified in the gut: slow waves and spikes (the latter named for the spiking appearance of these sudden increases in membrane potential). These two types of electrical wave patterns are shown in Fig. 35.10. Slow-wave electrical activity represents an ongoing tonic oscillation in membrane potential that occurs in the smooth muscle of the GI tract, especially in the muscle in the longitudinal layer. Normally, between 3 and 12 slow waves occur per minute, ranging from 40 to 50 millivolts (mV) in amplitude. Slow waves can be any degree of intensity and are

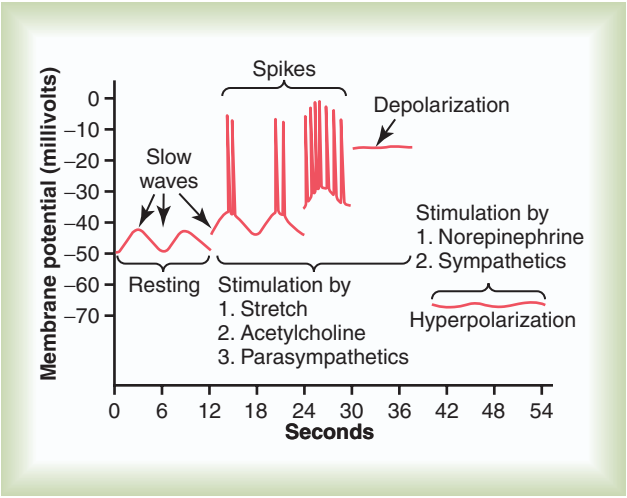


FIG 35.10 Membrane potentials in intestinal smooth muscle. (From Hall JE: *Guyton and Hall textbook of medical physiology,* ed 13, Philadelphia, 2016, Saunders.)

not the "all-or-nothing" type of action potential seen in other smooth muscle fibers in the body. In contrast to these nearly continuous slow waves, spikes occur under certain circumstances. When the muscle layer in the GI tract is stimulated by being stretched or by the effects of acetylcholine or parasympathetic excitation, the intracellular resting membrane potential of the muscle fibers becomes relatively more positive. The entire potential level of the slow waves is raised—an effect called *depolarization*. As shown in Fig. 35.10, when depolarization rises above a certain level (around −40 mV), sudden increases in the membrane potential, or spikes, start to appear on the peaks of the slow waves. If the resting potential rises further, spikes appear more frequently. With very strong stimulation, the spikes generally disappear because the membrane now remains entirely depolarized. Fig. 35.10 also illustrates the response of smooth muscle fibers to stimulation by norepinephrine or sympathetic excitation. In this situation, the resting membrane potential is decreased, or hyperpolarized, and electrical activity is almost abolished.

Hormonal Control

Hormones play a pivotal role in controlling GI secretion, in many cases affecting several different portions of the GI tract. Gastrin, which is secreted by specialized endocrine cells (G cells) of the stomach mucosa in response to food entry, increases stomach motility and is the primary mediator of gastric acid secretion. In addition, it promotes increased constriction of the LES, which serves to prevent reflux of stomach contents into the esophagus. Gastrin may also have a small effect in increasing motility of the small intestine and gallbladder.

Cholecystokinin (CCK), which is secreted mainly by I cells of the jejunum in response to the entry of fatty substances, has an extremely strong effect on gallbladder contractility. This stimulation of gallbladder activity results in an outpouring of bile, which plays an important role in fat digestion and absorption. CCK also stimulates pancreatic secretion, helps regulate gastric emptying and bowel motility, and induces satiety.

Secretin, which is produced by the mucosa of the duodenum in response to the entry of acidic gastric juice from the stomach, stimulates pancreatic fluid and bicarbonate secretion, with the effect of neutralizing the acidity of intestinal contents. It also has a mild inhibitory effect on motility in most of the GI tract.

Other important GI polypeptide hormones include vasoactive intestinal polypeptide, glucagon, glucose-dependent insulinotropic polypeptide (also called *gastric inhibitory polypeptide,* or GIP), somatostatin, motilin, leptin, and ghrelin. Somatostatin analogs (e.g., octreotide) have been used therapeutically to treat some forms of diarrhea, endocrine tumors, and parietal hemorrhage and to reduce fluid output from pancreatic fistulas.

Ghrelin is a recently discovered peptide secreted by the stomach that increases appetite, stimulates growth hormone secretion, and produces weight gain. It helps regulate mealtime hunger and meal initiation, and as such has been the focus of intense research as a mediator for obesity, with the potential for a therapeutic antiobesity drug that would block ghrelin's effects on the body.

Movement in the Gastrointestinal Tract

Contraction of Gastrointestinal Smooth Muscle

In general, most contraction in the GI tract occurs in response to spike potentials; slow waves without superimposed spikes ordinarily do not lead to contraction. Spike potentials occurring in GI smooth muscle are analogous to action potentials in cardiac muscle and are responsible for the membrane changes that initiate contraction. As calcium enters the cell membrane and passes to the interior of the smooth muscle, it initiates a reaction between actin and myosin, a process described in detail in Chapter 17.

The electrical activity occurring in the smooth muscle of the gut develops into tonic contractions and rhythmic contractions, both of which occur in most types of smooth muscle. Tonic contraction is continuous, instigated by pacemaker cells that reside at the interface between the longitudinal and circular muscle layers. The intensity of tonic contraction varies with the frequency of spike potentials and determines the amount of pressure in that segment. Thus the degree of contraction exerted by the pyloric, ileocecal, and anal sphincters serves to regulate the movement of nutrients through the GI tract.

The degree of rhythmic contraction varies in different parts of the GI tract. These differing rhythmic frequencies are dependent on the rate of slow wave activity in a particular segment and may occur at rates of 3 to 12 times per minute. These slow wave–dependent contractions are responsible for the mixing and peristaltic propulsive movements present in the GI tract.

Two types of muscular activity are involved in the digestive and absorptive functions of the GI tract: mixing movements and propulsive movements. In different portions of the GI tract these movements may serve different functions to achieve proper digestion and absorption of nutrients. For example, mixing movements in the stomach and small intestine promote digestion by mixing the digestive juices with the food that enters from the esophagus. In the small intestine and proximal segment of the large intestine, mixing movements facilitate absorption by exposing newly arrived intestinal contents into contact with absorbing surfaces. In the case of propulsion, the rate at which nutrients are propelled through the GI tract depends on the function of the different organs of the tract. For example, the passageway for nutrients from the mouth through the pharynx and esophagus is simply a conduit; essentially no digestive or absorptive function occurs there, and the transit of nutrients is quite rapid. In contrast, transit from the stomach and through the small and large intestines is quite slow. This slow rate of passage allows for completion of the digestive and absorptive processes that occur in these portions of the GI tract.

Although the characteristics of mixing and propulsive movements differ in various parts of the GI tract and will be described separately in the next section, a description of the general characteristics of these movements is presented here.

Propulsive Movements

The basic propulsive movement of the GI tract is called *peristalsis* (Fig. 35.11). Nutrients are propelled by the slow advancement of a circular

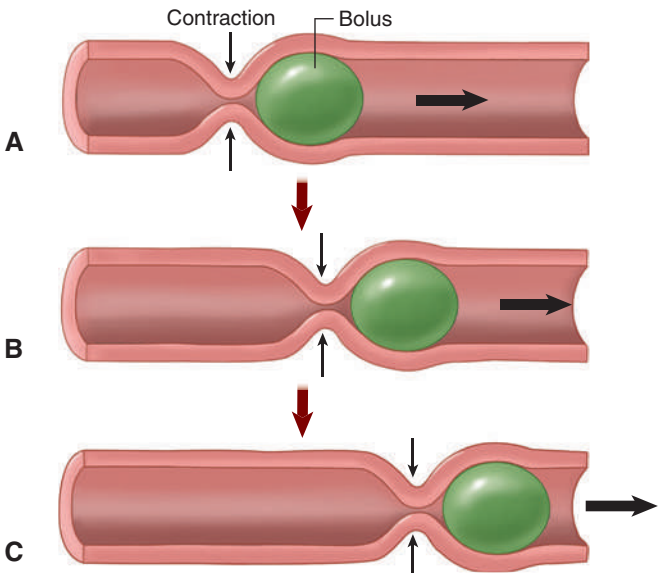


FIG 35.11 Peristalsis is a progressive type of movement, propelling material from point to point along the GI tract. **A,** A ring of contraction occurs where the GI wall is stretched, pushing the bolus forward. **B,** The moving bolus triggers a ring of contraction in the next region, which pushes the bolus even farther along. **C,** The ring of contraction moves like a wave along the GI tract, pushing the bolus forward. (From Patton KT, Thibodeau GA: *Essentials of anatomy and physiology,* St Louis, 2012, Mosby.)

constriction that squeezes the materials in front of the constricted area forward. Peristalsis is an inherent property of any smooth muscle tube that, like the intestine, is a functional syncytium. However, effective intestinal peristalsis requires the presence of an intact myenteric nerve plexus. The usual stimulus for peristalsis is distention of the intestinal walls. The entry and subsequent stretching of the intestinal wall by a bolus of food will have the effect of stimulating the gut wall 2 to 3 cm above this point, and a circular constriction will then occur and propel the food with a peristaltic movement. Although peristalsis can move in both directions in the gut, it normally moves toward the anus. It is thought that the myenteric plexus is organized in such a way that preferential transmission of signals downward occurs simultaneously with relaxation of the distal portion of the intestine below the distended stimulus point.

Mixing Movements

Segmental contractions serve to keep the intestinal contents thoroughly mixed on a constant basis. These movements may vary according to the specific function of each portion of the GI tract (see the Secretory Function section).

Movement of Nutrients

The path taken by foods ingested into the GI tract, as these nutrients travel down the tract and are digested and absorbed, will be traced beginning with the mouth. Although this process is described here as a linear sequence, it is important to note that several steps may occur simultaneously. The individual steps involved in nutrient ingestion constitute a synergistic process, and an inability to perform one phase of the process will ultimately have a profound effect on the entire GI tract. In addition, individuals manifest a great deal of variability in such aspects of digestive function as tolerance of certain nutrients and defecation patterns. Such variations may represent the influence of diet over a long period, conditioned responses to environmental cues, or age-related changes.

Chewing

The entry of solid food into the mouth results in the action of chewing, an important first step in the process of nutrient digestion. The process of moving the food around in the mouth and mixing it with saliva results in stimulation of the taste buds and olfactory epithelia; this sensory input greatly increases the subjective enjoyment of eating. As the food is mixed with saliva, it becomes softened and formed into a mass of appropriate size (bolus) that can be swallowed. The action of the molars and premolars in crushing more rigid forms of foods serves to prepare rough substances for transport down the esophagus. Although the act of chewing is under voluntary control, it is also partly reflexive in nature. The entry of food into the mouth has been shown to stimulate chewing in animals in the absence of full cerebral function. The movements of the skeletal muscles responsible for chewing are coordinated by impulses traveling through cranial nerves V, VII, IX, X, XI, and XII. Interruption of the proper transmission of impulses through these nerve tracts places an individual at risk for decreased voluntary control of the chewing function, with a resultant risk of aspiration (improper entry of oral or esophageal contents into the airways).

Swallowing

Swallowing is the transport of material from the mouth to the stomach. The process of swallowing has been divided into three stages that describe the regions through which the bolus of nutrients passes on its way to the stomach: (1) the oral stage, (2) the pharyngeal stage, and (3) the esophageal stage.

During the oral stage, the bolus is passed from the mouth to the pharynx through the space called the *fauces*. The bolus, either solid or liquid, is rolled toward the back of the tongue, and the front of the tongue is then pushed up against the hard palate. Respiration is inhibited briefly in this phase, while the pharyngeal muscles constrict to force the bolus of food into the pharynx. In the pharyngeal stage the bolus is passed through the pharynx into the esophagus, a process taking about 0.2 seconds. Continued contraction of the pharyngeal muscles and the position of the tongue prevent reentry of the bolus into the oral cavity. The soft palate is pulled upward to block the nasopharynx; simultaneously, food is prevented from entering the larynx by elevation of the larynx and approximation of the vocal cords, both actions serving to close the glottis. As these openings are blocked, the pharyngeal constrictors contract and force the bolus of food into the esophagus. Respiration is now resumed, and pressure in the pharynx rises as a result of the muscular activities that have occurred.

The muscular characteristics of the esophagus are of particular importance in effecting the third, or esophageal, stage of swallowing. The upper one-third of the esophagus consists of skeletal muscle, whereas the lower two-thirds consist of predominantly smooth muscle. In the normal resting stage, the upper part of the esophagus is closed by the tonic contraction of a band of skeletal muscle that serves as the pharyngoesophageal sphincter. The pressure exerted by the pharyngoesophageal sphincter in this region is normally about 20 to 40 cm H_2O above atmospheric pressure; this zone of high pressure keeps air from entering the esophagus during inspiration. Almost immediately after initiation of a swallow, the sphincter relaxes and pressure in the region drops to atmospheric pressure, thus allowing the bolus to be forced into the esophagus by the pressure generated in the pharynx. Pressure in the pharyngoesophageal junction region then rises as a result of contraction of skeletal muscle in this area, thus preventing reflux of food from the esophagus back to the pharynx. Pressure in this region then gradually subsides to a resting level while muscular relaxation occurs.

If the bolus being swallowed is a liquid, it is propelled through the esophagus by the initial force of swallowing and travels by gravity to the stomach in about 1 second. If the bolus is a semisolid mass, it is

propelled down the esophagus by means of a peristaltic wave. This esophageal peristalsis is caused by a contraction of circular muscle that forces the bolus ahead of it toward the stomach, with a transit time of about 4 to 6 seconds.

Although no well-differentiated muscular structure is located in the area where the esophagus joins the stomach, the region approximately 2 to 5 cm above the juncture with the stomach is referred to as the LES as described in the Esophagus section earlier. Almost immediately after initiation of a swallow, pressure at the LES drops and remains low during the time that a peristaltic wave is passing down through the lower end of the esophagus. Once the bolus has passed through the lower esophageal region and pressure in the lower portion of the esophagus has fallen to a resting level, the pressure in the LES rises and remains elevated for about 10 seconds before declining to a resting level once again.

Neural control of swallowing. Fig. 35.12 illustrates the neural pathways involved in the swallowing mechanism. Swallowing receptors in the posterior of the mouth and throat transmit impulses in response to a stimulus to the mucous membranes in the mouth, such as the presence of a moderate amount of fluid. These impulses travel mainly through the trigeminal nerve (cranial nerve V) into the reticular substance of the medulla oblongata, where the swallowing center is located. Once this center has been activated, the sequence of muscular reactions described earlier occurs automatically and usually cannot be voluntarily stopped. The swallowing center then sends impulses over a number of efferent nerves to the numerous skeletal and smooth muscles involved in the swallowing process to allow the complete act of swallowing to occur in the appropriate sequence. The glossopharyngeal (cranial nerve IX) and hypoglossal nerves (cranial nerve XII) are primarily concerned with the oral and pharyngeal stages, whereas the vagus nerve (cranial nerve X) is important in activating the esophageal stage.

The gag reflex is an important protective mechanism that aborts the normal swallowing response and helps eliminate potentially harmful ingestions. The afferent limb of the reflex is controlled by the glossopharyngeal nerve, and the efferent limb is regulated by the vagus nerve.

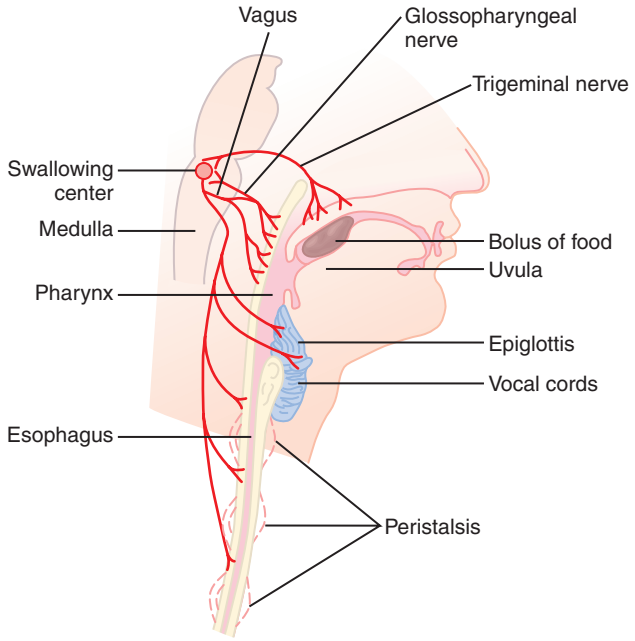


FIG 35.12 Neural pathways of the swallowing mechanism. (From Hall JE: *Guyton and Hall textbook of medical physiology,* ed 13, Philadelphia, 2016, Saunders.)

Disorders of the gag reflex such as stroke or neuropathy can lead to life-threatening obstruction or aspiration.

Motor Functions of the Stomach

The motor functions of the stomach include the storage of ingested nutrients for variable lengths of time and the discharge of gastric contents into the small intestine at an appropriate rate for optimal digestion and absorption. The stomach also aids in the digestive process by its mixing movements, which convert large pieces of food to a finer, liquid consistency.

Gastric filling and storage. On entering the stomach from the esophagus, newly arrived food forms concentric circles in the body and fundus of the stomach, with the most recently ingested food lying closest to the esophagus and older food lying closer to the stomach wall. The smooth muscle in the fundus and body of the stomach can adapt to the volume of contents so that relatively large contents can be introduced with little increase in intragastric pressure. The fundus and body of the stomach maintain a consistent pressure at all times. This tonic contraction continually presses on the food mass and aids in its delivery to the pyloric antrum.

Peristaltic contractions occur in the stomach once every 20 seconds. These rippling peristaltic waves begin in the corpus and move at a velocity of about 1 to 2 cm/sec. When they reach the more thickly walled pyloric antrum, they become much more vigorous and also increase in speed. These strong peristaltic contractions in the pyloric antrum are largely responsible for mixing ingested nutrients with gastric secretions. As ingested food is churned and mixed to a greater degree of fluidity, the mixture takes on a milky white sludge appearance and is then called *chyme*.

Emptying. As pressure in the antrum rises momentarily because of peristaltic contraction, a pressure differential exists between pressure in the antral pylorus and pressure in the duodenal bulb. The higher pressure in the antrum is sufficient to overcome the resistance of the pyloric sphincter, and the contents of the stomach are then propelled into the duodenum. Concurrently, the degree of constriction of the pyloric sphincter may increase or decrease, depending on several factors discussed in the next section. Because this process is dependent on the muscular activity of the antrum as well as the muscular tone of the pylorus, gastric emptying is largely regulated by mechanisms that affect each of these regions.

Regulation of gastric emptying. Factors that may affect the rate at which the stomach empties include the degree of distention of the gastric wall and the release of the hormone gastrin in response to certain types of food in the stomach. Both of these factors increase the rate of gastric emptying by increasing the force of antral contractions, while simultaneously inhibiting pyloric constriction. Distention of the gastric wall results in stimulation of mechanoreceptors in the stomach with subsequent activation of reflexes over the vagus and the intrinsic nerve plexuses. These neural influences, along with contractile activity as a direct response to the stretch of gastric muscle, constitute a major stimulus for gastric emptying. Gastric emptying time can vary depending on the food ingested (e.g., whether solid or liquid) and generally ranges from 1 to several hours.

Gastrin is released from the antral mucosa in response to stretching of the gastric wall, as well as the presence of certain foods, particularly meat. The role of gastrin in promoting the secretion of highly acidic gastric juices will be discussed later. With respect to stomach emptying, gastrin has a key role in enhancing peristalsis while at the same time relaxing the pylorus.

In addition to these influences, many of the mechanisms that affect gastric emptying are initiated in the duodenum. Reflex nervous signals are transmitted from the duodenum back to the stomach in

response to intraluminal stimuli; these signals likely help control both peristaltic activity and the degree of pyloric constriction. Stimulation of the duodenum in a variety of ways has the effect of slowing gastric emptying; both the chemical and the physical properties of chyme entering the duodenum may affect the rate of gastric emptying. A variety of both duodenal cells and duodenal receptors, including osmoreceptors, mechanoreceptors, and chemoreceptors, respond to intraluminal stimuli to produce hormonal and reflex inhibition of gastric motor activity and enhancement of pyloric tone. The presence in the duodenum of chyme containing the breakdown products of proteins, and to a lesser extent fats, may impede gastric emptying. Also, the presence of highly acidic or highly hypertonic or hypotonic chyme in the duodenum may inhibit the rate of gastric emptying. The degree of distention of the duodenum, as well as the presence of any degree of irritation of the duodenum, may also slow emptying of the stomach. These inhibitory mechanisms have a protective function and are effective in preventing the intestinal mucosa from overloading its digestive and absorptive abilities and potentially being damaged by chemical or mechanical sources.

Although regulation of gastric emptying is largely dependent on factors in the stomach and duodenum, gastric motility may be stimulated or inhibited reflexively from a variety of regions of the body. For example, stomach emptying is inhibited when the ileum is full and when the anus is mechanically distended. Stimulation of visceral and somatic pain receptors may result in inhibition of gastric motility. Various strong emotions such as anger, fear, and anxiety may produce changes in motility of the stomach, but whether these states tend to predispose an individual to inhibition or excitation of gastric motility is not always predictable.

Vomiting. Vomiting is rapid emptying of the contents of the stomach through the esophagus and into the mouth. The major force for vomiting is supplied by the skeletal muscle of the diaphragm and abdomen, rather than by contraction of the muscles of the stomach wall. Vomiting is the result of an extremely complex set of neural events coordinated by the nucleus tractus solitarius (NTS), a center located in the medulla. Afferent impulses from receptors in various regions of the body, including the sensory nerve endings of the pharynx, abdominal viscera, and the labyrinths, arrive at this center to initiate the vomiting reflex. This reflex causes closure of the glottis and trachea, relaxation of the gastroesophageal sphincter, and contraction of the diaphragm and the abdominal muscles, which forcibly expels the contents of the stomach.

Motility of the Small Intestine

After intact food entering the mouth has been liquefied and partially digested in the stomach, it enters the small intestine, where the major part of digestion and absorption occurs. As in other parts of the GI tract, movements of the small intestine can be described as propulsive and mixing movements, which in the small intestine generally occur simultaneously.

Propulsion. Chyme is propelled through the small intestine by peristaltic waves that move at a rate of 0.5 to 2 cm/sec, with a faster rate at the proximal part of the intestine and a slower rate in the terminal portion. Approximately 3 to 5 hours is normally needed for the passage of chyme from the pyloric sphincter to the ileocecal valve, but this period may vary in some disease states. Peristaltic activity in the small intestine is greatly increased after the ingestion of a meal. The increase in contractile activity in the stomach caused by distention of the stomach wall is conducted principally through the myenteric plexus down along the wall of the small intestine. This so-called *gastroenteric reflex* serves to increase the activity of the small intestine, with an enhancement of both intestinal motility and intestinal secretion.

The usual stimulus for peristalsis in the small intestine is distention of the intestinal walls; stretch receptors in the gut wall are sensitive to circumferential stretch and initiate a local myenteric reflex in response to this stimulation. The resulting contraction of longitudinal muscle, followed by the contraction of circular muscle, spreads downward in a peristaltic motion.

The peristaltic waves in the small intestine not only propel chyme downward toward the ileocecal valve but also spread the chyme along the intestinal mucosa, thus facilitating the process of absorption of nutrients (Fig. 35.13). As additional chyme enters the small intestine, this spreading process intensifies while peristalsis increases. When chyme reaches the ileocecal valve, it is sometimes stationary for several hours until the individual eats another meal and a new gastroenteric reflex intensifies the peristaltic process and propels the remaining chyme through the ileocecal valve.

Certain disease states, particularly those that involve intense irritation of the intestinal mucosa, may result in a peristaltic rush, a powerful peristaltic wave that travels long distances in the small intestine in a short period. The peristaltic rush clears the contents of the small intestine into the colon, thus relieving the small intestine of either irritating substances or excessive distention.

Mixing. In addition to propulsive peristaltic movements, a set of movements characterized as segmentation contractions occur in the small intestine. The primary effect of these contractions is progressive mixing of solid chyme particles with secretions of the small intestine. As their name implies, segmentation contractions involve contraction of the small intestine in regularly spaced segments that have the appearance of sausages (Fig. 35.14). As one set of segmentation contractions is completed, a new set begins, with contractile points located at different locations along the small intestine. Segmentation contractions occur at a rate of 7 to 12 times per minute and effectively chop and mix the chyme, as well as assist in propelling the chyme toward the ileocecal valve (see Fig. 35.14).

Control of motility. The electrical and mechanical activities of the small intestine are closely associated. Slow waves, as described previously in this chapter, occur at the membranes of the longitudinal smooth muscle, with frequencies of 11 to 12 per minute in the duodenum decreasing to 7 to 9 per minute in the terminal ileum. Slow waves do not directly produce muscular contractions in the small intestine, but provide the conditions under which contractions can occur. Although slow waves determine the velocity and direction of peristalsis, other factors determine whether action potentials and thus contraction will occur. Local mechanical and chemical stimulation by chyme is probably largely responsible for the initiation and continuance of contraction in the small intestine. Thus when the intestinal tract becomes overly distended or when the mucosa becomes irritated, myenteric reflexes enhance the electrical activity of the gut and spike potentials are superimposed on the slow waves. These spike potentials then spread through both longitudinal and circular muscle, and contraction results.

Intestinal motility may also be influenced by stimulation from sources extrinsic to the colon. Stimulation of the vagus nerve generally causes increased intestinal motility, with sympathetic stimulation resulting in inhibition. Intestinal motility can be altered reflexively by stimulation of many sensory areas. For example, trauma to organs outside the GI tract, such as irritation of the peritoneum or urinary tract, may cause intestinal inhibition. A condition called *paralytic ileus,* in which intestinal motility is inhibited as the result of reflex inhibition, may occur as a response to intraabdominal infection or surgery in this area.

Much current research is focused on the involvement of GI hormones in the regulation of GI tract motility. CCK, a hormone released from the mucosa of the jejunum in response to fatty substances in chyme, has been shown to block the increased gastric motility caused by gastrin. Another hormone, secretin, which is released mainly from the duodenal mucosa in response to gastric acid entering the duodenum, has the

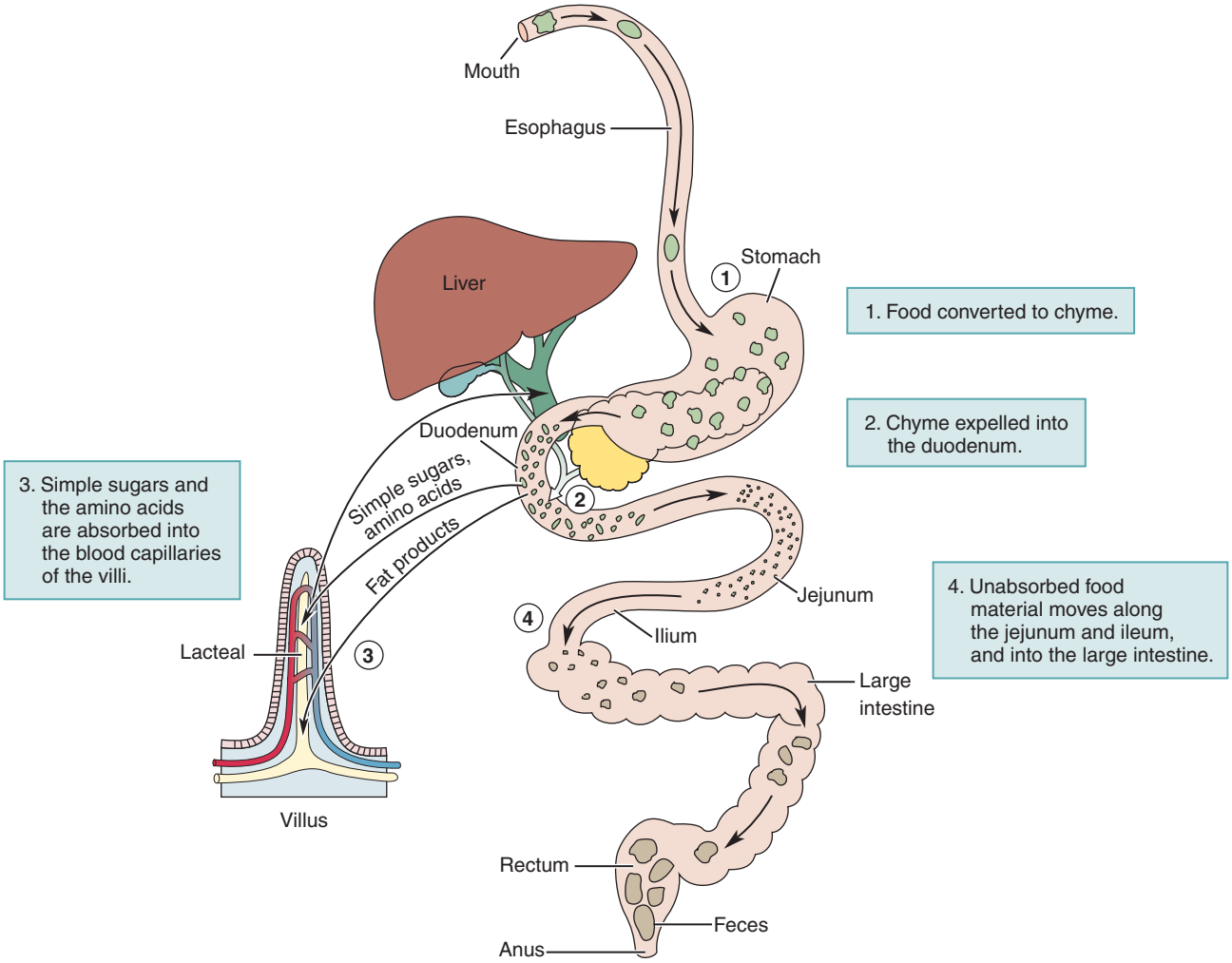


FIG 35.13 Chyme and the process of absorption of nutrients. (Modified from Herlihy B: *The human body in health and illness,* ed 4, Philadelphia, 2011, Saunders.)

general effect of decreasing GI motility. The hormone GIP, which is released from the upper portion of the small intestine in response to fat in chyme, as well as to carbohydrates, is known to inhibit gastric motility under some conditions. These hormones will be described in more detail in the Secretory Function section.

Ileocecal Sphincter

Chyme from the small intestine is eventually propelled downward to the terminal ileum immediately proximal to the cecum, where the last 2 to 3 cm of the muscular coat is thicker than that in the rest of the ileum. This region, called the *ileocecal sphincter,* has a high resting pressure (about 20 cm H_2O above atmospheric pressure) and is normally closed. Distention of the distal ileum lowers the pressure in the ileocecal sphincter. Thus when intestinal contents are present in the terminal ileum at sufficient quantity and are ready to be propelled into the cecum, the sphincter reflexively relaxes and the intestinal contents are pushed into the cecum by the propulsive movements of the distal small intestine. Conversely, distention of the cecum after it is filled with contents passing through the ileocecal valve results in increased pressure in the sphincter, which prevents reflux back into the ileum (Fig. 35.15).

Motility of the Colon

The movements of the colon are effective in promoting the two major functions of the colon: (1) absorption of water and electrolytes from chyme and (2) storage of the fecal mass until it can be expelled from the body by defecation.

Colonic movements. For most of the time, the large intestine in humans is inactive. However, the presence of material in the proximal end of the colon results in a type of mixing movement in the haustra (the outpouchings in the colon wall), termed *haustral churning,* that is similar to the segmenting movements in the small intestine. This movement is the major type of motility in the large intestine. Haustral churning exposes the contents of the large intestine to the mucosa, thus promoting the absorption of water. Normally, about 500 mL of chyme enters the proximal part of the colon each day. Out of this total volume, 400 mL—mostly water and electrolytes—is reabsorbed before defecation takes place, with an average volume of 100 mL of feces remaining for eventual disposal from the body.

At infrequent intervals of about three to four times a day, a strong peristaltic movement termed a *mass movement* occurs and propels the fecal material long distances. These strong contractions may reach a peak of 100 cm H_2O pressure in the segment undergoing the contraction. Fecal material may be transported all the way from the ascending colon to the descending colon by a mass movement. Feces are then stored in the distal end of the colon until defecation occurs.

Defecation. Under normal conditions, it takes about 18 hours for intestinal contents to reach the distal end of the colon after leaving the small intestine. Fecal material is stored in the distal part of the colon for

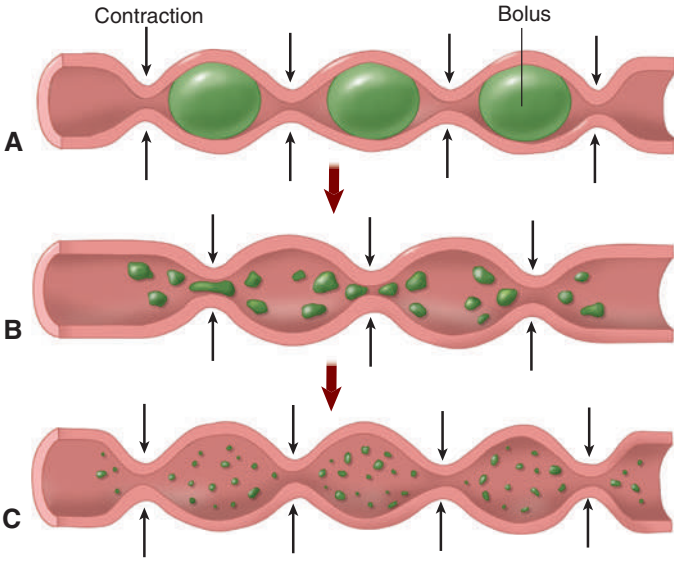


FIG 35.14 Segmentation. Segmentation is a back-and-forth action that separates chunks of food and mixes in digestive juices. **A,** Ringlike regions of contraction occur at intervals along the GI tract. **B,** Previously contracted regions relax and adjacent regions now contract, effectively "chopping" the contents of each segment into smaller chunks. **C,** Locations of the contracted regions continue to alternate back and forth, chopping and mixing the contents of the GI lumen. (From Patton KT, Thibodeau GA: *Essentials of anatomy and physiology,* St Louis, 2012, Mosby.)

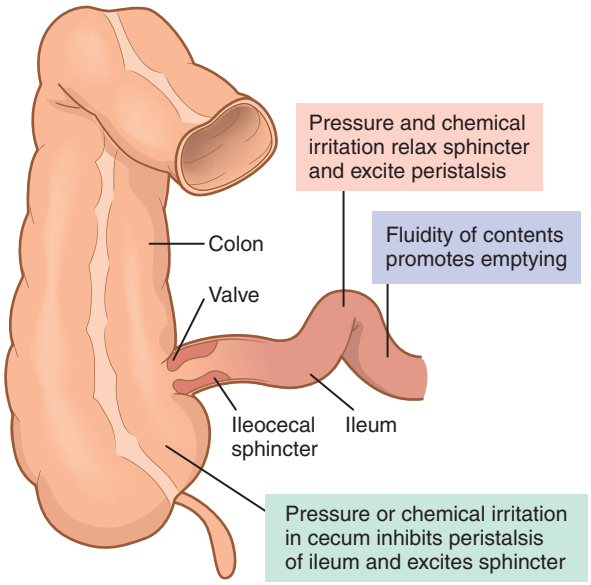


FIG 35.15 Emptying of the ileocecal valve. (From Hall JE: *Guyton and Hall textbook of medical physiology,* ed 12, Philadelphia, 2011, Saunders.)

varying lengths of time; defecation may take place 24 hours or longer after the ingestion of food. Ordinarily the rectum is empty, but fecal material is occasionally shifted into it after one of the mass movements, and the resulting distention of the rectum initiates the urge to defecate. The act of defecation is a combination of voluntary and involuntary movements. Contraction of the distal end of the colon and relaxation of the internal anal sphincter, which are regions composed of smooth muscle, are involuntary movements. Relaxation of the external anal sphincter, which consists of striated muscle, is a voluntary movement. Other voluntary movements that may assist in the act of defecation are contraction of the abdominal muscles and forcible expiration with closure of the glottis (the Valsalva maneuver, often referred to as *bearing down*).

Regulation of colonic motility. Movements in the proximal portion of the colon are largely initiated by distention in the colonic walls, which stimulates contractile activity by triggering short reflexes through the intrinsic nerve plexuses. Although the proximal part of the colon receives extrinsic innervation via the vagus nerve, it functions in a relatively autonomous manner in the absence of extrinsic motor innervation and is thus a somewhat self-regulating structure. Extrinsic nerves may occasionally modify proximal colonic activity; for example, entry of food into the stomach or duodenum may result in a mass contraction in the proximal end of the colon. Sometimes termed the *gastrocolic* or *duodenocolic reflexes,* these strong mass movements are most evident after the first intake of nutrients in the morning and are often followed by a strong need to defecate.

In contrast, the distal part of the colon is somewhat more dependent on its extrinsic nerve supply, so movements in this region, including the act of defecation, may be entirely abolished after injury to these nerves. However, weak movements return eventually, and defecation can still occur without voluntary control after the initial response to injury has passed.

KEY POINTS

- Movements of the GI tract are due to contraction of two layers of smooth muscle (i.e., the longitudinal and circular layers). Smooth muscle exhibits two types of electrical potentials: basic oscillations (slow waves), which do not result in contraction, and action potentials (spikes), which trigger calcium entry and result in contraction. Contraction of smooth muscle results in two types of intestinal motility: propulsive (peristalsis) and mixing (segmental).
- GI motility is regulated by the enteric nervous system, the autonomic nervous system, and hormonal mediators. The enteric nervous system has two branches—myenteric and submucosal—that coordinate reflexive contraction and relaxation along the entire GI tract.
- Luminal distention is an important stimulus for reflexive motility. Sympathetic nervous system activity is generally inhibitory to GI motility (and secretion). Parasympathetic nervous system activity generally enhances motility. Regulatory hormones include gastrin (increases gastric motility), GIP (decreases gastric motility), CCK (stimulates gallbladder contraction), and secretin (decreases GI motility).
- Swallowing is a complex function coordinated by a swallowing center in the medulla.
- Swallowing is partially voluntary and partially involuntary. Cranial nerves IX, X, and XI mediate the various stages of swallowing.
- Regulation of gastric emptying involves gastric and duodenal factors. Gastric distention and the release of gastrin from gastric mucosa promote gastric emptying. Duodenal distention, acidity, hypertonicity, and high protein and fat concentrations inhibit gastric emptying.
- Chyme remains in the small intestine for 3 to 5 hours, where it is continually mixed by segmental contractions and slowly propelled toward the ileocecal valve by peristalsis. Distention of the terminal ileum results in relaxation of the ileocecal sphincter, which allows contents to enter the large intestine.
- Segmental contractions (haustra) in the large intestine promote water absorption. About 18 hours is required for the contents to traverse the large intestine and reach the distal end of the colon. Three to four times a day a peristaltic mass movement sweeps fecal material along the colon. Mass movements may be initiated by entry of food into the stomach and duodenum (gastrocolic reflex).
- Contraction of the distal end of the colon and relaxation of the internal anal sphincter occur involuntarily as feces enter the rectum, and an urge to defecate occurs. The external anal sphincter is under voluntary control and inhibits defecation until voluntarily relaxed.

SECRETORY FUNCTION

Secretion of Gastrointestinal Juices

The many glands associated with the GI tract produce enzymes that break down the major nutrient components of carbohydrates, fats, and proteins as part of the digestive process. The somewhat archaic term *juices* is still used to describe the fluids secreted in the GI tract, which contain a complex mixture of salts and protein enzymes. Secretion of these digestive juices is stimulated by various factors, including mechanical and chemical stimulation by chyme, parasympathetic stimulation (in certain regions of the GI tract), and various hormones.

Gastrointestinal Hormones

Table 35.1 lists the major hormones of the GI tract and their sources, target organs, major actions, and factors that stimulate release. These hormones are released from the GI mucosa in response to distention or the presence of certain nutrient substances. They are then absorbed into the blood and carried to glands in target tissues (i.e., tissues on which they exert their effects), where they stimulate secretion. Chemically, GI hormones are polypeptides or polypeptide derivatives. Receptors for the peptide hormones are widely distributed throughout the body, including in the brain and central nervous system, and much current research is focused on the varied effects of these complex chemicals.

In addition to their effects on motility, as mentioned previously, gastrin, secretin, CCK, and GIP have critical roles in mediating secretion of GI juices. Gastrin is secreted by the stomach mucosa and stimulates the exocrine (secretory) cells of the gastric glands to produce their specific products, including hydrochloric acid (HCl). Research over the last two decades has shown that gastrin exerts its primary effect on enterochromaffin-like cells (ECL cells), provoking them to release histamine. Histamine, an amine with multiple roles in human physiologic processes, including an ability to constrict bronchial smooth muscle, diffuses readily into nearby parietal cells to induce acid release (Fig. 35.16). The development of specific medications that block the action of histamine (H_2 antagonists) led to the first effective treatment for peptic ulcers by reducing gastric acid secretion.

Secretin was one of the first of the body's many hormones to be discovered. The most potent stimulus for secretin release is HCl, and the presence of acidic chyme in the duodenum promotes its release into the blood from the duodenal mucosa. It is carried to the pancreas, where it stimulates the secretion of a large volume of alkaline juice rich in sodium bicarbonate. In the duodenum, sodium bicarbonate then neutralizes the HCl of the chyme, thus protecting the duodenal mucosa from potential damage and creating a slightly alkaline medium that is optimal for chemical digestion by pancreatic intestinal enzymes. Although the liver produces bile continuously, secretin is effective in increasing the rate of bile secretion. Hormonal regulation is the most important mechanism governing the activity of the pancreas, and CCK has a key role in stimulating the release of large amounts of digestive enzymes from the pancreas. CCK also stimulates the gallbladder to release the bile it stores. GIP acts to slow stomach emptying by decreasing gastric motor activity.

Stimulation of the parasympathetic nerves to certain regions of the GI tract will also increase the rates of glandular secretion. Those glands in the upper portion of the GI tract that are innervated by the vagus and other cranial parasympathetic nerves (particularly the salivary, esophageal, and gastric glands; the pancreas; and some duodenal glands) are especially subject to parasympathetic stimulation. Glands in the distal portion of the large intestine are also affected by parasympathetic stimulation because this region is innervated by the pelvic parasympathetic nerves. In the small intestine, the major stimulus for intestinal secretion is local and mechanical stimulation of the intestinal wall, which initiates the excitation of local myenteric reflexes and subsequent release of secretions.

KEY POINTS

- Major secreting glands and secretions in the GI tract can be summarized as follows:
- Salivary glands: Secrete salivary amylase.
- Gastric glands: Chief cells secrete pepsinogen; parietal cells secrete HCl and IF. HCl activates the conversion of pepsinogen to pepsin, and IF enhances vitamin B12 absorption. Parietal cell secretion is stimulated by acetylcholine, histamine, and gastrin. G cells secrete gastrin into the bloodstream. Gastrin increases gastric motility and stimulates chief and parietal cell secretion.
- Intestinal epithelium: Secretes brush border enzymes (peptidases, lipases, sucrase, lactase), secretin (stimulates pancreatic secretion), and CCK (stimulates gallbladder contraction).
- Pancreas: Secretes bicarbonate-rich fluid containing amylase, trypsin, chymotrypsin, and lipase into the duodenum when stimulated by secretin.
- Gallbladder: Secretes concentrated bile salts into the duodenum when stimulated by CCK.

DIGESTION AND ABSORPTION

Substances contained in foods that are important to maintenance of the body include carbohydrates, fats (also called *lipids*), proteins, vitamins, inorganic salts, and water. Many of the nutrient constituents that compose

TABLE 35.1 Major Hormones of the Gastrointestinal Tract

Hormone	Source	Target Organ	Major Actions	Stimulated By
Gastrin	Stomach (mucosa)	Stomach (gastric glands)	Stimulates gastric glands to secrete specific substances and HCl	Distention of stomach by food; other specific substances (e.g., partially digested proteins, caffeine)
Secretin	Duodenum (mucosa)	Pancreas	Stimulates release of alkaline component of pancreatic juice	Acidic chyme acting on duodenal mucosa
		Liver	Increases bile secretion rate	
Cholecystokinin	Duodenum (mucosa)	Pancreas	Stimulates release of digestive enzymes	Presence of fatty acids and partially digested proteins in duodenum
		Gallbladder	Stimulates gallbladder contraction and emptying	
Glucose-dependent insulinotropic polypeptide (gastric inhibitory peptide)	Duodenum (mucosa)	Stomach	Reduces motor activity of stomach; slows rate of gastric emptying	Presence of fat or carbohydrate in duodenum

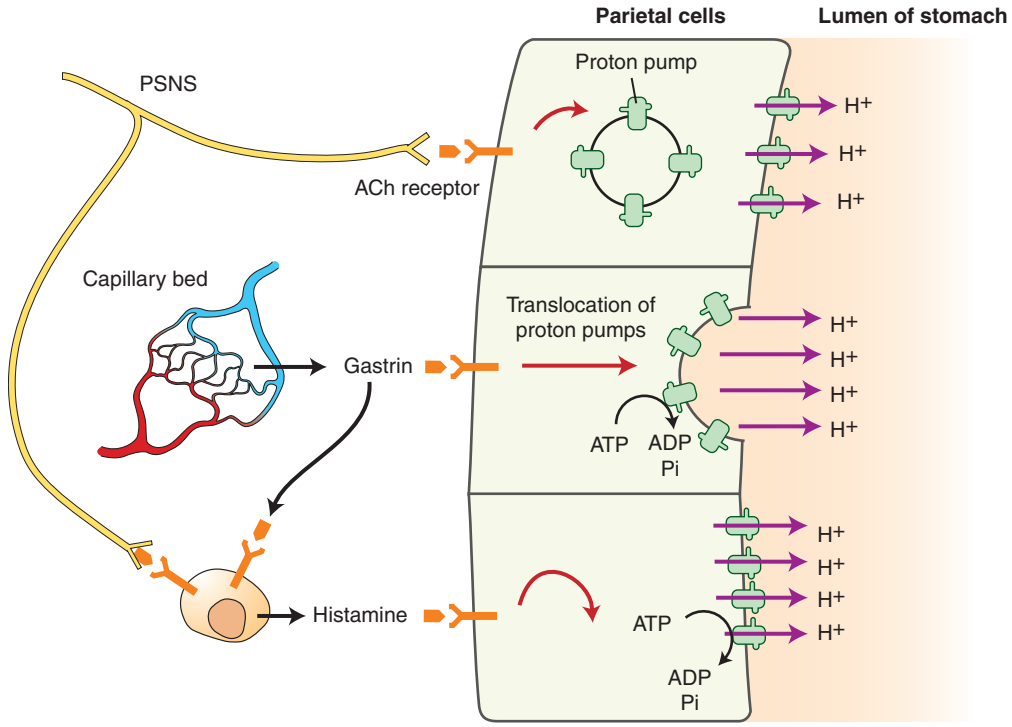


FIG 35.16 Schematic diagram of the complex regulation of acid (H^+) secretion from parietal cells in the gastric pits. Gastrin stimulates enterochromaffin-like cells (ECL cells), which in turn release histamine onto H_2 receptors on the parietal cell. Gastrin also stimulates acid secretion directly, as does acetylcholine. *ACh,* Acetylcholine; *ADP,* adenosine diphosphate; *ATP,* adenosine triphosphate; *Pi,* inorganic phosphate; *PSNS,* parasympathetic nervous system.

intact food substances are structurally complex and cannot be easily absorbed from the GI tract in their original forms. During the process of digestion, digestive juices and the enzymes contained in these secretions convert these complex organic molecules to smaller molecules (Fig. 35.17). These simpler compounds are then capable of absorption across the wall of the small intestine into the blood and lymph and subsequent transfer to the cells, the primary task of the GI tract. This section describes the mechanisms of digestion and absorption of the three major groups of nutrients: carbohydrates, lipids, and proteins.

Digestion of Carbohydrates

In terms of calories, carbohydrates account for approximately half of the American diet. The major digestible carbohydrate in food is plant starch, a large polysaccharide composed of straight and branched chains of glucose. A summary of carbohydrate digestion is presented in Table 35.2.

Digestion of starch begins in the mouth, as salivary amylase breaks down polysaccharides to the much smaller disaccharide molecules maltose and dextrin. In the stomach, this action of salivary amylase continues until the enzyme is eventually inactivated by acidic gastric juice. In the duodenum, the pancreatic enzyme amylase completes the task of splitting any remaining undigested polysaccharides and dextrins to small maltose units. Then maltase, an enzyme located in the brush border of the epithelial cells lining the duodenum, hydrolyzes each maltose molecule to two molecules of glucose. Other carbohydrates that are present in the diet in smaller quantities are the disaccharides sucrose, which is table sugar (glucose-fructose), and lactose, which is milk sugar (glucose-galactose). These two carbohydrates remain chemically unaltered until they reach the duodenum, where the enzyme sucrase in the brush border converts sucrose to the monosaccharides glucose and fructose. The enzyme lactase hydrolyzes lactose into the monosaccharides glucose and galactose.

Glucose, the major product of carbohydrate digestion, accounts for about 80% of the monosaccharides obtained from food, whereas fructose and galactose account for the other 20%. Humans do not secrete an enzyme capable of digesting cellulose—a plant polysaccharide present in large amounts in the cell walls of fibrous vegetables. Although cellulose consists of glucose molecules, it contains molecular linkages different from those of starch. Consequently, much of this complex carbohydrate passes through the digestive tract without being digested and is excreted in the feces.

Digestion of Lipids

Lipids in the diet are mostly in the form of triglycerides but also include phospholipids, cholesterol, and the fat-soluble vitamins A, D, E, and K. Digestion of lipids occurs in the small intestine, where fats are emulsified by the action of bile; neither salivary nor gastric enzymes appear to have any effect on triglycerides. As the lipid particles enter the duodenum from the stomach, bile exerts a detergent action on them in which the surface tension of the particles is decreased. This decrease in surface tension promotes fragmentation of the particles into smaller particles as they are blended by the mixing movements of the small intestine. The emulsification process is an entirely mechanical action because bile contains no enzymes and thus performs no chemical digestion.

Eventually, the detergent action of bile salts reduces the particles of fat to tiny droplets so that their surface area is greatly increased. This enhancement of surface area allows for maximal exposure to pancreatic lipase, an enzyme that (along with intestinal lipase, to a lesser extent) hydrolyzes the triglycerides to free fatty acids and glycerol. Some monoglycerides (glycerol with one fatty acid still attached) may remain; in fact, some fat may escape digestion entirely or be reduced only to diglycerides (glycerol with two fatty acids attached). A summary of triglyceride digestion is presented in Table 35.2.

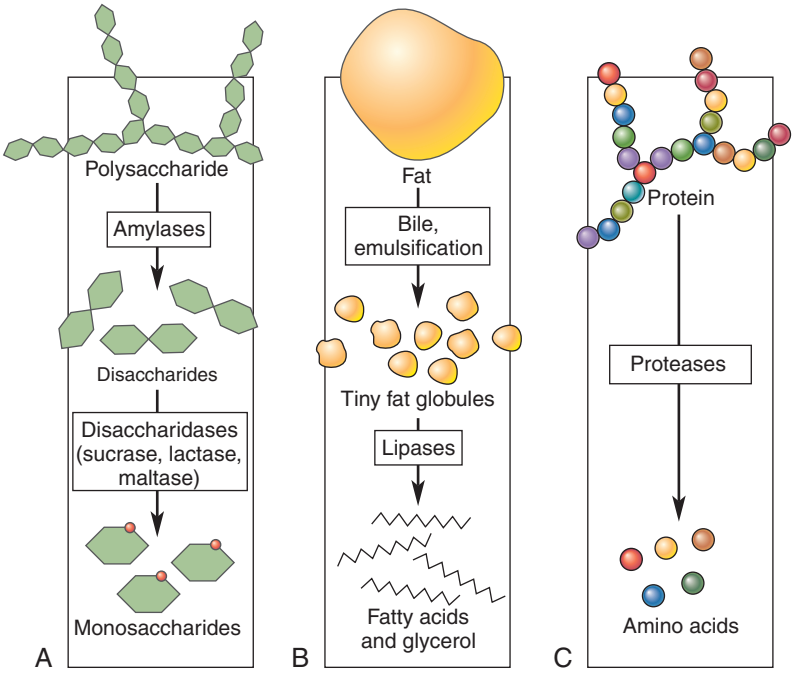


FIG 35.17 Chemical digestion. **A,** Amylases and disaccharidases break carbohydrates down into monosaccharides. **B,** Lipases break fats down to fatty acids and glycerol. The large fat globule must first be emulsified by bile. **C,** Proteases and peptidases break proteins down into amino acids. (From Herlihy B: *The human body in health and illness,* ed 4, Philadelphia, 2011, Saunders.)

TABLE 35.2 Summary of Carbohydrate, Protein, and Lipid Digestion

Location of Digestive Process	Source of Digestive Enzyme or Substance	Basic Digestive Process
Carbohydrates		
Mouth, stomach	Salivary glands (salivary amylase)	Polysaccharides → salivary amylase → maltose + dextrin
Small intestine lumen	Pancreas (pancreatic amylase)	Undigested polysaccharides/dextrins → pancreatic amylase → maltose
Brush borders	Intestine (maltase, sucrase, lactase)	Maltose → maltase → glucose + glucose Sucrose → sucrase → glucose + fructose Lactose → lactase → glucose + galactose
Lipids		
Small intestine	Liver Pancreas	Lipid particle → bile salts → emulsified fat (triglycerides) Triglyceride → lipase → fatty acids + glycerol
Proteins		
Stomach	Stomach (gastric glands)	Protein → pepsin → polypeptides
Small intestine lumen	Pancreas	Polypeptides → trypsin, chymotrypsin → tripeptides + dipeptides → carboxypeptidase → free amino acids
Brush borders (and within cytoplasm of epithelial cells)	Small intestine	Tripeptides and dipeptides → peptidase → free amino acids

Cholesterol, a steroid type of lipid, is ingested in the form of cholesterol esters, which cannot be directly absorbed. An esterase in pancreatic juice degrades cholesterol esters to cholesterol and fatty acid, which then undergo absorption.

Digestion of Proteins

Proteins are composed of molecular subunits called *amino acids* that are linked together by peptide bonds. Proteins that undergo digestion in the small intestine include both protein from food and protein from desquamated cells and the many enzymes of the GI tract. This protein of endogenous origin constitutes a sizable portion of the total protein subjected to digestion and absorption.

Protein digestion involves breakage of the peptide bonds by hydrolysis and release of free amino acids. It begins in the stomach with the action of the enzyme pepsin, which is secreted by the gastric glands. By its action on peptide bonds, pepsin reduces most proteins to intermediate-sized polypeptides. Pepsin is also capable of breaking down collagen, a protein component of intercellular connective tissue, thus rendering cellular proteins more accessible to enzymatic action in the GI tract. In the duodenum, the trypsin and chymotrypsin contained in pancreatic juice reduce the polypeptides to small peptides (tripeptides and dipeptides). Carboxypeptidase, which has its source in the pancreas, and peptidases in the brush borders of the intestinal epithelial cells split some of these peptides into free amino acids. Free amino acids, in addition to dipeptides

and tripeptides, are absorbed into the intestinal epithelial cells. Within the cytoplasm of epithelial cells the small peptides are then hydrolyzed by various peptidases into free amino acids before their passage into the circulation. Numerous proteolytic enzymes are involved in protein digestion, and each enzyme acts on a slightly different type of peptide linkage. Protein digestion is summarized in Table 35.2.

Absorption

Intestinal absorption is the movement of water and dissolved materials, such as the products of nutrient digestion, vitamins, and inorganic salts, from the inside of the small intestine through the semipermeable intestinal membrane and into the blood and lymph. A major feature of the intestinal absorptive surface is the villus, the small fingerlike projection lined with epithelial cells that was described earlier in this chapter. Within each villus is a network of capillaries that branch from a miniscule artery and empty into a miniscule vein. A central lymph vessel called a *lacteal* is also located in the villus. In the process of absorption, nutrient molecules pass through the single layer of epithelial cells lining the villus and through the single layer of cells forming the wall of the capillary or lacteal. A number of transport systems specific to certain nutrient components function in the intestinal epithelium to promote this process of absorption.

Operation of the intestinal transport systems is dependent on ATP production by the epithelial cells. These systems are capable of moving the products of nutrient digestion and inorganic salts from the intestinal lumen into the blood against electrochemical gradients (active transport). In addition to active transport, some molecules may move across the intestinal epithelium by passive means when a difference in concentration on the two sides of the epithelium exists. The rate of molecular transfer based on diffusion gradients is dependent not only on the magnitude of the difference in concentration, but also on the size of the molecules and the lipid solubility of the substances involved.

Almost all substances capable of intestinal absorption disappear from the lumen of the small intestine by the time that the intestinal contents reach the midjejunum. The ileum is not involved in absorption to any significant degree because the proximal regions of the small intestine have usually completed the process of absorption before the intestinal contents reach the ileal region. Nevertheless, the distal end of the small intestine has the capability of absorption and may do so in situations in which absorption has not taken place in the proximal part of the small intestine. Thus about 50% of the small intestine can be surgically removed without compromising absorptive ability. However, it is important to note that vitamin B12 and bile salts are absorbed specifically in the terminal ileum, and surgical removal of this portion of the small intestine will result in impaired absorption of these substances. IF produced by the parietal cells of the gastric antrum is also required for B12 absorption; IF deficiency leads to a condition known as *pernicious anemia*.

The intestinal contents arriving at the terminal ileum contain no digestible carbohydrate, very little fat, and only 15% to 17% nitrogen-containing substances. Most of the contents of the terminal ileum consist of bacteria, desquamated epithelial cells, digestive secretions, and the residue of foods that are undigested and therefore unabsorbed, such as the cellulose walls of fibrous plants and connective tissue from animal sources.

Carbohydrates

Carbohydrates are absorbed in the form of monosaccharides. The intestinal epithelium is impermeable to carbohydrates of high molecular weight such as disaccharides and polysaccharides, and no transport systems exist for these types of carbohydrate molecules. The monosaccharides glucose and galactose are absorbed by an active, energy-requiring process in which a carrier molecule located on the luminal border of epithelial cells transports them across the border. The same carrier molecule that ferries glucose and galactose also carries sodium, which also enhances the carrier affinity for monosaccharides. In contrast to the other monosaccharides, the monosaccharide fructose is absorbed passively by means of a diffusion gradient.

Lipids

Absorption of lipids occurs by a highly complex, unique process. As fatty acids and monoglycerides are freed during digestion, they become dissolved in bile salt micelles, which are colloidal particles composed of many molecules. Within the micelles, the products of lipid digestion are now soluble and can be absorbed far more efficiently. The bile salt micelles transport the lipid products to the epithelial brush borders, where the monoglycerides or fatty acids, which are highly soluble in the lipid cell membrane, diffuse into the epithelial cells and leave the micelle behind. The micelle is now emptied of its cargo and can pick up more fatty acids and monoglycerides and transport them to the cell membrane.

Bile salts, which are required for micelle formation, are absorbed mostly in the terminal ileum and then recycled in the liver. In the absence of bile, the amount of lipid absorbed in this manner is reduced by more than 25%. In this situation, the absorption of fat-soluble vitamins (vitamins A, D, E, and K) is compromised. Several cholestatic conditions, such as primary biliary cirrhosis and primary sclerosing cholangitis, may be associated with deficiencies of fat-soluble vitamins.

Monoglycerides may be further degraded into glycerol and fatty acids by the enzyme lipase within the epithelial cell. Short-chain fatty acids (those with fewer than 12 carbon atoms) can be absorbed directly into the blood at this point. Long-chain fatty acids and glycerol, however, are reassembled into triglycerides by the endoplasmic reticulum. These newly synthesized triglycerides are aggregated into droplets that become progressively larger during passage through the cell. These lipid droplets are stabilized by enclosure with absorbed cholesterol and phospholipids and encased by a protein coat. The final product, called a *chylomicron,* passes out of the cell and into the lacteal of the villus. From the lacteal, chylomicrons pass through a series of lymph vessels that eventually drain into the general circulation.

Proteins

Amino acids are transported across the epithelial membrane by means of an active transport carrier system in much the same way as monosaccharides. It is currently thought that different carrier systems exist to carry the different chemical classes of amino acids (i.e., neutral, basic, dicarboxylic, and imino acids). As is the case for the transport of monosaccharides, brush border membrane carriers are involved in the transfer of amino acids across the intestinal epithelial cell; these carriers require energy and are coupled to the transport of sodium. After being transported to the epithelial cells of the villi, amino acids diffuse through the base of the cell and into the blood. Both amino acids and monosaccharides are transported directly to the liver by the hepatic portal vein.

Water and Electrolytes

Water and inorganic ions, which are in the GI tract as a result of ingestion and secretion, are absorbed mainly from the small intestine and, to a lesser extent, from the colon. The process of absorption of water and ions is the same in both the small and large intestines: sodium is actively transported to the blood, and water follows passively in response to the osmotic gradient created by the removal of sodium from the intraluminal fluid. About 8000 mL of water is absorbed every day by the small intestine and about 300 to 400 mL by the colon. Frequently, diarrhea is the result of failure of the small intestine to absorb water appropriately. If large quantities of water are allowed to enter the colon from the small intestine because of some malfunction of the small intestine's absorptive ability, the colonic absorptive mechanism may be overwhelmed, and diarrhea is the result (Fig. 35.18).

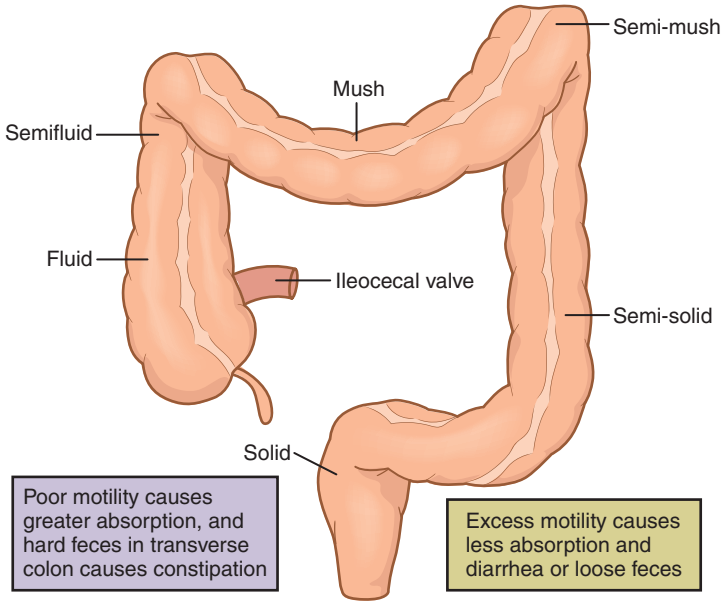


FIG 35.18 Absorptive function of the large intestine. (From Hall JE: *Guyton and Hall textbook of medical physiology,* ed 13, Philadelphia, 2016, Saunders.)

KEY POINTS

- Digestion, the process of converting large molecules to simpler forms, is accomplished by mechanical and enzymatic processes. Digestion is a necessary prelude to absorption because only simple molecules can cross the intestinal epithelia.
- Digestion of complex carbohydrates is initiated in the mouth, where salivary amylase begins to cleave polysaccharides into disaccharides. Pancreatic amylase continues this process in the small intestine. Disaccharides (e.g., maltose, sucrose, lactose) are cleaved into monosaccharides (e.g., glucose, fructose, galactose) by brush border enzymes (e.g., maltase, sucrase, lactase) on the intestinal epithelia. Glucose and galactose are absorbed across the intestinal epithelia by a sodium-dependent cotransporter. Fructose is absorbed passively by facilitated diffusion. Monosaccharides then travel via the bloodstream to the liver.
- Lipid digestion begins in the small intestine, where bile salts from the gallbladder mix and emulsify the fatty substances. Emulsification mechanically separates the lipids into small drops that are more accessible to enzymatic digestion. Pancreatic lipase and brush border lipases digest the lipids into free fatty acids and glycerol, which remain associated with the bile salts and form micelles. Cholesterol is digested by pancreatic esterase. Fatty acids are transported to the intestinal epithelia by micelles. Free fatty acids diffuse out of the micelle and into the epithelial cell passively. Epithelial cells synthesize large protein-lipid complexes (chylomicrons) that enter the lymphatic system.

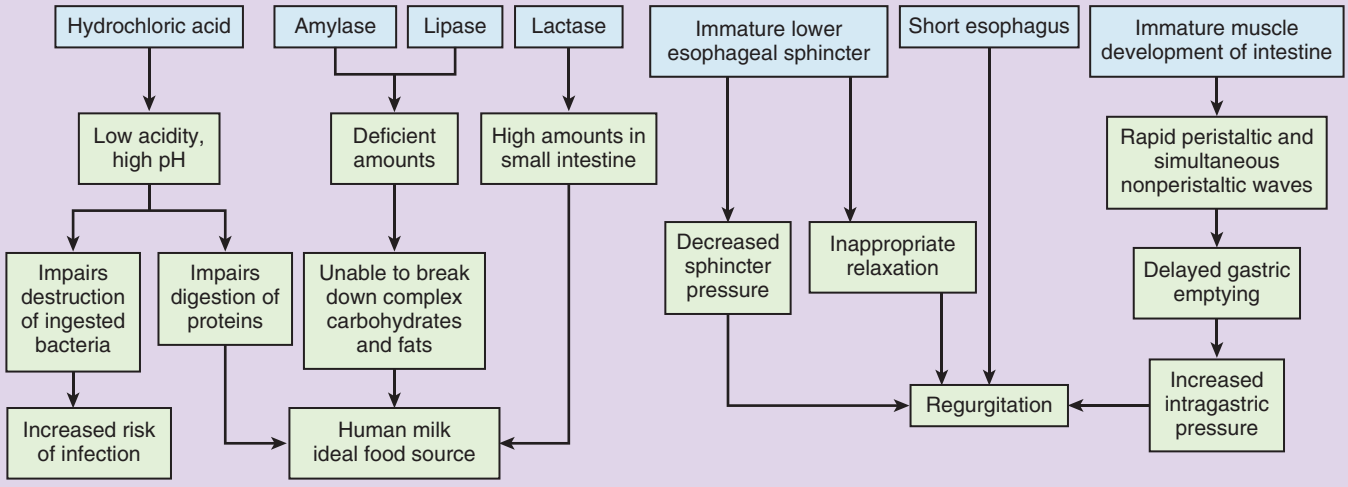
PEDIATRIC CONSIDERATIONS

Changes in the Gastrointestinal System in the Infant

The acid and enzymes present in the infant's GI system are different from those in the adult's GI tract and affect the infant's ability to digest nutrients. The stomach depends on hydrochloric acid to begin digestion of human milk. Although hydrochloric acid is present, the stomach acidity is low. This low acidity limits the GI system's ability to destroy ingested bacteria, which causes the infant to be more susceptible to GI infections. The acidity is also too low to digest protein in the stomach via the enzyme pepsin. As the child matures, the acidity of the gastric contents continues to increase until it reaches adult levels in adolescence.

Lipase and amylase in the intestine are limited for the first 4 months of the infant's life. The infant is unable to digest fats and complex carbohydrates until the levels of these enzymes are sufficient. Levels of the enzyme lactase, found in the small intestine, are extremely high in the newborn. This is essential for digestion of the human milk diet of the infant. The lactase levels decline after infancy, and for most individuals, lactase levels are absent by adulthood. Because of the enzyme composition in the infant GI system, human milk is the ideal food source for an infant.

Several factors of the GI system predispose the infant to regurgitation. The lower esophageal sphincter is immature in the infant. This immaturity leads to inappropriate relaxation of the sphincter, and the pressure of the sphincter is decreased. The infant also has a shorter esophagus, which results in less distance for gastric contents to travel and increases the likelihood of regurgitation. Muscle development is immature in the intestine. This immaturity leads to rapid peristaltic waves and simultaneous nonperistaltic waves. This inconsistency of the intestine to move food efficiently leads to delayed gastric emptying. As the gastric contents increase, the pressure of the stomach increases and exceeds the pressure of the lower esophageal sphincter. Once the pressure of the sphincter is exceeded, regurgitation occurs. As the infant grows, the GI system develops and is fairly complete by the beginning of toddlerhood.



- Protein digestion begins in the stomach, where HCl from parietal cells activates the conversion of pepsinogen to pepsin. Pepsin cleaves proteins into smaller polypeptides. Pepsin is neutralized in the duodenum, and pancreatic trypsin, chymotrypsin, and carboxypeptidase take over protein digestion. Brush border peptidases split tripeptides and dipeptides into single amino acids. Amino acid transport into intestinal epithelial cells is mediated by a sodium-dependent cotransport system similar to monosaccharide transport. Small peptides may also undergo endocytosis and be cleaved into amino acids within the epithelial cells.
- Amino acids pass into the bloodstream and travel to the liver.
- Absorption of water occurs passively by osmosis. An osmotic gradient for water absorption is created as electrolytes are absorbed.

GASTROINTESTINAL FUNCTION ACROSS THE LIFE SPAN

Maturation

During the first months of life, the newborn's GI tract undergoes many maturational changes. In the first 3 to 4 months of life, sucking reflexes are present, and extrusion reflexes protect against the ingestion of solids.

The pressure in the LES remains low during this time, and "spitting up" of gastric contents is common because intragastric pressure often exceeds LES pressure. Gastric motility is not well coordinated for the first 3 to 4 months, so antral mixing is inadequate for the digestion of solid foods. At about 12 weeks of age, intestinal peristalsis similar to that in adults begins to develop, but it is one-third slower. This slower transit in infants may serve to improve nutrient digestion and absorption by increasing the exposure of nutrients to the intestinal mucosa. The motor function of the large intestine appears to be fully developed at birth. During the first 2 years of life, the secretory and absorptive functions of the intestine mature and begin a pattern of functioning that continues into senescence. See Pediatric Considerations box.

Age-Related Changes

Changes in GI function in older adults occur simultaneously with other age-related changes such as a decrease in lean body mass and impaired homeostasis of multiple body systems. Within the GI tract, a variety of changes occur that may place an aging individual at risk for health problems related to GI functioning and nutrition. Important elements of this process are summarized in Geriatric Considerations: Changes in the Gastrointestinal System.

Loss of dentition and reduced taste and smell acuity may promote a decreased interest in food intake as chewing becomes difficult and the sensory enjoyment associated with food becomes impaired. A condition called *presbyesophagus*, in which esophageal motility is slowed or disorganized, may develop in older adults. Presbyesophagus may be manifested as difficulty in swallowing and may cause discomfort as food passes through the esophagus. The incidence of hiatal hernia, where part of the stomach moves into the chest cavity through an enlarged diaphragmatic opening, is also increased in the aging population,

GERIATRIC CONSIDERATIONS

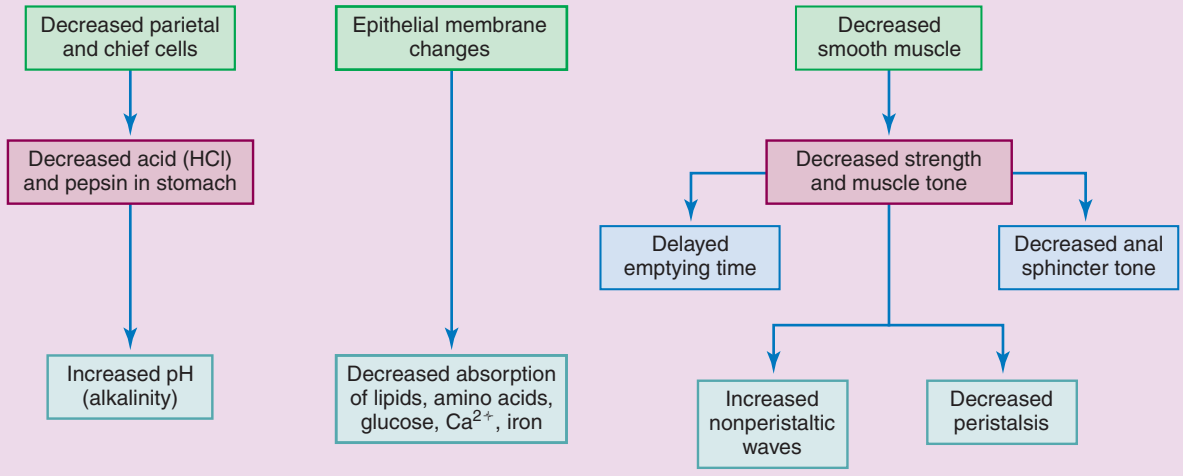
Changes in the Gastrointestinal System

As a person ages, gastrointestinal muscle strength and movement decrease, leading to reduced peristalsis and decreased gastrointestinal motility throughout the system.

In the esophagus, the elderly person experiences greater numbers of muscle movements that do not propel the contents onward. These nonperistaltic waves are common in the lower esophagus. The phenomenon of presbyesophagus—in which the esophageal sphincter fails to relax and the lower esophagus becomes dilated—may not necessarily be normal to the elderly.

In the stomach, decreased numbers of parietal and chief cells result in diminished acid (HCl) and pepsin secretion. This leads to increased pH and a more alkaline secretion. The amount of protective alkaline viscous mucus in the stomach is also decreased. The loss of smooth muscle in the stomach can delay emptying time, which increases and prolongs the exposure of gastric epithelial cells to the gastric contents.

The amount of small intestinal smooth muscle, Peyer patches, and lymphatic follicles is decreased. Normal intestinal absorption in the elderly is not well understood and may be influenced by a number of factors, including bowel motility, epithelial membranes, vascular perfusion, and gastrointestinal membrane transport. However, absorption of lipids, amino acids, glucose, calcium, and iron is known to be decreased. Normal changes in the large intestine have been difficult to determine. As a result of smooth muscle changes, anal sphincter tone decreases.



affecting two-thirds of persons older than 70 years. The transit time for intestinal contents to pass through the GI tract is increased in older persons; this factor, coupled with a decreased perception of the sensory stimuli that produce the urge to defecate, may promote constipation in the aging population. Conversely, a confused or neurologically impaired older individual may experience fecal incontinence because the sensation and tone of the rectum diminish with aging.

KEY POINTS

- Infants may experience GI dysfunction because of immaturity of the GI tract. Motility is not well coordinated until 3 to 4 months of age, making digestion of solids difficult in infancy. Pressure in the LES is low, which leads to "spitting up" and gastric distention. Maturation of the GI tract is complete by about 2 years of age.
- Elderly individuals may experience GI dysfunction for a number of reasons. Poor dentition, loss of taste and smell acuity, and reduced esophageal motility may lead to poor intake of nutrients. Hiatal hernia and constipation are common in the elderly.

SUMMARY

This chapter describes the structure of the human GI system and the process by which it provides nutrients for the body. A thorough understanding of the structure and motility of the GI tract, secretion of digestive juices, and digestion and absorption of nutrients is needed as a basis for understanding other principles of health and disease.

GI motility is a complex process involving a set of carefully timed autoregulatory action responses (Fig. 35.19). You may wish to trace the path and destiny of the apple you ate for lunch as an example of this process. As you track the movement of nutrients through the GI tract, consider the ways in which secretion of digestive juices occurs in response

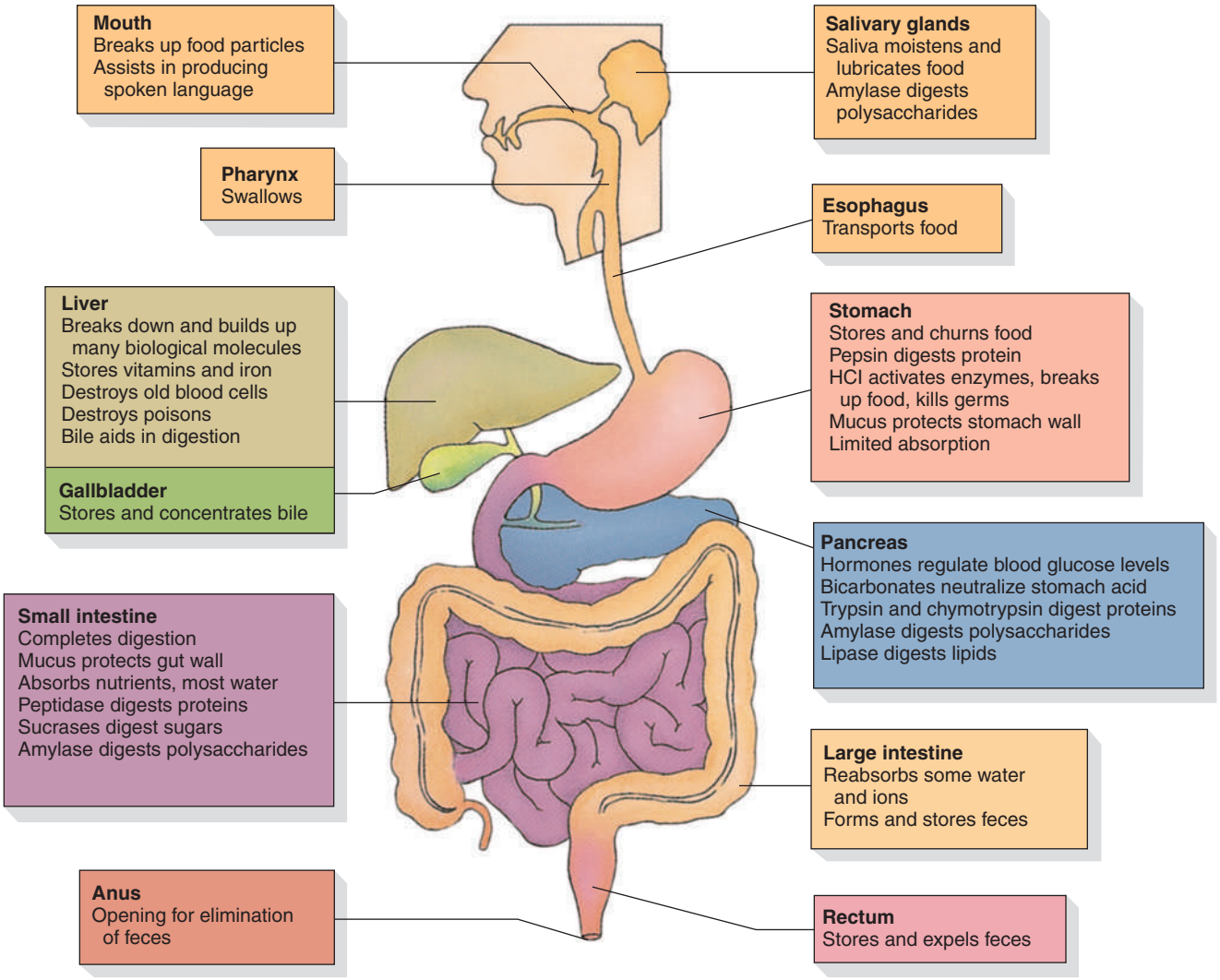


FIG 35.19 Summary of digestive function. (From Patton KT, Thibodeau GA: *Essentials of anatomy and physiology,* St Louis, 2012, Mosby.)

to the ingestion of your apple, which contains a great deal of carbohydrate (fructose), small amounts of protein, and minimal lipid. Consider also how digestion and absorption of these nutrients are occurring. What part of the apple will you use, for example, for energy to study this text? What part of the apple will your body "throw away," and how will this be accomplished? Finally, will your GI tract respond the same way to eating an apple when you are 85 years old? A careful review of the elegant and nearly automatic function of the human GI tract will prepare you to care for individuals experiencing interruptions in proper nutrient digestion and absorption.

RESOURCES

Barocelli E, Ballabeni V: Histamine in the control of gastric acid secretion: a topic review. *Pharmacol Res* 47(4):299–304, 2003.

Berna MJ, Jensen RT: Role of CCK/gastrin receptors in gastrointestinal/metabolic diseases and results of human studies using gastrin/CCK receptor agonists/antagonists in these diseases. *Curr Top Med Chem* 7(12):1211–1231, 2007.

Burns AJ, Thapar N: Advances in ontogeny of the enteric nervous system. *Neurogastroenterol Motil* 18(10):876–887, 2006.

de Herder WW, Lamberts SW: Somatostatin and somatostatin analogues: diagnostic and therapeutic uses. *Curr Opin Oncol* 14(1):53–57, 2002.

Feldman M, Friedman LS, Brandt LJ: *Sleisenger and Fordtran's gastrointestinal and liver disease*, ed 10, Philadelphia, 2015, Saunders Elsevier.

Gebruers EM, Hall WJ: Role of the gastrointestinal tract in the regulation of hydration in man. *Dig Dis* 10(2):112–120, 1992.

Hall JE: *Guyton and Hall textbook of medical physiology*, ed 13, Philadelphia, 2015, Saunders Elsevier.

Herrera JL, Lyons MF, 2nd, Johnson LF: Saliva: its role in health and disease. *J Clin Gastroenterol* 10(5):569–578, 1988.

Hockenberry MJ, Wilson D: *Wong's nursing care of infants and children*, ed 9, St Louis, 2011, Mosby.

Hofmann AF: Cholestatic liver disease: pathophysiology and therapeutic options. *Liver* 22(Suppl 2):14–19, 2002.

Hornby PJ: Central neurocircuitry associated with emesis. *Am J Med* 111(Suppl 8A):106S–112S, 2001.

Jaffe BM, Berger DH: The appendix. In Schwartz SI, Brunicardi CF, editors: *Schwartz principles of surgery*, ed 8, New York, 2005, McGraw-Hill.

Johnson LR: *Gastrointestinal physiology*, ed 8, Philadelphia, 2013, Mosby Elsevier.

MacGregor J: *Introduction to the anatomy and physiology of children: a guide for students of nursing, child care and health*, ed 2, New York, 2008, Routledge.

Nonogaki K: Ghrelin and feedback systems. *Vitam Horm* 77:149–170, 2007.

Orr WC, Chen CL: Aging and neural control of the GI tract, IV: clinical and physiological aspects of gastrointestinal motility and aging. *Am J Physiol Gastrointest Liver Physiol* 283(6):G1226–G1231, 2002.

Pedersen AM, et al: Saliva and gastrointestinal functions of taste, mastication, swallowing and digestion. *Oral Dis* 8(3):117–129, 2002.

Podolski DK, et al, editors: *Yamada's textbook of gastroenterology*, ed 6, Oxford, 2015, Wiley-Blackwell.

Rehfeld JF, Friis-Hansen L, Goetze JP, et al: The biology of cholecystokinin and gastrin peptides. *Curr Top Med Chem* 7(12):1154–1165, 2007.

Ross AC, et al, editors: *Modern nutrition in health and disease*, ed 11, Philadelphia, 2012, Lippincott Williams & Wilkins.

Roy CC, Siverman A, Alagille D, editors: *Pediatric clinical gastroenterology*, ed 4, St Louis, 1995, Mosby.

Thomson AB, et al: Small bowel review: normal physiology part 2. *Dig Dis Sci* 46(12):2588–2607, 2001.

Wiedmer P, Nogueiras R, Broglio F, et al: Ghrelin, obesity and diabetes. *Nat Clin Pract Endocrinol Metab* 3(10):705–712, 2007.

36

Gastrointestinal Disorders

Jeffrey S. Sartin

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- What are the common causes of these general manifestations of gastrointestinal disorders: pain, nausea, vomiting, diarrhea, and constipation?
- What are the predisposing factors and characteristics common to inflammatory disorders of the gastrointestinal tract?
- What are the common causes of and clinical findings in functional and mechanical bowel obstructions?
- What are the common causes of and clinical findings in gastrointestinal malabsorption disorders?
- What are the warning signs that may indicate cancer of the gastrointestinal tract?

CHAPTER OUTLINE

Manifestations of Gastrointestinal Tract Disorders, 721
Dysphagia, 721
Categories, 721
Esophageal Pain, 721
Abdominal Pain, 721
Vomiting, 723
Intestinal Gas, 723
Alterations in Bowel Patterns, 723
Constipation, 723
Diarrhea, 723
DISORDERS OF THE MOUTH AND ESOPHAGUS, 724
Oral Infections, 724
Stomatitis, 724
Esophageal Disorders, 724
Gastroesophageal Reflux Disease, 724
Hiatal Hernia, 725
Mallory–Weiss Syndrome, 725
Esophageal Varices, 725
ALTERATIONS IN THE INTEGRITY OF THE GASTROINTESTINAL TRACT WALL, 726
Inflammation of the Stomach and Intestines, 726
Gastritis, 726
Gastroenteritis, 726
Peptic Ulcer Disease, 726
Inflammatory Bowel Disease, 728
Ulcerative Colitis, 729
Crohn Disease, 730
Enterocolitis, 731
Antibiotic-Associated Colitis (Pseudomembranous Colitis), 731
Necrotizing Enterocolitis, 731
Appendicitis, 731
Diverticular Disease, 731
ALTERATIONS IN MOTILITY OF THE GASTROINTESTINAL TRACT, 733
Motility Disorders, 733
Irritable Bowel Syndrome, 733
Intestinal Obstruction, 733
Volvulus, 734
Intussusception, 734
Megacolon, 734
Hirschsprung Disease, 734
Disorders of Malabsorption, 735
Mucosal Disorders, 735
Celiac Disease, 735
Tropical Sprue, 735
Malabsorption Disorders After Surgical Intervention, 735
Dumping Syndrome, 735
Short-Bowel Syndrome, 736
NEOPLASMS OF THE GASTROINTESTINAL TRACT, 737
Esophageal, Gastric, and Small Intestinal Cancers, 737
Esophageal Cancer, 737
Gastric Carcinoma, 737
Small Intestinal Neoplasms, 737
Colonic Polyps and Colon Cancer, 738
Colon Polyps, 738
Colon Cancer, 738
Psychosocial Aspects of Gastrointestinal Disorders, 739
Stress of Lifestyle Changes, 739

Alterations in function of the gastrointestinal (GI) tract may have far-reaching consequences in an individual's life. The ability to take in nutrients, convert them to usable forms for body functions, and dispose of their waste products goes beyond physiologic function and is intimately associated with social and psychological health. A person with an alteration in GI function may experience great emotional distress and be unable to participate fully in social activities, which in many societies are often centered on food consumption. Certain symptoms that may accompany GI disorders, such as chronic diarrhea and abdominal pain, may severely limit an individual's ability to maintain employment. It has been estimated that 200,000 people miss work daily because of GI-related problems. In addition, GI diseases account for more hospital admissions in the United States than any other category of disease. Because many chronic GI conditions begin in midlife and continue into old age, their prevalence will likely increase as the U.S. population continues to age.

This chapter describes the pathophysiologic processes associated with the most common GI disorders, along with current treatment recommendations for these conditions. Because GI disorders represent an important focus of current research, knowledge about these disorders is expanding rapidly. Some of the most current evidence-based treatment options are described. A discussion of the emotional and psychological aspects of GI disturbances is provided within the context of patient integrity and well-being.

MANIFESTATIONS OF GASTROINTESTINAL TRACT DISORDERS

Common manifestations of GI disorders include dysphagia, esophageal and abdominal pain, vomiting, intestinal gas, and alterations in bowel patterns.

Dysphagia

Dysphagia is a subjective difficulty in swallowing (Fig. 36.1). It may include the inability to initiate swallowing or the sensation that the swallowed solids or liquids "stick" in the esophagus. In certain disorders, odynophagia, or pain with swallowing, may accompany dysphagia. The physiologic mechanism of normal swallowing is described in Chapter 35.

Categories

The pathophysiologic basis for dysphagia usually falls into three major categories: (1) problems in delivery of the bolus of food or fluid into the esophagus as a result of neuromuscular incoordination; (2) problems in transport of the bolus down the body of the esophagus as a result of altered esophageal peristaltic activity; and (3) problems in bolus entry into the stomach as a result of lower esophageal sphincter (LES) dysfunction or obstructing lesions.

In the first category of dysphagia, individuals have a decreased ability to accomplish the initial steps of swallowing in an orderly sequence. The normal sequence of contraction of the pharynx, closure of the epiglottis, relaxation of the upper esophageal sphincter, and initiation of peristalsis by contraction of the striated muscle in the upper portion of the esophagus is altered, or certain steps in the sequence may be absent. Persons experiencing this type of dysphagia may cough and expel the ingested food or fluids through their mouth and nose or aspirate when they attempt to swallow. With this type of swallowing dysfunction, symptoms are usually worse with the consumption of liquids than with solids.

The second type of dysphagia may be the result of any disorder, structural or neuromuscular, in which the peristaltic activity of the body of the esophagus is altered. Some conditions leading to disordered peristalsis include esophageal diverticula, or outpouchings of one or more layers of the esophageal wall; achalasia, a disorder of esophageal smooth muscle function; and structural disorders such as neoplasms or strictures. This alteration in peristalsis may be simply weak peristaltic activity, aperistalsis (the absence of all peristaltic activity), or disorganized and therefore ineffective peristalsis. With this type of dysphagia the individual may have the sensation that food is "stuck" behind the sternum. Initially, dysphagia may be noted with solid foods; if the underlying pathologic process fosters a worsening of peristaltic ability, the passage of liquids may also become impaired.

The third category of dysphagia, which results from problems of bolus entry into the stomach, is secondary to any condition in which the LES functions improperly or is obstructed by a lesion. Tumors of the mediastinum, lower part of the esophagus, or gastroesophageal junction may produce an obstruction at the LES or invade the myenteric plexus, thus interrupting normal LES function. In addition, motor disorders resulting from neuromuscular diseases or chronic lower esophageal inflammation from the reflux of acidic gastric contents may limit the ability of the LES to function properly. This type of dysphagia may be manifested as tightness or pain in the substernal area during the swallowing process.

Esophageal Pain

Two types of pain occur in the esophagus: (1) heartburn (also called *pyrosis*) and (2) pain located in the middle of the chest, which may mimic the pain of angina pectoris. Heartburn is a substernal burning sensation that may radiate to the neck or throat, caused by the reflux of gastric contents into the esophagus. Two common mechanisms contribute to the manifestations of heartburn. First, the highly acidic gastric contents may be a noxious stimulant to sensory afferent nerve endings in the esophageal mucosa. Second, spasm of the esophageal muscle instigated by acid stimulation may produce esophageal pain.

Chest pain other than heartburn may be the result of esophageal distention or powerful esophageal contractions. These stimuli may arise from esophageal obstruction or a condition called *diffuse esophageal spasm,* in which high-amplitude, simultaneous contractions in the smooth muscle portion of the esophagus alternate with normal peristalsis. This type of esophageal pain is similar to that of angina pectoris, particularly in its pattern of radiation into the neck, shoulder, arm, and jaw. Odynophagia may accompany diffuse esophageal spasm and can be indistinguishable from esophageal chest pain, except that it is triggered specifically by swallowing.

Infections of the esophagus attributable to herpes simplex virus (HSV), cytomegalovirus, or *Candida* species occur in immunocompromised patients. Patients with infectious esophagitis may experience a dull, aching chest pain. Swallowing generally worsens the sensation of heartburn or chest pain.

Abdominal Pain

Pain in the abdominal region may be the first sign of a disorder of the GI tract and is often an important impetus for seeking medical care. Although abdominal pain may result from GI tract disorders, it may also be the result of reproductive, genitourinary, musculoskeletal, or vascular disorders, as well as toxins or drug use. Abdominal pain is usually categorized into three types, which may manifest separately or in combination: (1) Visceral pain develops from stretching or distending an abdominal organ or from inflammation. The pain is diffuse and poorly localized and has a gnawing, burning, or cramping quality. (2) Somatic pain arises from injury to the abdominal wall, the parietal peritoneum, the root of the mesentery, or the diaphragm. In contrast to visceral pain, it is sharper, more intense, and generally well localized to the area of irritation. (3) Referred pain is felt at a location distant from the source of the pain but in the same dermatome or neurosegment.

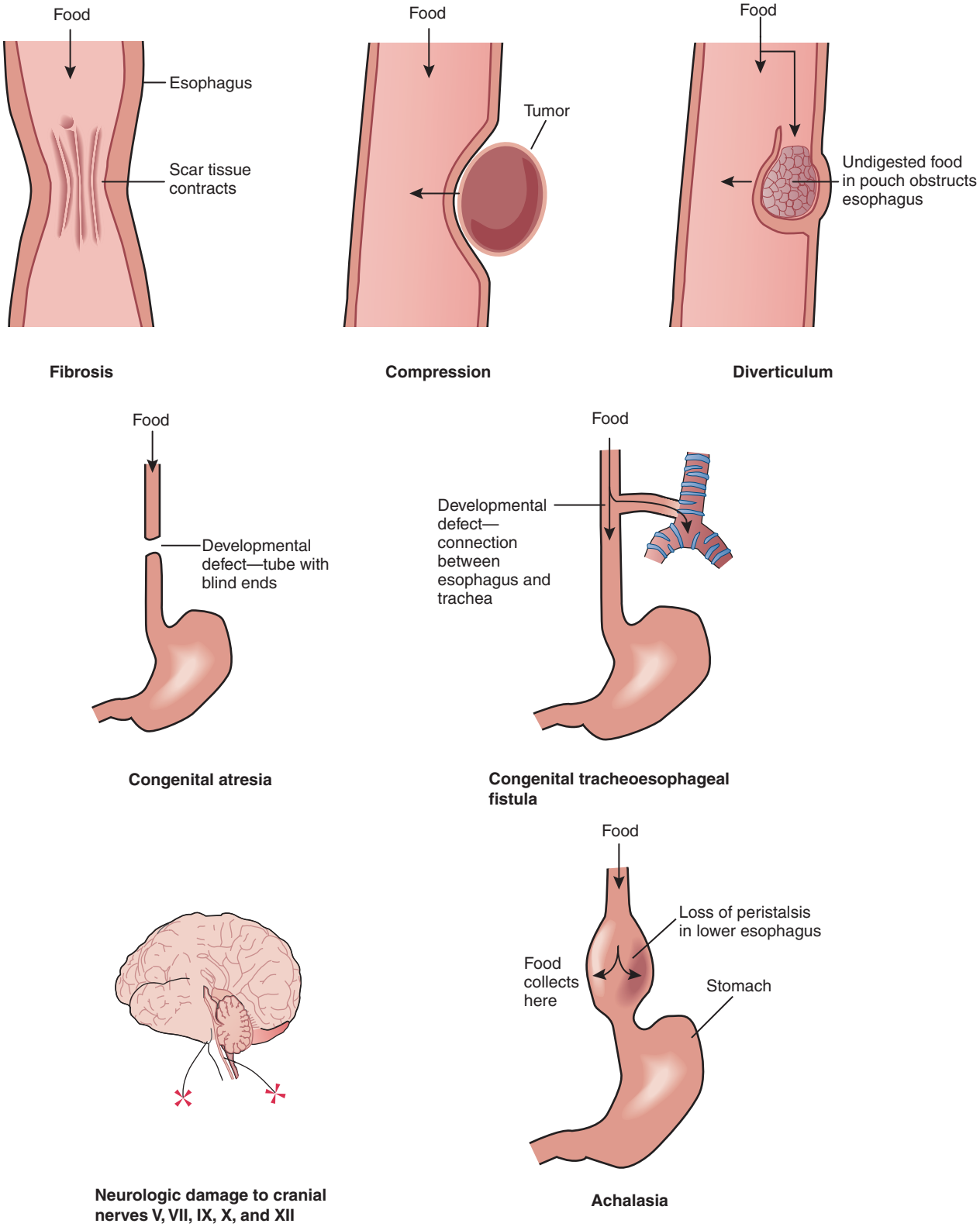


FIG 36.1 Causes of dysphagia. (From Gould BE: *Pathophysiology for the health professions,* ed 6, Philadelphia, 2018, Saunders.)

Referred pain is usually sharp and well localized and may be felt in the skin or deeper tissues.

Abdominal pain may be acute with instantaneous onset, such as pain caused by a perforated ulcer or a ruptured internal organ. A more gradual development of abdominal pain may accompany such chronic states as diverticulitis or ulcerative colitis (UC). Abdominal pain seldom occurs as a solitary manifestation of GI disorders; it is usually accompanied by other manifestations such as vomiting or alteration in bowel patterns to a variable degree.

Vomiting

Vomiting is the forceful expulsion of gastric contents through the mouth. Vomiting is usually accompanied by a feeling of nausea and is the result of a coordinated sequence of abdominal muscle contractions and reverse esophageal peristalsis. Although vomiting is a common sign of GI disorders, it may also occur with metabolic, endocrine, vestibular (inner ear), and cardiac disorders, as well as infection and fluid and electrolyte imbalances. It is also associated with such nonpathologic causes such as pharmacologic agents, surgery, and the first trimester of pregnancy.

Vomiting associated with GI disorders may be the result of alterations in the integrity of the GI tract wall, such as gastroenteritis, or alterations in the motility of the GI tract, such as intestinal obstruction. The characteristics of the vomitus and the presence of blood or fecal matter may suggest the nature of the GI disorder and the level of the GI tract at which the disorder is located.

Intestinal Gas

Gas is a normal occurrence in the GI tract as the result of the swallowing of air or bacterial and digestive action on intestinal contents, diffusion from the blood, or the neutralization of acids by bicarbonate within the upper GI tract. The manifestations of excess intestinal gas include distention of the abdomen and prominent belching and flatus. These manifestations may occur singly or in combination and may stem from a variety of causes. One prominent cause of abdominal distention is failure to adequately digest a particular nutrient, such as the carbohydrate lactose. In the absence of adequate lactase (the digestive enzyme that breaks down lactose into glucose and galactose in the intestine), lactose undergoes bacterial fermentation, which results in gas production in the intestinal lumen. In some individuals, abdominal distention from excess gas may result from a defect in intestinal motility in which the intestinal contents are not propelled in a regular fashion, rather than from the production of too much gas. Belching is a normal phenomenon caused by the eructation of swallowed air, but may also be the result of a motility disorder or gastric outlet obstruction. Excessive flatus has causes similar to those of abdominal distention. Most commonly, it is the result of the action of bacteria on nutritional substrates that are particularly gas producing, such as certain vegetables and legumes. Some individuals are particularly sensitive to the flatulent effects of beans.

Alterations in Bowel Patterns

Because there is a wide variation in bowel patterns in individuals, both constipation and diarrhea are difficult to define with precision. In addition, cultural and family socialization may play a role in the way in which an individual perceives bowel patterns. Alterations in bowel patterns may be the result of a change in GI tract motility or may be a component of a functional GI disorder such as irritable bowel syndrome (IBS).

Constipation

Constipation may be defined as small, infrequent, or difficult bowel movements. Authorities have agreed on a norm of fewer than three stools per week as a guideline for defining constipation. (Many textbooks suggest that everyone should have at least one bowel movement no longer than 72 hours apart.) Dietary factors, particularly a diet low in fiber, have been shown to contribute to constipation. The presence of cellulose, the carbohydrate component of dietary fiber that is indigestible in the human intestine, may be effective in promoting regular peristaltic movement in the GI tract by forming bulk within the intestinal lumen to stimulate propulsion. Because exercise stimulates intestinal peristalsis, a lack of exercise has been implicated in the development of constipation. In elderly persons the slowed rate of peristalsis that occurs with the aging process coupled with a decreased level of physical activity may promote chronic constipation. These factors may eventually lead to fecal impaction, a condition in which a firm, immovable mass of stool obstructs the lower GI tract. Constipation may also be the result of pathologic conditions, including processes that alter the motility of the GI tract (such as intestinal obstruction) or processes that alter the integrity of the GI tract wall (such as diverticulitis).

Diarrhea

Diarrhea is defined as an increase in the frequency and fluidity of bowel movements and is often a primary sign of GI tract disorders. Although stool weight in excess of 200 g in 24 hours is an easily obtainable, objective definition of diarrhea, most persons consider increased liquidity as the primary determinant. Diarrhea may occur as an acute or chronic condition. Acute diarrhea may be the result of an acute infection, emotional stress, or leakage of liquid stool around impacted feces. Chronic diarrhea is usually defined as symptoms lasting longer than 4 weeks and may be the result of a chronic GI tract infection (often associated with immune system compromise), alterations in the motility or integrity of the GI tract, malabsorption disorders, or certain endocrine disorders. Diarrhea that occurs on an episodic basis may be related to a food allergy or may be due to the ingestion of irritants to the GI tract, such as caffeine. Diarrhea in children frequently results from infection, although malabsorption disorders, anatomic defects, and allergy issues can also be causative factors.

Pathophysiologic mechanisms. Four major pathophysiologic mechanisms have been identified in the development of diarrhea: (1) In osmotic diarrhea, increased amounts of poorly absorbable, osmotically active solutes such as a carbohydrate or magnesium sulfate cause sodium and water influx into the bowel lumen, resulting in diarrhea. (2) In secretory diarrhea, a pathophysiologic event such as the presence of a bacterial toxin causes enhanced secretion of chloride ion and water in the small intestine by simultaneously stimulating active secretion and inhibiting resorption. Diarrhea of 1 L or more per day may result from this inappropriate secretion of fluid across the intestinal mucosa. Causes of secretory diarrhea include enterotoxins produced by such organisms as *Vibrio cholerae* and *Staphylococcus aureus.* (3) Exudative diarrhea is the result of exudation of mucus, blood, and protein from sites of active inflammation into the bowel lumen. This creates an increased osmotic load and a subsequent shift of water across the epithelium. In addition, if a large surface area of the bowel has an alteration in its integrity, intestinal absorption will be severely impaired, further compounding the diarrhea produced. Crohn disease and UC are prominent causes of exudative diarrhea. (4) Diarrhea related to motility disturbances is a result of the decreased contact time of chyme with the absorptive surfaces of the intestinal lumen. If inadequate absorption takes place in the small intestine, large amounts of fluid will be delivered to the colon and may overwhelm the absorptive capability of the colon and cause diarrhea. In addition, if the fatty acids and bile salts present in chyme have not been adequately absorbed in the small intestine, they may induce a secretory diarrhea once they reach the colon, further compounding the process of diarrhea formation. Diarrhea associated

with postgastrectomy dumping syndrome and IBS are examples of this type of diarrhea.

KEY POINTS

- Dysphagia is the perception of difficulty in swallowing. Dysphagia caused by neuromuscular disorders may be accompanied by coughing and aspiration, particularly with liquid ingestion. Altered esophageal peristalsis is associated with the sensation that food has become "stuck" behind the sternum. Lower esophageal sphincter (LES) dysfunction may be manifested as substernal pain.
- Pain is a common symptom of GI disorders. A heartburn type of pain is associated with esophageal reflux. Chest pain similar to anginal pain may result from esophageal distention and obstruction. Abdominal pain may be visceral (diffuse, poorly localized), somatic (sharp, well localized), or referred (at a distance from the source but in the same dermatome).
- Nausea and vomiting are manifestations of many GI and other disorders. Alterations in bowel motility or integrity are causative factors. Excess gas may result from altered motility or lack of digestive enzymes. Gas is generated by swallowed air and bacterial action on nutritional substrates.
- Constipation is defined as small, infrequent (fewer than three per week), or difficult bowel movements. Lack of exercise, lack of dietary fiber, slowed peristalsis, and pathologic conditions that alter motility (e.g., obstruction) may produce constipation.
- Diarrhea is defined as an increased frequency and fluidity of bowel movements. Acute infection, stress, fecal impaction, malabsorption disorders, and ingestion of bowel irritants may produce diarrhea. Osmotic diarrhea is due to increased amounts of poorly absorbed solutes in the intestine. Secretory diarrhea is usually due to toxins that stimulate intestinal fluid secretion and impair absorption. Exudative diarrhea (mucus, blood, protein) results from inflammatory processes. A decreased transit time in the small intestine results in diarrhea because the absorptive capacity of the large intestine is exceeded.

DISORDERS OF THE MOUTH AND ESOPHAGUS

The mouth and the esophagus are the portals of entry for nutrients into the GI tract. Impairment in the proper functioning of these structures may have a profound effect on the ability of the individual to ingest adequate nutrients and begin the initial steps of the digestive process. Although disorders of the mouth and esophagus may not be acute, life-threatening emergencies, they may have severe long-term consequences for the well-being of the individual experiencing them.

ORAL INFECTIONS

Stomatitis

Etiology. Stomatitis is defined as an ulcerative inflammation of the oral mucosa that may involve the buccal mucosa, lips, and palate. Among its many causes are pathogenic organisms, including bacteria and viruses; mechanical trauma; exposure to such irritants as alcohol, tobacco, and other chemical substances; certain medications, particularly chemotherapeutic agents; radiation therapy; autoimmune disorders; and nutritional deficiencies, especially vitamin deficiencies. Stomatitis is a central manifestation of several autoimmune disorders, including Reiter syndrome and Behçet syndrome. Stomatitis may also be idiopathic, that is, without identifiable cause.

One of the most commonly encountered types of stomatitis is acute herpetic gingivostomatitis, called more colloquially *cold sores.* The HSV has an affinity for the skin, particularly mucous membranes, and nervous system. This type of stomatitis is commonly acquired by children between the ages of 1 and 3 years, although it can occur at any age. In primary infection, a brief period of prodromal tingling and itching may occur, along with fever and pharyngitis. Vesicles may erupt on any part of the oral mucosa, particularly the tongue, gums, and cheeks. They form on an erythematous base, eventually rupture, and leave a painful ulcer. Once HSV is acquired, it remains latent in the dorsal ganglia of the spinal cord and may reactivate spontaneously or be triggered by physical or emotional stressors.

Treatment. The pharmacologic therapy used for stomatitis depends on its cause. The antiviral drugs acyclovir, famciclovir, and valacyclovir have been approved for treating acute herpetic stomatitis. Stomatitis due to autoimmune conditions such as pemphigus or Behçet syndrome may respond to topical or systemic steroids or other immunomodulating agents. Unfortunately, in a significant number of cases stomatitis is idiopathic or not amenable to specific therapy (e.g., stomatitis attributable to chemotherapy). In all types of stomatitis, measures designed to provide adequate oral hygiene and increase comfort in the oral cavity will be helpful in preventing decreased nutritional intake during the period of inflammation and assist in promoting the healing process. Topical mucosal barriers are often of benefit—one popular newer treatment is Gelclair, a bioadherent oral gel containing maltodextrin, polyvinylpyrrolidone, and sodium hyaluronate. Coating agents with topical steroids (e.g., triamcinolone [Kenalog] in Orabase) are also useful, whereas pentoxifylline, colchicine, dapsone, and thalidomide have been used for recalcitrant cases of idiopathic stomatitis.

ESOPHAGEAL DISORDERS

Gastroesophageal Reflux Disease

Gastroesophageal reflux disease (GERD) is the backflow of gastric contents into the esophagus through the LES. GERD may or may not be symptomatic.

Pathogenesis. GERD is a function of a multifactorial process. Any condition or agent that alters the closure strength and efficacy of the LES or increases intraabdominal pressure may predispose an individual to GERD. For example, the closure strength of the LES may be adversely affected by the intake of fatty foods, caffeine, and alcohol; cigarette smoking; sleep position; or obesity. In addition, pharmacologic agents such as progesterone-containing medications (e.g., birth control pills), narcotics, benzodiazepines, calcium channel blockers, and theophylline may decrease the pressure of the LES. Pregnancy increases the risk of reflux both by increasing intraabdominal pressure and by the effect of hormones on LES muscle tone. Certain anatomic features, especially hiatal hernia, have been associated with GERD. The extent and severity of damage to the esophagus from GERD reflect the frequency and duration of exposure to refluxed material, as well as the volume and acidity of the gastric juices being refluxed. The role of *Helicobacter pylori,* a cause of gastric and duodenal ulceration, in GERD is poorly understood and controversial. In fact, eradication of *H. pylori* for treatment of peptic ulcer disease (PUD) may increase reflux symptoms by unmasking hidden esophagitis.

Clinical manifestations. Most GERD symptoms are attributed to reflux esophagitis, which is esophageal inflammation caused by the highly acidic refluxed material. The most common manifestations of this condition are heartburn, regurgitation, chest pain, and dysphagia. Complications of persistent GERD include esophageal strictures; Barrett esophagus (see Complications section); and pulmonary symptoms related to reflux esophagitis, such as cough, asthma, and laryngitis.

Treatment. Appropriate therapy is directed to increasing LES pressure, enhancing esophageal clearance, improving gastric emptying, and suppressing gastric acidity. Dietary and behavioral changes, such as avoiding tobacco and aggravating food and drink, are advised for all patients, whereas over-the-counter antacids and histamine (H_2)-blocking medications may be effective for occasional GERD. Proton pump

inhibitors (PPIs) are the mainstays of treatment for chronic GERD and have proven very successful in halting and even reversing the changes of chronic GERD. Long-term use of PPIs can lead to malabsorption of vitamin B_{12}, iron, and magnesium, as well as calcium. Studies show a 30% increase in the incidence of hip fracture due to osteoporosis in patients on long-term PPIs.

When reflux esophagitis has progressed in severity, tissue damage, including ulceration, fibrotic scarring, and strictures, may be present in the distal third of the esophagus. Upper GI endoscopy is indicated for patients with ongoing symptoms, and some patients with stricture may require endoscopic dilatation. Surgical intervention, such as thoracoscopic Nissan fundoplication, may be helpful for intractable GERD.

Complications. Barrett esophagus is a serious complication of chronic GERD in which the normal squamous epithelium of the distal esophagus, exposed chronically to acid reflux, is replaced by columnar tissue. It carries a significant risk for esophageal cancer, and patients with Barrett esophagus should undergo regular endoscopic screening for cancer, along with pharmacologic control of their reflux. For patients with documented dysplastic changes, endoscopic eradication therapy (usually with radiofrequency ablation) is a relatively low-morbidity option for treatment.

Hiatal Hernia

A hiatal hernia is a defect in the diaphragm that allows a portion of the stomach to pass through the diaphragmatic opening into the thorax. Two types of hiatal hernia are commonly recognized: (1) a sliding hernia, in which both a portion of the stomach and the gastroesophageal junction slip up into the thorax so that the gastroesophageal junction is above the diaphragmatic opening; and (2) a paraesophageal hernia, in which a part of the greater curvature of the stomach rolls through the diaphragmatic defect (Fig. 36.2). "Mixed" hiatal hernias with features of both of these types may also occur. Sliding hernias are 3 to 10 times more common than paraesophageal and mixed hernias combined. The incidence of hiatal hernia increases with age and occurs more often in women than in men.

Etiology. Although the cause of the anatomic deformity leading to hiatal hernia is not well understood, certain conditions seem to predispose to loosening of the muscular band around the esophageal and diaphragmatic junction. Conditions in which intraabdominal pressure increases, such as ascites, pregnancy, obesity, and chronic straining or coughing, have been associated with the development of hiatal hernia.

Clinical manifestations and treatment. Individuals with hiatal hernia are predisposed to GERD and may experience symptoms such as heartburn, chest pain, and dysphagia. Ulcerations can develop along the mucosal surface of the stomach as it slides through the diaphragmatic opening, so-called *Cameron ulcers.* This is a fairly uncommon cause of chronic upper GI blood loss. A potentially life-threatening situation can develop if a large portion of the stomach becomes caught above the diaphragm and is incarcerated, although this is extremely rare. Medical therapy for hiatal hernia is the same as that for GERD, detailed previously. Indications for surgery include acute incarceration or intractable reflux.

Mallory–Weiss Syndrome

Etiology. Mallory–Weiss syndrome is bleeding caused by a tear in the mucosa or submucosa of the cardia or lower portion of the esophagus. The tear is usually longitudinal, caused primarily by forceful or prolonged vomiting during which the upper esophageal sphincter fails to relax. Approximately 75% of individuals with Mallory–Weiss syndrome are men with a history of excessive ingestion of alcohol or salicylates. Other factors and conditions that may contribute to the development of esophageal tearing in Mallory–Weiss syndrome are coughing, straining during bowel movements, trauma, esophagitis, gastritis, and use of polyethylene glycol as a preparation for colonoscopy. Hiatal hernia is an important predisposing condition.

Clinical manifestations and treatment. Manifestations of Mallory–Weiss syndrome include vomiting blood and passing large amounts of blood per rectum after an episode of forceful vomiting. Epigastric or back pain may also be present. Bleeding may range in severity from mild to massive. It is often profuse when the tear is near the cardia of the stomach and may proceed to fatal shock in this circumstance. Identification is made by endoscopic examination during an episode of acute upper GI bleeding. The majority of patients require at least one blood transfusion, but in most cases bleeding stops spontaneously. Control of active bleeding may be achieved through endoscopic multipolar electric coagulation or similar techniques, epinephrine injection, or interventional radiologic procedures (e.g., vasopressin infusion, Gelfoam embolization). In selected cases, surgical intervention may be necessary.

Esophageal Varices

Esophageal varices represent a complication of portal hypertension, which in Western society is generally the result of cirrhosis attributable to alcoholism or viral hepatitis. In developing tropical countries, chronic infection with the Schistosoma species of liver flukes is a major cause of portal hypertension, along with cirrhosis attributable to chronic hepatitis B infection. Varices will affect more than half of cirrhotic patients, and approximately 30% of these patients experience an episode of variceal hemorrhage within 2 years of the diagnosis of varices. The diagnosis and management of varices are discussed in detail in Chapter 38.

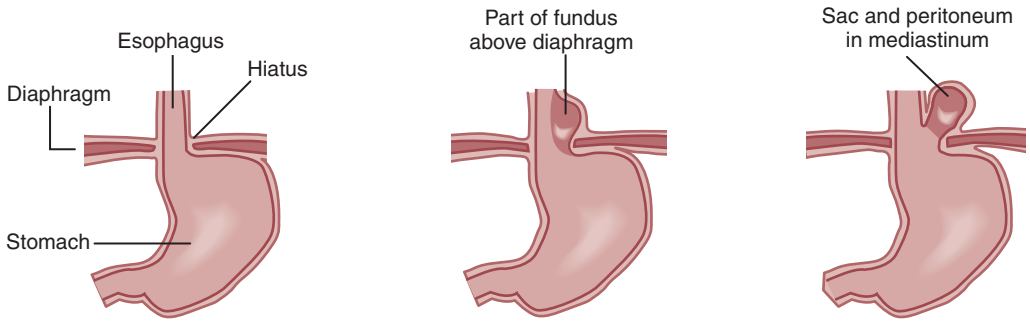


FIG 36.2 Types of hiatal hernia. (From Gould BE: *Pathophysiology for the health professions,* ed 6, Philadelphia, 2018, Saunders.)

KEY POINTS

- Stomatitis is inflammation of the oral mucosa. It may result from pathogenic organisms, trauma, chemical irritants, chemotherapy, radiation therapy, or nutritional deficiencies.
- Common esophageal disorders are gastroesophageal reflux disease (GERD) with esophagitis, hiatal hernia, and bleeding. Reflux esophagitis is manifested as heartburn, chest pain, and dysphagia and may be precipitated by gastric overdistention or poor lower esophageal sphincter (LES) tone. Fatty foods, cigarettes, morphine, theophylline, and progesterone may inhibit LES tone. Chronic GERD may cause Barrett esophagus.
- Hiatal hernias may be sliding or rolling (paraesophageal). Conditions that increase intraabdominal pressure predispose to the development of hiatal hernia. Esophageal reflux often accompanies hiatal hernia, and the manifestations are similar: heartburn, chest pain, and dysphagia.
- Bleeding from the esophagus may pose a life-threatening situation. Mallory–Weiss syndrome is bleeding caused by tears in the lower end of the esophagus or upper part of the stomach. Alcohol and salicylate ingestion appear to be factors. Esophageal bleeding may also be precipitated by coughing, straining, or esophagitis. Rupture of esophageal varices is a dreaded complication of cirrhosis with portal hypertension and carries a high mortality.

ALTERATIONS IN THE INTEGRITY OF THE GASTROINTESTINAL TRACT WALL

Alterations in the integrity of the GI tract may occur at any location along the approximately 30 feet of its length, resulting from infection, an inflammatory process, or weakness of the intestinal wall. Such alterations may present as an acute, life-threatening situation or as a chronic, disabling condition. When the integrity of the GI tract wall is compromised, the ability to perform digestive and absorptive functions may also be compromised because the surface area or motility (or both) is altered.

INFLAMMATION OF THE STOMACH AND INTESTINES

Gastritis

Etiology. Gastritis is defined as inflammation of the stomach lining. Acute gastritis may be triggered by ingestion of toxins such as alcohol, aspirin, or other irritating substances or may occur as a consequence of viral, bacterial, or autoimmune illnesses. (Some experts prefer use of the term "gastropathy" for toxic gastric inflammation, with "gastritis" reserved for gastric inflammation due to infection or autoimmune disorders.) In Western countries, overuse of nonsteroidal antiinflammatory drugs (NSAIDs) and overindulgence in alcohol and tobacco are preeminent causes of acute gastritis.

Pathogenesis. Chronic gastritis is currently the focus of extensive research. The factors promoting the condition have always been poorly understood. However, in 1983 identification of the bacterium *H. pylori* proved to be a landmark event. Since that time, *H. pylori* has generated worldwide attention for its role in the promotion of chronic gastritis, PUD, and gastric carcinoma and lymphoma. Circumstantial evidence suggests that the mode of transmission of *H. pylori* is primarily person to person. Some studies suggest a fecal-oral route, with the possibility of a reservoir in water sources.

It is now known that *H. pylori* causes chronic, superficial gastritis in virtually all infected persons. Once established in the gastric mucosa, *H. pylori* establishes a destructive pattern of persistent inflammation. This persistent inflammation may resolve spontaneously, with clearance of the organism over time, as reflected in a decreased prevalence of *H. pylori* infection among older individuals. Consequences of *H. pylori* gastritis include PUD (discussed in a later section), atrophic gastritis, gastric adenocarcinoma, and mucosa-associated lymphoid tissue lymphoma. The diagnosis and management of *H. pylori* infection will be discussed later.

Clinical manifestations. Although gastritis may be asymptomatic, manifestations of acute gastritis include anorexia, nausea, vomiting, and postprandial discomfort. Occasionally, hematemesis may occur in response to damage to the gastric epithelial mucosa. These manifestations usually disappear when the causative agent is removed and the gastric epithelium undergoes a process of renewal after sloughing off the layer of damaged cells.

Gastroenteritis

Etiology. Gastroenteritis is an inflammation of the stomach and small intestine, and may occur on an acute or chronic basis. Chronic gastroenteritis is usually the result of a primary inflammatory disorder, such as Crohn disease, and is discussed in a later section. Acute gastroenteritis commonly occurs as the result of direct infection of the GI tract lining by a pathogenic organism such as the Norwalk virus; it can also occur from ingestion of preformed bacterial toxins (e.g., *S.* aureus, Bacillus cereus) or bacteria that produce toxins (e.g., Clostridium perfringens). An imbalance in the normal bacterial flora of the GI tract, such as the introduction of an unusual bacterial strain during travel, can trigger GI symptoms as well.

Clinical manifestations and treatment. Acute gastroenteritis in adults is usually a self-limited disease with diarrhea, abdominal discomfort and pain, nausea, and vomiting. An elevated temperature and malaise may also be present. The manifestations vary according to the type of causative pathologic organism and the region of the GI tract affected. Many pathogenic organisms induce a severe secretory type of diarrhea (see the earlier discussion on the pathophysiology of secretory diarrhea). In children and the elderly, fluid losses from diarrhea and vomiting can have serious consequences and could prove life threatening, particularly in underdeveloped countries. Supportive treatment designed to provide fluid and electrolyte replacement is recommended for patients experiencing severe acute gastroenteritis.

Peptic Ulcer Disease

The term *peptic ulcer disease* (PUD) refers to disorders of the upper GI tract caused by the action of hydrochloric acid and pepsin. These disorders are characterized by injury to the mucosa of the esophagus, stomach, or duodenum and may range from a slight irritation to severe ulceration (Figs. 36.3 and 36.4). The GI mucosa is continuously exposed to caustic substances, and mucosal health depends on a number of protective mechanisms. The presence of an intact gastric mucosal barrier and the ability of the mucosa to renew its epithelium serve to protect it against injury. On the other hand, the presence of hydrochloric acid, which potentiates the actions of pepsin and other injurious substances such as aspirin and NSAIDs, will promote injury to the mucosa. PUD thus is the result of an excess of factors that tend to injure the mucosa relative to factors that protect it.

Previously, PUD was attributed to a stressful lifestyle and an irritating diet, and treatment revolved around removing spices from the diet (the much-despised "bland diet") and promoting a more relaxing lifestyle. In recent years, however, research has suggested that the organism *H. pylori* is the major precipitant of PUD, along with NSAIDs. A brief review of the current understanding of PUD pathogenesis will form the basis for further discussion of the manifestations and management of this disease.

Etiology and pathogenesis. Most peptic ulcers are found in the stomach and duodenum. Although the precise mechanisms of ulcer

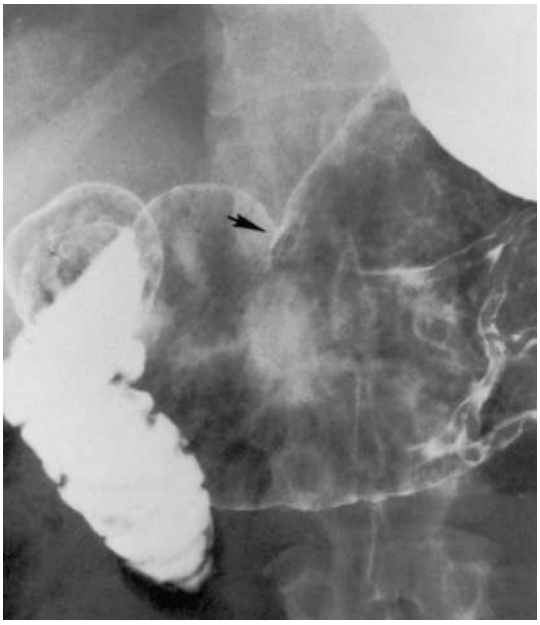


FIG 36.3 Radiograph of an ulcer in the lesser curvature of the stomach *(arrow).* (From Laufer I: *Double contrast gastrointestinal radiology with endoscopic correlation,* Philadelphia, 1979, Saunders.)

formation remain not well understood, the process involves the interplay of mucosal defense mechanisms, pepsin, and acid. The primary event in ulcer formation is thought to be a breakdown in the normally protective epithelial lining of the stomach (Fig. 36.5). The barrier of the epithelial layer and the slightly alkaline layer of mucus may be interrupted by the chronic presence of such injurious substances as aspirin, NSAIDs, alcohol, and bile acids, which may be regurgitated from the duodenum. These substances strip away the surface mucus and cause degeneration of the epithelial cell membranes, with diffusion of hydrochloric acid into the gastric epithelial wall.

Inappropriate excess secretion of acid is a major factor in the development of PUD in the duodenum (Fig. 36.6). Studies have documented that the basal activity of the vagus nerve is increased in persons with PUD of the duodenum, particularly during a fasting state and at night. Vagal activity stimulates the pyloric antrum cells to release gastrin, which travels via the bloodstream and acts on the gastric parietal cells to release hydrochloric acid (HCl). The result is an inappropriately high level of HCl in the duodenum.

H. pylori has a key role in promoting both gastric and duodenal ulcer formation (Fig. 36.7). Up to 75% of persons with duodenal ulcers and 60% of persons with gastric ulcers have *H. pylori* infection. *H. pylori* thrives in acidic conditions; infection renders a person with PUD subject to a slow rate of ulcer healing and a high rate of recurrence, and clearance of *H. pylori* promotes ulcer healing. Although the precise

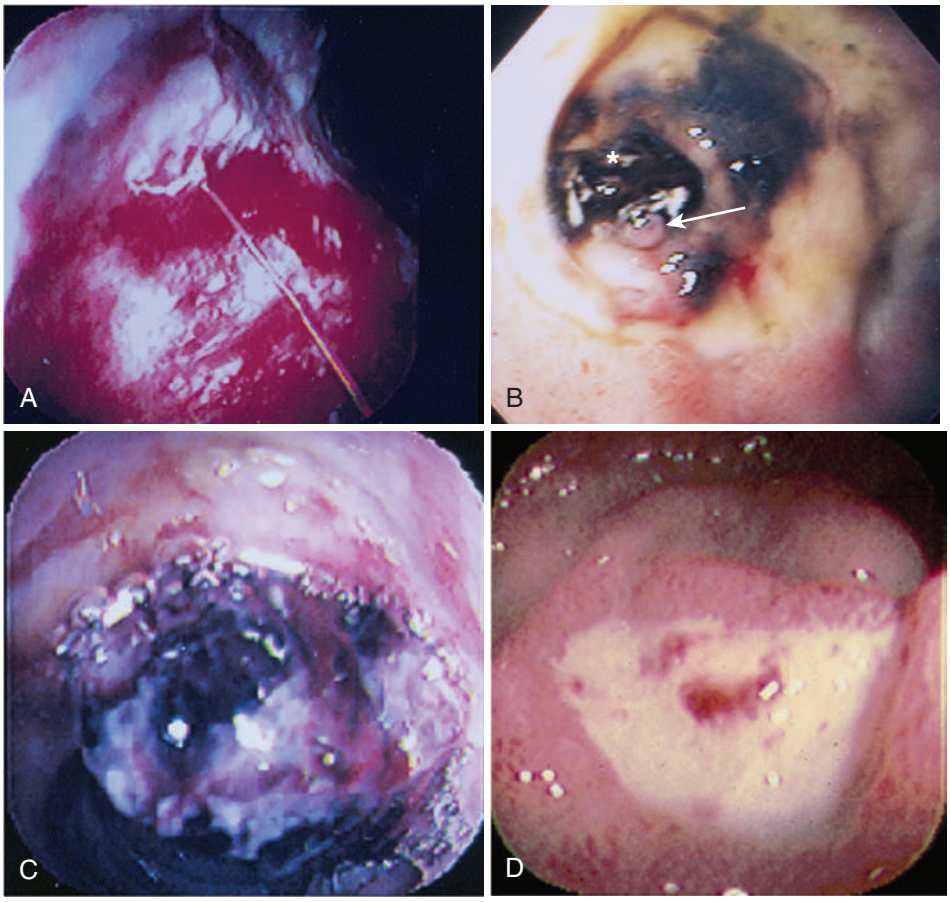


FIG 36.4 Endoscopic stigmata of recent peptic ulcer bleeding. **A,** Active bleeding with spurting. **B,** Visible vessel *(arrow)* with adjacent clot. **C,** Adherent clot. **D,** Slight oozing of blood after washing in the center of the ulcer, without clot or a visible vessel. (From Feldman M, Friedman LS, Brandt LJ, editors: *Sleisenger and Fordtran's gastrointestinal and liver disease,* 10e, Philadelphia, 2016, Saunders.)

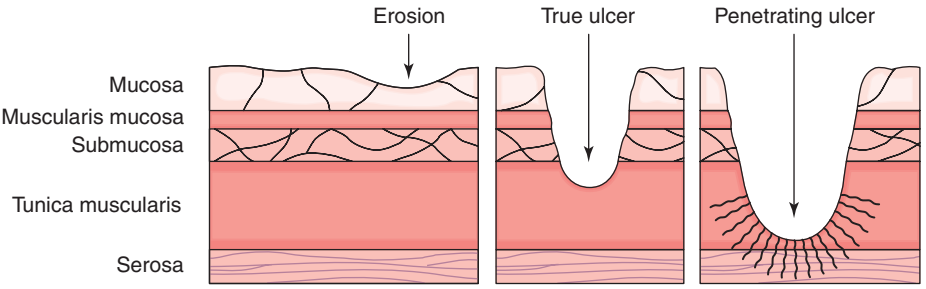


FIG 36.5 Lesions caused by peptic ulcer disease. (From Monahan FD, Sands JK, Neighbors M, Marek JF, Green-Nigro CJ: *Phipps medical-surgical nursing: health and illness perspectives,* ed 8, St Louis, 2007, Mosby.)

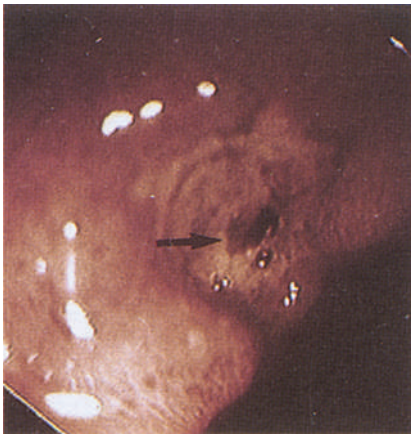


FIG 36.6 Duodenal bulbar ulcer. (From Sleisenger MH, Fordtran JS, editors: *Gastrointestinal disease,* ed 5, Philadelphia, 1993, Saunders.)

mechanisms for the development of PUD remain complex and poorly understood, *H. pylori* virulence is associated with several factors, including the presence of unique, lengthy DNA sequences known as *pathogenicity islands,* particularly cytotoxin-associated gene A (CagA). PUD patients have lower bicarbonate levels in the duodenum as a consequence of the effect of *H. pylori* on the duodenal mucosa.

Other cofactors in the development of PUD have been investigated. Stress has long been considered a key factor in PUD. Glucocorticoid release in response to stress may have a role in promoting excess acid production and inhibiting gastric mucosal defenses. Smoking is also an important environmental risk factor, as identified by epidemiologic studies showing that PUD is twice as likely to develop in smokers as in nonsmokers. In addition, smoking is related to poor ulcer healing and high rates of recurrence. Heredity is thought to have a role in the development of PUD. Certain patterns of gastrin release and pepsin secretion have been identified as genetic traits in families with an increased incidence of PUD. Somewhat surprisingly, given the historical context, there is little evidence of a pathogenic role for alcohol, spicy foods, and caffeine.

Clinical manifestations and diagnoses. Manifestations of PUD include epigastric burning pain that is usually relieved by the intake of food (especially dairy products) or antacids. The pain of gastric ulcers typically occurs on an empty stomach, but may present soon after a meal. Duodenal ulcer pain classically occurs 2 to 3 hours after a meal and is relieved by further food ingestion. Other manifestations that may occur in individuals with PUD include nausea, abdominal upset (dyspepsia), and chest discomfort. A significant proportion of ulcers are asymptomatic, and life-threatening complications, such as GI bleeding, may occur in patients with no warning. The symptoms of PUD are not specific enough to allow for a diagnosis without testing, and malignant conditions can mimic benign PUD.

Diagnosis can be accomplished by upper GI barium contrast radiography or by endoscopy. The finding of a duodenal ulcer indicates a high probability of *H. pylori* and a low probability of malignancy, and the condition can be managed on this basis. All gastric ulcers should be visualized with endoscopy and biopsied to rule out malignancy and confirm the presence of *H. pylori*. Testing for *H. pylori* in patients with uncomplicated duodenal ulcers, given the very high pretest probability, is not usually recommended. Most authorities do recommend testing for this organism in the case of gastric ulcers. Noninvasive testing modalities include the urease breath test, fecal antigen testing, and serologic analysis; the first two are generally preferred, as serologic testing can be misleading in persons with a low pretest prevalence. Invasive endoscopic tests include the tissue urease test, histologic analysis, and bacterial culture.

Treatment. The major treatment objectives for PUD are to encourage healing of the injured mucosa by reducing gastric acidity and to prevent recurrence. PPIs are generally given to block acid secretion. Agents such as sucralfate form a protective coating over the injured mucosa and may be useful under some circumstances. Eradication of *H. pylori* infection with antibiotics is recommended for patients in whom this organism is found, and treatment has led to a marked reduction in the recurrence rate of PUD to less than 10%.

In addition to these pharmacologic strategies, such measures as cessation of smoking, avoidance of aspirin and other NSAIDs, and reduction of stress are all part of a comprehensive program to manage PUD. At the present time, no conclusive research has demonstrated that any specific diet has a therapeutic effect. Susceptible people are generally advised to avoid foods that seem to exacerbate symptoms, including caffeinated beverages and alcohol.

INFLAMMATORY BOWEL DISEASE

The term *inflammatory bowel disease* (IBD) refers to the two separate disease entities of UC and Crohn disease. IBD is generally a life-altering chronic illness with serious consequences for people and their families who must cope with it, particularly when it occurs early in life. Both UC and Crohn disease have onsets most commonly in childhood or young adulthood, with a possible second peak between age 50 and 80 years. There is a slight female predominance in the incidence of Crohn disease and a higher incidence of UC among males.

IBD is typically characterized by exacerbations and remissions. Its causes are poorly understood, but recent research has focused on genetic, environmental, and immunologic factors. About 10% to 25% of IBD patients have a first-degree relative with either UC or Crohn disease. Both IBD illnesses are more common in individuals with Jewish ancestry and in whites compared with black and Hispanic populations.

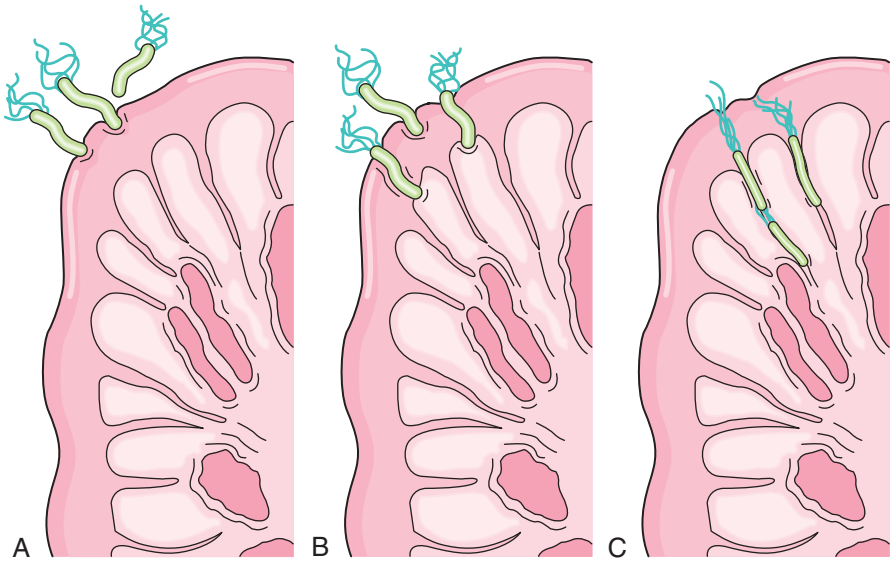


FIG 36.7 Penetration of the mucosal layer by *Helicobacter pylori.* **A,** After penetration, *H. pylori* forms clusters near membranes of surface epithelial cells. **B,** Some attach to the cell membrane. **C,** Others lodge between the epithelial cells.

FIG 36.8 Ulcerative colitis. (From Sleisenger MH, Fordtran JS, editors: *Gastrointestinal disease,* ed 5, Philadelphia, 1993, Saunders.)

Environmental factors have also been investigated. Cigarette smoking has been shown to increase the risk for Crohn disease, whereas it may be protective of the development of UC. Appendectomy may protect against the development of UC, but not Crohn disease. Some features of a "Western" style diet, with highly processed foods high in carbohydrates, are associated with an increased risk of developing Crohn disease and possibly UC. Obesity has been shown to increase disease activity in patients with Crohn disease. An important immunologic basis for IBD is supported by the fact that it frequently accompanies other autoimmune conditions such as thyroid disease and pernicious anemia.

Ulcerative Colitis

Ulcerative colitis (Fig. 36.8) is an inflammatory disease of the mucosa of the rectum and colon. Approximately one fifth of patients have total colitis, one-third have subtotal disease extending beyond the sigmoid, and one-half have disease limited to the rectum and rectosigmoid. The changes are usually most severe in the rectum and extend for a variable extent around the colon, although there are several exceptions to this general rule. The annual incidence in North America has been estimated between 2 and 14 per 100,000 persons.

Etiology and clinical manifestations. UC begins as an inflammation at the base of the crypts of Lieberkühn. Damage to the crypt epithelium results, with eventual invasion of leukocytes and the formation of abscesses in the crypts. When multiple abscesses form in close proximity and begin to coalesce, large areas of ulcerations develop in the epithelium. Concurrent with this destructive process are attempts at repair of damaged tissue, along with the development of fragile and highly vascularized granulation tissue. The manifestations of UC are the result of these processes and include abdominal pain, diarrhea, and rectal bleeding. Bleeding occurs as a result of mucosal destruction and ulceration, as well as damage to newly developed granulation tissue. Diarrhea is a result of the mucosal destruction in the colon, which leads to a decreased ability of the bowel to absorb water and sodium and thus to an increased volume of fluid in the intestinal contents.

The progression of UC is highly variable. In some individuals it may have very mild manifestations; in others it may rapidly progress to a life-threatening disorder. Approximately 5% to 10% of persons with UC have only one attack, with no further recurrence. However, 65% to 75% of those with UC experience an intermittent series of exacerbations and remissions. Rarely, patients with UC will manifest toxic megacolon, a life-threatening condition in which the colon becomes massively enlarged. This condition requires urgent treatment and may necessitate emergency colectomy.

A number of conditions in other organ systems complicate UC, the most devastating of which is the relentlessly progressive liver condition primary sclerosing cholangitis (PSC), which occurs in 3% of UC patients. An additional concern is increased risk for the development of colon cancer in persons who have had UC for more than 7 to 10 years. Authorities recommend monitoring these individuals carefully with regular endoscopy and biopsy. The presence of high-grade dysplasia should prompt consideration of prophylactic complete colectomy. Recent surgical advances, such as the ileoanal pouch, have allowed colectomy patients to avoid colostomy and have close to normal bowel function.

Treatment. Management of UC is complex and ever evolving. Corticosteroids have long been the mainstay of treatment of acute

exacerbations, but side effects limit their long-term use. Patients with signs of systemic toxicity, especially those with impending or full-blown toxic megacolon, should receive broad-spectrum antibiotics, including an agent effective against anaerobic bacteria (e.g., metronidazole). Important categories of disease-modifying agents include the salicylate analogs, and immunomodulating agents such as azathioprine and mercaptopurine. Intravenous followed by oral cyclosporine is a relatively recent treatment advance for steroid-refractive UC that may help patients avoid colectomy. Antitumor necrosis factor (anti-TNF) therapy with infliximab and adalimumab has also been shown to be effective in patients with moderate to severe UC, though there are important cost and side effect issues.

Crohn Disease

Crohn disease, also called *regional enteritis* or *granulomatous colitis,* is an inflammation of the GI tract that extends through all layers of the intestinal wall (Fig. 36.9). It most commonly affects the proximal portion of the colon and, less often, the terminal ileum. It may affect multiple portions of the colon, with intervening normal areas left between the affected regions. The manifestations of Crohn disease differ in some respects from those of UC, although some overlap may occur and distinction may be difficult, even after pathologic examination. In Crohn disease, abdominal pain is often constant and in the right lower quadrant of the abdomen. A palpable abdominal mass may be present in the right lower quadrant, reflecting significant ileocecal involvement. The stool may be bloody, although not usually to the extent of that seen with UC. The cause of Crohn disease is unknown at the present time. There are fascinating parallels with UC as well as unexpected distinctions. For instance, smoking has been shown to protect against UC but to increase the risk of Crohn disease. The annual incidence of Crohn disease in North America has been estimated from 3 to 15 per 100,000, and up to 5% of people with Crohn disease have one or more affected relatives.

Etiology and pathogenesis. Crohn disease appears to result primarily from a process in which the lymphoid and lymphatic structures of the GI tract become blocked. Subsequent engorgement and inflammation of surrounding tissue lead to the development of deep linear ulcers in the bowel wall. Eventually, all layers of the GI tract wall may become involved, and the portion of intestine that is affected may become thickened by fibrous scar tissue. Deep fissures may develop into fistulas, which may extend into adjacent tissue of other organs such as the bladder wall or even the skin. One of the cardinal features of Crohn disease on histopathologic analysis is granulomas, which is generally diagnostic of this disorder.

Clinical manifestations. The manifestations of Crohn disease are the result of the pathologic changes in which the bowel becomes incapable of adequately absorbing the intestinal contents. Complications such as perianal fissures, fistulas, and abscesses are common in Crohn disease

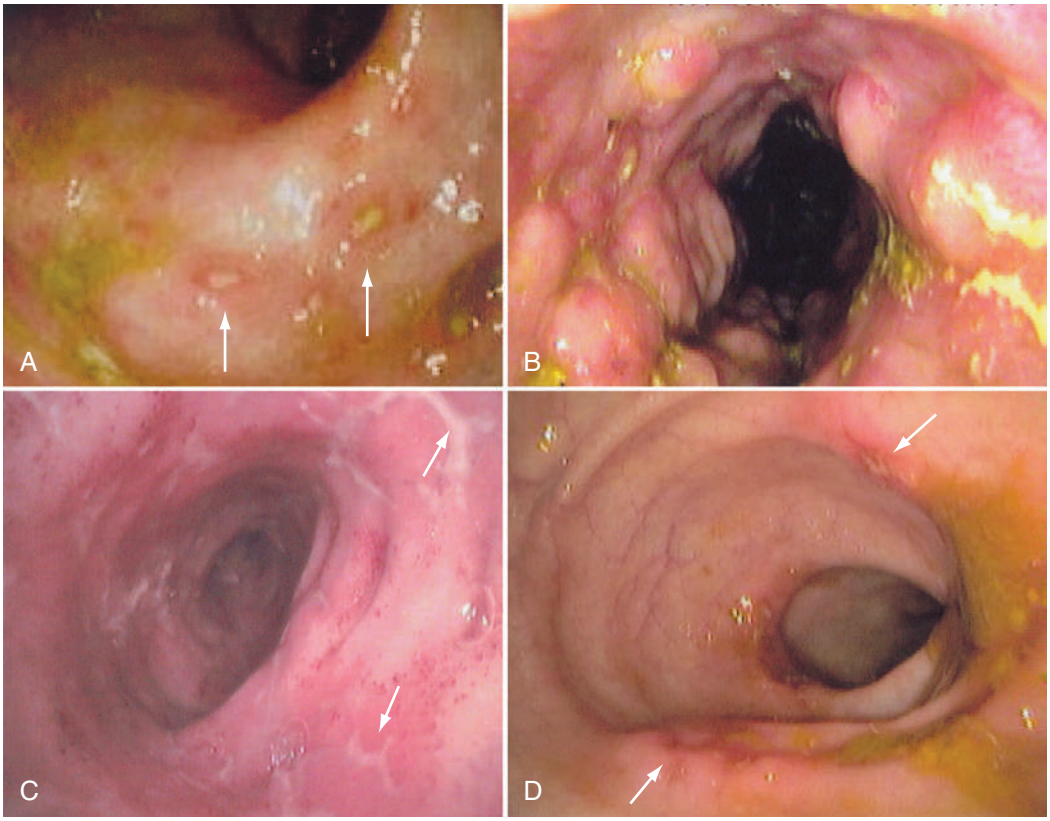


FIG 36.9 Endoscopic appearance of Crohn disease. A wide variety of findings may be visualized on endoscopy, in part depending on the duration and severity of the inflammation. **A,** Typical aphthous ulcers *(arrows),* consisting of a central white depression surrounded by a slightly elevated, erythematous rim only a few millimeters in diameter. **B,** Findings more typical of advanced disease, with erythema, edema, and a cobblestoned appearance. **C,** Stellate ulcers *(arrows)* in the terminal ileum. **D,** Discrete ulcers *(arrows)* with normal intervening mucosa, typical of the patchy inflammation seen in Crohn disease. (From Feldman M, Friedman LS, Brandt LJ, editors: *Sleisenger and Fordtran's gastrointestinal and liver disease,* 10e, Philadelphia, 2016, Saunders.)

and may be the symptoms that lead individuals to seek health care. The onset and course of Crohn disease may vary a great deal, and the symptoms present during a period of exacerbation may be subtle but persistent. At the present time, it is unclear whether a significantly increased incidence of intestinal cancer occurs in persons with Crohn disease. However, when Crohn disease involves the large bowel, the risk of colorectal cancer appears to be similar to that for UC of similar extent. There is known to be an increased risk of lymphoma in patients on long-term thiopurines like azathioprine, in particular, in combination with anti-TNF agents. Toxic megacolon and PSC also occur with Crohn disease, but are much less frequent than in patients with UC. Many extraintestinal manifestations occur, including spondyloarthritis and uveitis.

The diagnosis of Crohn disease is typically made on the basis of the clinical history, radiographic changes, and typical biopsy findings of granulomatous intestinal inflammation. Several immunologic tests are also available, in particular, testing for antibodies directed against nuclear cytoplasmic antigens, the yeast Saccaromyces cerevesiae, and the Omp-C antigen, but these are expensive and generally reserved for cases where the diagnosis is uncertain.

Treatment. Because the etiology of Crohn disease is unknown, therapeutic strategies are focused on alleviating and reducing inflammation and symptoms. Therapeutic drug categories are similar to those for UC. Prednisone or sulfasalazine is generally used as initial therapy to achieve remission. The antibiotic metronidazole is particularly useful for colonic Crohn disease. Options for the treatment of refractory patients include azathioprine, 6-mercaptopurine, methotrexate, and biological therapies.

Newer treatment options include the anti-TNF agents infliximab, adalimumab, and certolizumab and the newer antiintegrin agents natalizumab, vedolizumab and ustekinumab. All of these newer options have shown success in trials and are indicated for refractory Crohn disease. All of these agents have the significant cost and side effect issues noted earlier in the section on UC.

Despite rapid progress in treatment modalities over the last decade, there is no cure for this challenging condition.

ENTEROCOLITIS

Antibiotic-Associated Colitis (Pseudomembranous Colitis)

Etiology. Antibiotic-associated colitis (AAC), also known as pseudomembranous colitis, is an acute inflammation and necrosis of the large intestine caused by Clostridium difficile, usually affecting the mucosa but sometimes extending to other layers. Exposure to antibiotics is the major factor predisposing to the development of this disorder, and patients with cancer or who have undergone abdominal surgery are at particular risk. The disease is mediated by bacterial toxins, leading to mucosal necrosis and the characteristic pseudomembrane composed of leukocytes, mucus, fibrin, and inflammatory cells.

Clinical manifestations and treatment. Resulting manifestations include diarrhea (often bloody), abdominal pain, fever, leukocytosis, and, rarely, toxic megacolon or colonic perforation. AAC is a major cause of fever and leukocytosis (elevated white blood cell count) among hospitalized patients receiving antibiotics. Treatment involves stopping the offending antibiotic, treating ischemia and other contributing conditions if present, and using antibiotics directed against *C.* difficile. Important antibiotics effective against this organism include metronidazole, oral vancomycin, and oral fidaxomicin. Recurrences are relatively common and may necessitate retreatment. In rare cases a fecal transplant (transfer of fecal material from another healthy person to the source patient via colonoscopy, enema, or gastric tube) or colectomy may be necessary to clear the infection. Much research is underway to prevent and treat this condition, and in particular vaccination for prevention and treatment with lyophilized fecal flora look promising.

Necrotizing Enterocolitis

Etiology. Necrotizing enterocolitis (NEC) is a disorder occurring most often in premature infants (less than 34 weeks' gestation) and infants with low birth weight (less than 5 lb or 2.25 kg). This disorder is characterized by diffuse or patchy intestinal necrosis accompanied by sepsis.

Clinical manifestations and treatment. Early manifestations include a distended abdomen and stomach. The major complication of NEC is intestinal perforation, which may necessitate surgery. Various theories regarding the etiologic progression of NEC include perinatal oxygen deficit with insufficient blood flow to the viscera and the use of hypertonic feeding formulas in newborn infants. More than 90% of infants with NEC have a history of milk feeding as well. A special form of NEC called *typhlitis* may afflict adult cancer patients with neutropenia and carries a grave prognosis. Management of neonatal and adult forms of NEC includes careful supportive care, including fluid management and administration of broad-spectrum antibiotics. Early surgical consultation is essential for this syndrome, and surgery is necessary for patients with evidence of significant ischemia or perforation.

Appendicitis

Etiology. The most common cause of emergency surgery on the abdomen, appendicitis is an inflammation of the vermiform appendix. The classic hypothesis suggests that obstruction of the appendiceal lumen by a fecalith causes most cases of appendiceal inflammation. Less commonly, lymphoid hyperplasia or parasitic worms may lead to appendicitis. In an unknown number of cases appendiceal inflammation may be self-limited and may remit (e.g., with relief of the obstruction). If left unchecked, inflammation generally leads to necrosis of the appendix, with subsequent abscess formation and life-threatening peritonitis. Rarely appendicitis may occur in a subacute or stuttering fashion over several days or weeks.

Clinical manifestations and treatment. Appendicitis is two times more likely to occur in individuals younger than age 45 compared with those 45 years and older, and it affects men more often than women. The peak incidence is between ages 10 and 19 years. The earliest manifestation of appendicitis is generalized periumbilical pain accompanied by nausea and, occasionally, diarrhea. The pain is often described as "migrating" or localizing to the lower right abdomen (McBurney point) because of distention of the serosa from inflammatory edema, at which time fever usually manifests. Experienced surgeons generally operate in suspicious cases. Less typical cases should be assessed with computed tomography (CT) or with ultrasound if the patient is a child or pregnant woman or if CT is not readily available. Such an approach yields a relatively low false-positive surgical rate of around 5%. Surgical removal of the appendix, either through an open procedure or laparoscopically, is the treatment of choice for appendicitis. Administration of antibiotics with replacement of fluid and electrolytes is usually necessary. Localized abscesses secondary to perforation may be managed with percutaneous tube drainage and antibiotics alone if there are no signs of diffuse peritonitis. Delayed appendectomy is usually carried out several weeks later, after the initial inflammation has subsided.

Diverticular Disease

Etiology. The term *diverticular disease* generally refers to diverticulosis, or the presence of diverticula in the colon. Diverticula are acquired herniations of the mucosa and submucosa through the muscular coat of the colon (Fig. 36.10) that likely result from a combination of structural and functional factors. In particular, areas of weakness in the bowel

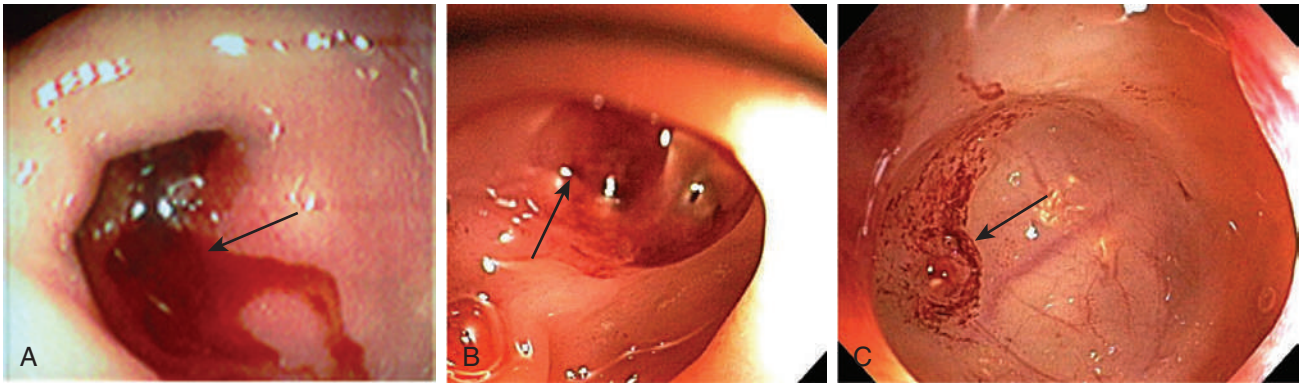


FIG 36.10 Endoscopic stigmata of recent colonic diverticular bleeding. **A,** Active bleeding *(arrow)*. **B,** Adherent clot *(arrow)*. **C,** Nonbleeding visible vessel *(arrow)*. (From Feldman M, Friedman LS, Brandt LJ, editors: *Sleisenger and Fordtran's gastrointestinal and liver disease,* 10e, Philadelphia, 2016, Saunders.)

BOX 36.1 Terminology of Diverticulosis

Diverticulum: A single pouchlike herniation through the muscular layer of the colon
Diverticula: More than one diverticulum (Latin plural form)
Diverticulosis: The presence of one or more diverticula
Diverticulitis: Inflammation of one or more diverticula
Diverticular disease: Complications related to the presence of diverticula

wall, especially where blood vessels enter, are subject to damage from high intraluminal pressures. Colonic diverticulosis is common in Western countries and is associated with a diet low in fiber; this lack of fiber presumably fails to provide enough bulk to dampen pressure variations in the intestine. The prevalence of diverticulosis increases with age; about 30% of the general population at 60 years of age and about 80% at 80 years will have diverticula in the colon. Most persons experience no manifestations of diverticulosis, and by itself diverticulosis is not considered a pathologic condition. However, when diverticula become inflamed, the condition is referred to as *diverticulitis* (see Box 36.1 for the terminology of diverticulosis).

Clinical manifestations and treatment. Inflammation of diverticula can lead to serious consequences such as intestinal obstruction and perforation with the development of abscesses in the bowel wall or peritonitis. Manifestations of diverticulitis include acute lower abdominal pain (usually left lower quadrant in location), fever, and leukocytosis. Constipation is common, but 25% of patients may have diarrhea. During an acute episode of diverticulitis, the administration of broad-spectrum antibiotics is indicated, and on occasion percutaneous or surgical drainage of an abscess may be necessary. Recurrence of diverticulitis is common. Long-term complications include colonic strictures and fistulas, which may necessitate surgery. Surgery is also recommended for patients with one or more recurrences of diverticulitis.

KEY POINTS

- Alterations in intestinal wall integrity are generally a result of infection, inflammation, or weakness of the muscular layers. General symptoms include pain, bleeding, and diarrhea.
- Gastritis may be acute or chronic. Acute gastritis is generally precipitated by the ingestion of irritating substances, including alcohol and aspirin. Chronic gastritis may lead to atrophy of the gastric mucosa and the subsequent decreased production of HCl and intrinsic factor. Acute gastroenteritis is usually due to the ingestion of pathogenic organisms or preformed bacterial toxins and is characterized by self-limited vomiting, diarrhea, and abdominal pain.
- Peptic ulcer disease (PUD) may affect the esophagus, stomach, and duodenum. Gastric ulcers are thought to be due to the breakdown of the protective mucous layer that normally prevents the diffusion of acids into gastric epithelia. Duodenal ulcers are caused by excessive acid secretion that is mediated by increased vagal activity. The organism *H. pylori* has been implicated in the pathogenesis of both gastric and duodenal ulcers. PUD is characterized by epigastric pain that is relieved by food or antacids. Perforation and bleeding are the major complications. Management of PUD is aimed at minimizing acid secretion and eradicating *H. pylori*.
- Ulcerative colitis (UC) and Crohn disease are chronic inflammatory disorders of the bowel. UC (inflammation and ulceration of the colon and rectal mucosa) is manifested as bloody diarrhea and abdominal pain. There is an increased risk for colon cancer in persons who have had UC for more than 7 to 10 years. Crohn disease generally affects the proximal portion of the colon or the terminal ileum. Involvement of all layers of the intestinal wall predisposes to fistula formation and malabsorption. Crohn disease may result from blockage and subsequent inflammation of lymphatic vessels. Chronic abdominal pain and diarrhea are common. Management of UC and Crohn disease is aimed at reducing inflammation and subsequently trying to maintain remission.
- Antibiotic-associated colitis (AAC) or pseudomembranous colitis is associated with *C. difficile* infection. Abdominal pain, diarrhea, fever, and sepsis may result. The use of broad-spectrum antibiotics has been implicated in the etiologic development of AAC. NEC, which occurs most often in infants, is thought to be due to bowel ischemia.
- Appendicitis is characterized by right lower quadrant pain, nausea and vomiting, and systemic signs of inflammation. Surgical removal of the appendix is necessary. Untreated appendicitis may result in rupture of the appendix and subsequent peritonitis; localized abscesses may be managed with tube drainage and antibiotics alone.
- Diverticula of the colon are common in Western society because of a low intake of dietary fiber. Low-bulk stools result in the development of high intraluminal pressure, which predisposes to diverticula formation. Diverticulosis is generally asymptomatic. Inflammation of the diverticula, or diverticulitis, is manifested as fever and lower abdominal pain. Antibiotics and surgery may be required for management of complicated diverticulitis.

ALTERATIONS IN MOTILITY OF THE GASTROINTESTINAL TRACT

Disorders of the GI tract that alter its regular propulsive ability may have a negative effect on nutrient absorption. In the case of increased motility, the transit time of substances passing through the GI tract may be too fast to allow for adequate absorption. Conversely, a blockage or constriction of the GI tract may result in slowed or absent motility, which also prevents normal ingestion and processing of nutrient substances. As with alterations in the integrity of the GI tract wall, these alterations in GI motility may be acute or chronic, with many implications for the lifestyle of the patient.

MOTILITY DISORDERS

Irritable Bowel Syndrome

Irritable bowel syndrome is a complex entity that remains incompletely understood despite decades of intensive research. A clear definition of this syndrome has not yet been decided by all authorities; nevertheless, certain defining characteristics have been established. Typically, IBS is the presence of alternating diarrhea and constipation accompanied by abdominal cramping pain in the absence of any identifiable pathologic process in the GI tract. (Other terms that have been used for this syndrome include *spastic colitis* and *irritable colon syndrome.*) Many authorities emphasize that the quantity of symptoms is not as important as their effect on the normal lifestyle of an individual. Persons with IBS may miss work, curtail their social life, and avoid sexual intercourse. This is an extremely common disorder, affecting up to 20% of the U.S. population. It is important to differentiate IBS, in which no pathologic process of the GI tract has been identified, from inflammatory bowel disease, in which a specific pathologic process is identifiable. There are two classifications of IBS: IBS diarrhea and IBS constipation.

Etiology and pathogenesis. The etiologic factors and pathogenesis of IBS are presently obscure. Most evidence seems to show that IBS is primarily a disorder of bowel motility. Studies have demonstrated that the myoelectric activity of the colon in persons with IBS is altered. In particular, the slow wave activity of the colon, which usually occurs at a rate of three to six times per minute, is markedly increased in IBS. Moreover, the sensory response to distention and stimulation seems to be heightened. Whether these findings are the result of genetic factors or such environmental factors as episodic infection, psychological stressors, or dietary patterns remains unknown. The role of sensitivity to substances in ingested foods such as gluten and the contribution of bacterial overgrowth to symptoms are among current areas of exploration.

Clinical manifestations and treatment. The manifestations of IBS may vary greatly, with some persons experiencing only diarrhea or constipation and others experiencing an alternating pattern of both. In addition to cramping abdominal pain, manifestations such as nausea and mucus in the stool may be present. The severity of manifestations ranges from barely noticeable to incapacitating. Current therapy focuses on dietary modification and on the use of antidiarrheal/antispasmodic agents and laxatives (for constipation-predominant IBS) as appropriate. The 5-hydroxytryptamine-3 (serotonin, 5-HT_3) receptor antagonist alosetron has shown favorable results for the diarrhea-predominant form of IBS, though it has been associated with significant side effects, necessitating the Food and Drug Administration to restrict use. Lubiprostone is a locally acting chloride channel activator that enhances chloride-rich intestinal fluid secretion and is useful for patients with primarily constipative IBS.

Ingestion of a diet with increased amounts of fiber has proved useful in many cases and is thought to promote a more normal pattern of myoelectric activity by providing a regular propulsive stimulus in the gut. Perhaps more than with most GI disorders, patients with IBS may benefit from support groups, Internet-based resources, and alternative therapies.

Intestinal Obstruction

Intestinal obstruction is partial or complete blockage of the intestinal lumen of the small or large bowel. Mechanical obstructions are caused by blockage of the intestine by adhesions, hernia, tumor, inflammation, stricture (as in Crohn disease), impacted feces, volvulus, or intussusception. (Volvulus and intussusception are covered in more detail in the following sections.) Functional obstruction or ileus refers to the loss of propulsive ability by the bowel and may occur after abdominal surgery or in association with hypokalemia, peritonitis, severe trauma, spinal fractures, ureteral distention, and the administration of medications such as narcotics. Ogilvie syndrome, also known as *intestinal pseudoobstruction,* is a rare severe motility problem characterized by recurrent bouts of ileus.

Etiology and pathogenesis. The most common location for GI obstruction is the small bowel (90% of cases). The most frequent contributing factors are previous abdominal surgery with adhesions and congenital abnormalities of the bowel. Metastatic carcinoma, particularly cancer of the intestinal tract or female reproductive organs, is an important cause of obstruction and should be considered in patients with obstruction who have never had abdominal surgery. The severity and types of symptoms initially accompanying an intestinal obstruction vary with its cause and location.

With obstruction of the bowel lumen, fluid and gas begin to accumulate proximal to the obstructed location. The distention produced by trapped fluid and gas causes water and electrolytes to be secreted into the obstructed lumen of the small bowel. Distention also results in the impedance of venous return, and the bowel wall becomes edematous. The absorptive ability of the bowel wall is compromised, and fluid and gas continue to accumulate as additional water and electrolytes are secreted into the lumen. The pressure on the bowel wall exerted by the excess fluid and gas may result in leakage of fluid through the wall into the peritoneum, as well as necrosis of the bowel wall.

In addition to the process just described, other complications may be present with blockage of the intestinal lumen. Impairment of bowel circulation leads to ischemia, a process referred to as *strangulation.* Bacteria and bacterial endotoxins may translocate across the bowel wall into the bloodstream to produce fever and other signs of sepsis. As blood escapes from the engorged veins, significant loss of blood and plasma from the affected segment may result in the rapid development of shock. In addition, the strangulated segment may become gangrenous, with resulting peritonitis, or become perforated, with the leakage of highly toxic bacterial material into the peritoneal cavity. If left untreated, a person with an intestinal obstruction of the small bowel has a high risk of death from shock and vascular collapse.

Clinical manifestations and treatment. The manifestations of an intestinal obstruction depend on its site and duration. Obstructions in the upper jejunal area usually result in vomiting, dehydration, and electrolyte depletion. In obstructions of the distal portion of the small bowel or ileum, constipation may be an early manifestation, with massive accumulation of fluid in the lumen occurring later. Dehydration may progress to hypovolemic shock if the obstruction is left untreated. In obstructions of the colon, massive gas distention may be present. The fluid and electrolyte losses associated with colonic obstruction may not be as severe as those seen in obstruction of the small bowel. Blockage of the colon by a tumor is the most common cause of colonic obstruction,

and perforation of the bowel wall adjacent to the tumor may occur in association with an obstruction.

Therapeutic strategies for intestinal obstruction include surgical intervention to correct or remove the source of a mechanical obstruction. Supportive therapy, including decompression of the bowel with specialized tubes or endoscopy, and fluid and electrolyte replacement therapy may be needed during an acute obstructive episode.

Volvulus

Volvulus is the twisting of the bowel on itself, causing intestinal obstruction and blood vessel compression (Fig. 36.11). The two most common sites for the development of volvulus are the cecum and the sigmoid colon. A volvulus may be the result of an anomaly of rotation, an ingested foreign body, or an adhesion; however, the cause cannot always be determined. Volvulus tends to occur in elderly individuals with coexistent medical conditions. With the sudden tight twisting of the bowel on its mesentery, blood flow to the bowel is impeded. Gangrene, necrosis, and perforation may develop, resulting in a life-threatening situation. If both ends of a bowel segment are twisted, a closed-loop obstruction results, with the manifestations described earlier for intestinal obstruction. Treatment varies according to the severity and location of the volvulus and includes the therapeutic approaches described for intestinal obstruction.

Intussusception

Intussusception is a telescoping or invagination of a portion of the bowel into an adjacent distal portion (Fig. 36.12). It is most common in infants and occurs three times more often in males than in females. In most cases involving infants, the actual cause is unknown, although intussusception has been linked to viral infections and the use of some forms of rotavirus vaccine. In older children, it may be associated with alterations in intestinal motility or a condition called *Meckel diverticulum,* in which a congenital abnormality consisting of a blind tube is present in the distal end of the ileum. In adults, intussusception usually results from the presence of benign or malignant tumors.

As a bowel segment undergoes intussusception, peristalsis acts to pull more bowel along with it. The resulting area of tightened, invaginated bowel becomes edematous; venous engorgement with hemorrhage may occur. Intestinal obstruction of the bowel may develop, with eventual gangrene, shock, and perforation of the bowel if surgical treatment is delayed.

Megacolon

Megacolon can be congenital or acquired at any age. Perhaps the most common cause in Western countries is prolonged constipation/obstipation, usually chronic in nature. This is particularly common in younger children who are dealing with the psychological aspects of toilet training and bowel control. Although most of these children are psychologically normal, a small number of children with encopresis have experienced sexual abuse, and its presence should be considered. Significant voiding issues also plague the other end of the age spectrum, and constipation/obstipation in the elderly may lead to megacolon, particularly when the sufferer has relied on regular enemas or laxatives for many years.

Hirschsprung disease (see the following section) is characterized by the congenital absence of autonomic smooth muscle ganglia. The aganglionic bowel segment contracts, but without the reciprocal relaxation needed to propel the intestinal contents forward. Stasis of stool and dilation of the proximal end of the colon result in megacolon, or massive dilation of the colon. Chagas disease caused by infection with Trypanosoma cruzi is a common cause of acquired colon neuronal dysfunction and megacolon in Central and South America, but is rarely seen in the United States, except among immigrants from endemic countries. As discussed earlier in the chapter, AAC (pseudomembranous colitis) may result in acute megacolon, which is a surgical emergency. Finally, the idiopathic syndrome of intestinal pseudoobstruction (Ogilvie syndrome) may rarely result in megacolon.

Hirschsprung Disease

Hirschsprung disease is a congenital disorder of the large intestine in which the autonomic nerve ganglia in the smooth muscle are absent or markedly reduced in number. In 90% of individuals with Hirschsprung disease, the aganglionic segment is in the rectosigmoid area, but occasionally the entire colon may be affected. Hirschsprung disease occurs in approximately 1 in 5000 live births and is occasionally familial. It is more common in males than in females, with a ratio of 3.8:1. The disease often coexists with other anomalies, particularly Down syndrome. Although Hirschsprung disease is most commonly identified in infants and children, it may be present in adults as a long-standing undiagnosed condition.

In infants, Hirschsprung disease may have severe, life-threatening effects. Fecal stagnation may result in enterocolitis with bacterial overgrowth, profuse diarrhea, hypovolemic shock, and intestinal perforation. Interventions such as colonic lavage may be performed to empty the bowel until the infant is stable enough to withstand surgical intervention, which is the definitive treatment.

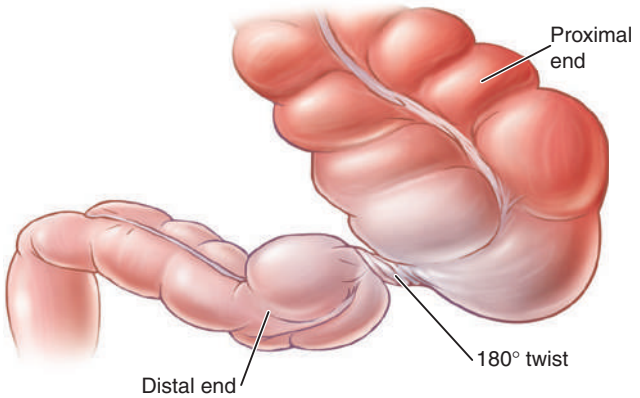


FIG 36.11 Volvulus. Intestine twists at least 180 degrees, causing obstruction and ischemia. (From Black JM, Hawks JH: *Medical-surgical nursing: clinical management for positive outcomes,* ed 8, Philadelphia, 2009, Saunders, p 714.)

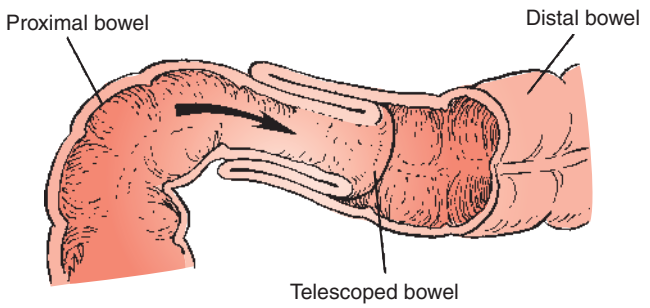


FIG 36.12 Intussusception. A portion of bowel telescopes into adjacent (usually distal) bowel. (From Black JM, Hawks JH: *Medical-surgical nursing: clinical management for positive outcomes,* ed 8, Philadelphia, 2009, Saunders, p 714.)

KEY POINTS

- Irritable bowel syndrome (IBS) is manifested by bouts of alternating diarrhea and constipation in the absence of an identifiable GI pathologic process. The cause is unclear; however, the slow wave activity of the bowel is markedly increased. A high-fiber diet and antidiarrheal agents may be recommended.
- Intestinal obstructions may be mechanical or functional. Mechanical obstructions are due to adhesions, hernia, tumors, impacted feces, volvulus (twisting), or intussusception (telescoping). Mechanical obstructions are characterized by increased bowel sounds initially, accompanied by abdominal pain, nausea, and vomiting. Functional obstructions are due to conditions that inhibit peristalsis, such as use of narcotics, exposure to anesthetic agents, surgical manipulation, peritonitis, hypokalemia, and spinal cord injuries. Functional obstructions are characterized by the absence of bowel sounds. Uncorrected obstruction may lead to intestinal wall edema, ischemia, and necrosis. Bowel gangrene, sepsis, and shock can result. Surgical intervention or decompression with an intestinal tube is often required.
- Hirschsprung disease is a familial, congenital disorder of the large intestine in which the autonomic ganglia are reduced or absent. Stasis of stool and megacolon may occur in the abnormally innervated section of bowel. Megacolon can also be acquired as an adult.

DISORDERS OF MALABSORPTION

Malabsorption refers to failure of the GI tract to absorb or normally digest one or more dietary constituents. It is typically manifested as diarrhea, with the passage of inappropriately processed intestinal contents resulting in impaired fluid absorption. A variety of pathologic processes produce malabsorption syndromes, including intestinal enzyme abnormalities (e.g., lactase deficiency), infection (e.g., AIDS enteritis), and radiation enteritis, among others. The types of malabsorption syndromes discussed here result from a mucosal disorder of the small bowel or from the surgical removal of portions of the stomach or small bowel.

MUCOSAL DISORDERS

Myriad disorders affect the mucosa of the small intestine. Because the small intestine is the principal site of digestion and absorption of nutrients, a defect in this area has the potential for causing malabsorption of fat, protein, carbohydrate, vitamins, and minerals. Crohn disease, described earlier, may result in damage to the mucosa of the distal portion of the ileum, which is the site of vitamin B_{12} and bile acid absorption. Other important mucosal disorders of the small intestine are celiac disease and tropical sprue.

Celiac Disease

Etiology. Celiac disease (also called *celiac sprue*) is characterized by intolerance of gluten, a protein in wheat and wheat products. Current research suggests that celiac sprue is an immune disorder triggered by exposure to gliadin (a specific wheat gluten) in genetically predisposed persons. Environmental, genetic, and immune factors play pivotal roles in determining the nature of symptoms. The main pathologic finding is villus atrophy, with a decrease in the activity and amount of surface epithelial enzymes. The resulting malabsorption of ingested nutrients may promote malnutrition and severe debilitation.

Celiac disease affects twice as many females as males and may have a familial inheritance pattern. More accurate serologic tests have shown the incidence in the general population to be much higher than previously believed; the current prevalence is about 1 in 300, with those affected primarily of northwestern European ancestry. The onset of celiac disease may present in infancy, when gluten-containing products are first introduced into the diet, but is more common in the fourth and fifth decades.

Diagnosis and treatment. In the past the diagnosis of celiac disease relied on intestinal biopsy showing the typical pathologic manifestations. New blood tests that assay for anti–tissue transglutaminase antibody and the more specific immunoglobulin A endomysial antibody have a high degree of sensitivity and specificity, although in general biopsy is still recommended for confirmation. Effective treatment includes the elimination of all gluten from the diet, which results in significant improvement in the intestinal mucosa, and the administration of supplemental iron, folate, and in specific cases vitamin B_{12} and the fat-soluble vitamins (A, D, E, K). Refractory cases may be treated with oral corticosteroids or other immunomodulating agents. Ulcerative jejunitis and lymphoma should be suspected among refractory patients who do not respond to corticosteroids. Importantly, although the overall risk for malignancy is no higher than the general population, the incidence of certain types of intestinal malignancy, especially lymphoma, is modestly increased among sprue patients.

Tropical Sprue

Etiology. Tropical sprue or enteropathy is a malabsorptive syndrome of unknown cause prevalent in equatorial countries. Current theory suggests that bacterial overgrowth of the large intestine produces products of fermentation that damage intestinal mucosa, although the exact etiology is obscure. In tropical sprue, the mucosa of the small intestine atrophies, with resulting malabsorption, malnutrition, and B_{12} and folic acid deficiency. Its incidence is high in persons living in or visiting tropical climates, and it appears to affect adults more often than children. Although environmental factors seem preeminent, a genetic component may be present in some cases.

Clinical manifestations and treatment. The atrophy of the small intestinal mucosa may have severe effects. Massive malabsorption may result from failure of the mucosa to produce the enzymes needed for digestion. Manifestations include severe diarrhea with blood-tinged stools, abdominal distention, and steatorrhea (the presence of excess fat in the stool). In the Caribbean, tropical sprue is strongly linked to the presence of enterotoxin-producing coliforms and responds well to broad-spectrum antibiotics. The response of patients from other areas (e.g., India) to treatment is less predictable than that of patients with Caribbean sprue. Treatment includes antidiarrheal medication and prolonged antimicrobial therapy, as well as replacement of deficient vitamins, particularly folic acid.

MALABSORPTION DISORDERS AFTER SURGICAL INTERVENTION

Surgical procedures in which a portion of the stomach or small bowel is removed may result in loss of the ability to absorb nutrients properly, either through a loss of appropriate motility patterns or through a loss of the surface area of the small bowel needed for adequate absorption. Two types of disorders of malabsorption may occur after surgical intervention on the stomach or small bowel: dumping syndrome and short-bowel syndrome.

Dumping Syndrome

Etiology. *Dumping syndrome* is a term used to describe the literal dumping of stomach contents into the proximal portion of the small intestine because of impaired gastric emptying (Fig. 36.13). This loss of normal, gradual pyloric emptying may occur after removal of all or part of the stomach (gastrectomy), a procedure performed commonly for PUD in previous years, but more recently primarily for control of

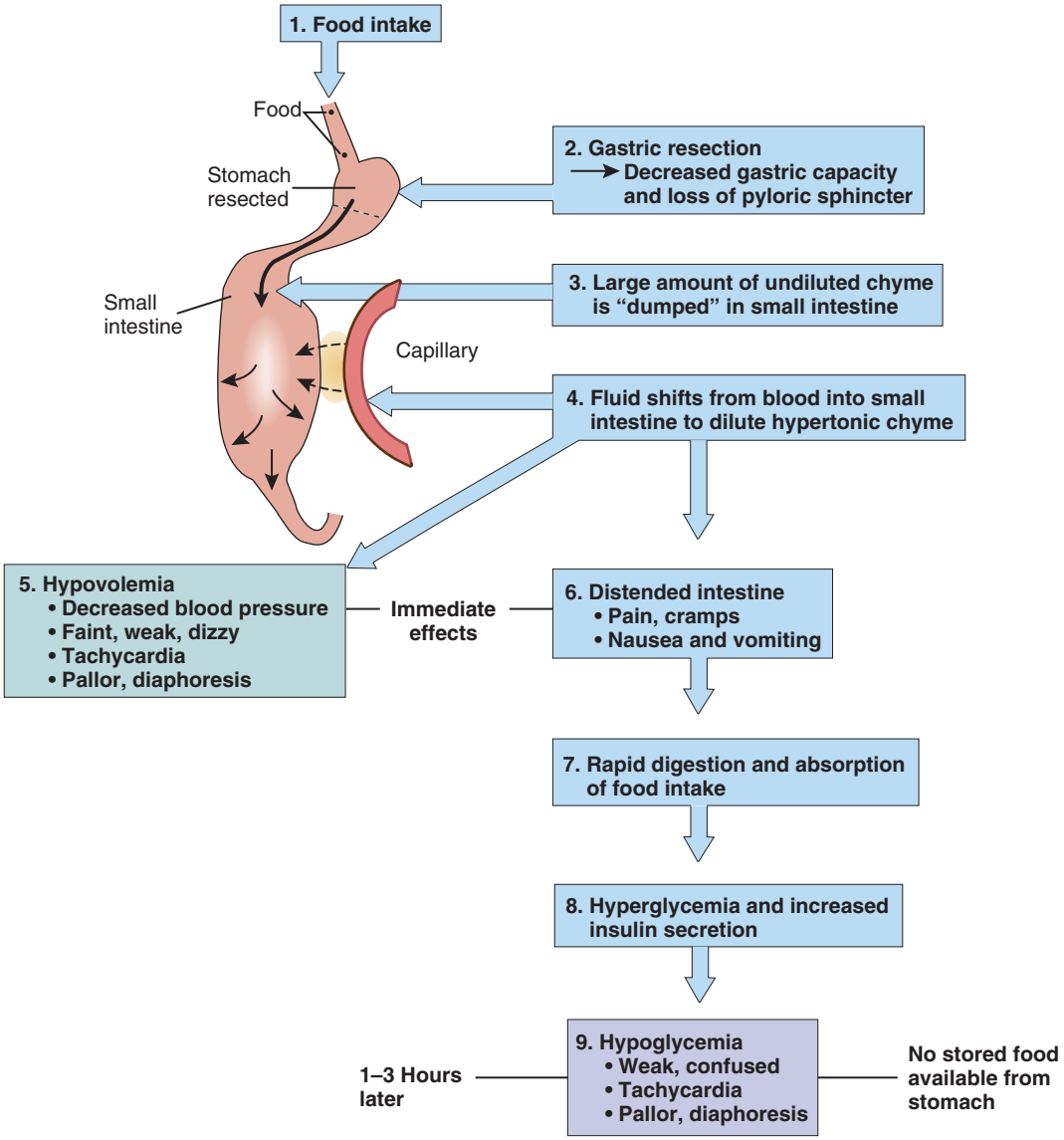


FIG 36.13 Dumping syndrome (postgastrectomy). (From Gould BE: *Pathophysiology for the health professions*, ed 6, Philadelphia, 2018, Saunders.)

obesity. Interestingly, dumping seems to occur only with Roux-en-Y gastric bypass procedures.

Pathogenesis and clinical manifestations. With the normal reservoir function of the stomach now impaired, a large volume of hyperosmolar food is dumped rapidly into the small intestine, with consequences that may be severe. The hyperosmolar contents of the small intestine draw water into the lumen and stimulate bowel motility, with manifestations of diarrhea and abdominal pain. In addition, the rapid absorption of a large amount of glucose and a subsequent rise in blood glucose levels promote an excessive rise in plasma insulin level. The elevated insulin level then causes a rapid fall in blood glucose levels 1 to 3 hours after a meal. This sudden reversal is referred to as *rebound hypoglycemia.*

Treatment. Persons who have undergone a gastrectomy procedure will require specific instruction regarding eating small meals six to eight times a day rather than three large meals. Restriction of carbohydrate intake may be needed to limit glucose absorption. Medications to reduce bowel motility have been helpful in promoting a more normal pattern of bowel function in this population.

Short-Bowel Syndrome

Etiology. Short-bowel syndrome refers to the severe diarrhea and significant malabsorption that develop after the surgical removal of large portions of the small intestine.

Pathogenesis and clinical manifestations. The severity of the manifestations depends on the amount and location of the bowel resected. In particular, removal of the distal two-thirds of the ileum and the ileocecal valve may result in severe malabsorption. Because the ileocecal valve serves to regulate the transit time of intestinal contents, its removal may promote a transit time that is too rapid for adequate absorption of nutrients. In addition, loss of large portions of the small intestine will result in a diminished ability to absorb water, electrolytes, protein, fat, carbohydrates, vitamins, and trace elements. Removal of the terminal ileum will adversely affect vitamin B_{12} malabsorption and will likely necessitate intramuscular replacement.

Treatment. The small intestine displays an amazing ability to adapt after bowel resection. The remaining villi may enlarge and lengthen, thus increasing the absorptive surface area of the bowel. The presence

of orally ingested nutrients is needed for this adaptive process to occur, and a gradual increase in oral intake after bowel resection may promote gradual improvement in absorptive ability. Intravenous nutritional support may be required temporarily or indefinitely after surgical foreshortening of the gut.

KEY POINTS

- Malabsorption occurs when the small bowel fails to absorb one or more dietary components. Diarrhea and abdominal discomfort are the usual manifestations. Malabsorption may occur because of mucosal dysfunction (Crohn disease, celiac disease, tropical sprue), enzyme deficiencies, or surgical alterations that affect transit time and absorptive surface area.
- Celiac disease appears to be caused by a familial intolerance of gluten-containing foods. Ingestion of gluten leads to inflammation and atrophy of the intestinal villi. A reduced surface area and a decreased number of brush border enzymes impair nutrient absorption.
- Dumping syndrome occurs with loss of pyloric sphincter regulation, generally after gastric surgery for ulcers or cancer. Rapid dumping of chyme into the duodenum causes an osmotic shift of water into the lumen and diarrhea. Glucose absorption may be rapid and lead to overshoot of insulin secretion and rebound hypoglycemia.
- Short-bowel syndrome follows surgical procedures involving removal of large sections of the small intestine. Rapid transit time and reduced surface area for absorption lead to diarrhea and malabsorption.

NEOPLASMS OF THE GASTROINTESTINAL TRACT

Neoplasms may develop in any region of the GI tract. They vary in their severity and in their ability to disrupt normal GI functioning. The most common neoplastic processes of the GI tract are summarized here; the reader may wish to refer to Chapter 7 as a background for understanding these specific types of neoplasms occurring in the GI tract.

ESOPHAGEAL, GASTRIC, AND SMALL INTESTINAL CANCERS

Esophageal Cancer

Etiology. Esophageal cancer accounts for 1% to 2% of all cancers (about 17,000 cases annually in the United States) and affects men three times more often than women. It usually develops in men older than 60 years, with an overall 5-year survival rate of 18%. Although the cause of esophageal cancer is presently unknown, several predisposing factors have been identified, including genetic predisposition, dietary habits (especially ingestion of foods high in nitrosamine content), environmental exposures, and chronic irritation of the esophagus. It has been estimated that smoking, alcohol consumption, and diets low in fruits and vegetables account for 90% of squamous cell esophageal cancers in the United States. Chronic severe reflux, especially that associated with achalasia, is a prominent risk factor for adenocarcinoma as well.

Most esophageal tumors worldwide are squamous cell carcinomas. The incidence of adenocarcinoma of the gastroesophageal junction has been increasing steadily in the United States, though, and is now more common than squamous cell cancer. This shift seems to reflect an increased prevalence of Barrett esophagus (discussed earlier in the chapter). As noted earlier, infection with *H. pylori* may actually protect against development of this form of esophageal cancer.

Pathogenesis and treatment. Tumors of the esophagus are usually insidiously infiltrating, and the disease may spread extensively to surrounding organs by way of the esophageal lymphatics at an early stage. Invasion of surrounding structures may lead to the formation of esophagobronchial or esophagopleural fistulas, with subsequent pneumonia or abscess. The tumor may partially constrict the lumen of the esophagus, and surgery, radiation therapy, or other measures may be considered to maintain a patent esophagus. Endoscopic procedures of benefit include stent placement and ablation of the tumor through heat probe and laser techniques. If the individual survives the initial extension of the tumor, the liver and lungs are the usual sites of distant metastasis. Regardless of the cell type, the prognosis is generally poor. Selected patients may be candidates for surgical resection, with or without adjuvant chemotherapy and radiation treatment, with a very high rate of complications.

Gastric Carcinoma

Gastric carcinoma is common throughout the world; however, certain population groups appear to be at higher risk than others. In Japan, the prevalence of gastric adenocarcinoma is about 10 times the prevalence in the United States. The incidence is higher in men older than 30 years than in other age and gender groups. This disease strikes about 25,000 Americans a year, with a male-to-female ratio of approximately 2 to 1. The overall 5-year survival rate is 28%, although the prognosis depends on the stage of the disease at the time of diagnosis. Early-stage gastric cancer has not penetrated the major muscle layer of the stomach wall and is associated with a more favorable survival rate than that seen in more advanced disease.

Etiology. The etiology of gastric cancer is a rapidly expanding area of research. A recent consensus committee of the World Health Organization affirmed the role that *H. pylori* plays in development of gastric cancer, with a twofold increased risk for infected individuals compared with their uninfected peers. In particular, the development of multifocal atrophic gastritis induced by persistent *H. pylori* infection is a critical step in the development of gastric cancer. Epstein–Barr virus is another viral oncogene that has been identified as contributing to the global burden of gastric cancer. Other risk factors are similar to those for esophageal cancer, with the exception that alcohol is not a significant contributor to gastric carcinoma. Aspirin use seems to be protective against stomach cancer. Small numbers of gastric neoplasms may have different histologic characteristics, including lymphoma and carcinoid tumors, and have distinct clinical courses.

Clinical manifestations and treatment. Unfortunately, early gastric cancer typically has no manifestations and is rarely identified in countries that do not have a widespread screening program. Gastric carcinoma extends rapidly to the regional lymph nodes and surrounding organs by way of the lymphatic system and the bloodstream and by direct extension through the wall of the stomach. Advanced gastric cancer (Fig. 36.14) has penetrated the muscle layer of the stomach and produces manifestations such as anorexia, weight loss, and GI bleeding.

Surgical resection of the tumor with appropriate surrounding margins remains the only effective treatment for this cancer, which, like esophageal cancer, has a poor long-term survival.

Small Intestinal Neoplasms

Neoplasms of the small intestine may be benign or malignant. Fairly unusual, they account for fewer than 5% of GI tumors. Tumors of the small intestine (usually adenocarcinomas) occur most often in persons older than 50 years. Carcinoid tumors, lymphoma, and sarcoma represent less common forms of intestinal tumors. Depending on the extent and type of tumor, partial or complete obstruction of the small bowel may occur.

Clinical manifestations and treatment. If the tumor is located near the ampulla of Vater, the common bile duct may become obstructed,

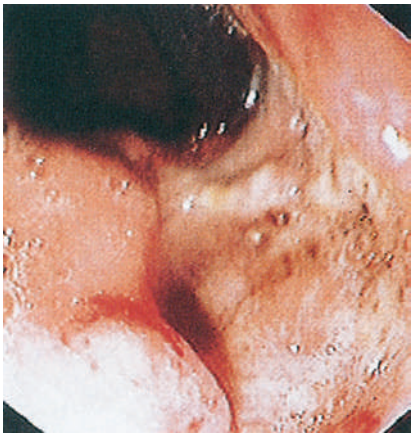


FIG 36.14 Ulcerating gastric cancer. (From Sleisenger MH, Fordtran JS, editors: *Gastrointestinal disease,* ed 5, Philadelphia, 1993, Saunders.)

with resulting biliary stasis and jaundice. Bleeding and ulceration of small intestinal tumors are common manifestations, as is obstruction and, less commonly, intussusception. Treatment may include surgical intervention to remove the tumor and the affected portion of the small intestine, as well as chemotherapy for lymphomas and carcinoid tumors.

COLONIC POLYPS AND COLON CANCER

Cancer of the colon and rectum is identified in approximately 140,000 men and women in the United States each year, with the incidence equally distributed between men and women. It is second only to lung cancer as a cause of cancer deaths. It is well accepted that adenomatous colon polyps represent the major precursor lesion in the development of colon cancer.

Colon Polyps

The term polyp refers to any protrusion into the lumen of the GI tract. Polyps may be benign or malignant, although most clinicians use the term *polyp* to refer to a benign or not-yet-malignant lesion. Polyps can have several forms; a sessile polyp is a raised protuberance with a broad base, whereas a pedunculated polyp is attached to the bowel wall by a stalk that is narrower than the body of the polyp. Benign adenomatous polyps of the colon predispose to malignant adenocarcinoma of the colon through dysplasia and neoplastic degeneration. In fact, some adenomatous polyps may already contain a focus of carcinoma (carcinoma in situ). Most persons with polyps have no manifestations, although polyps may cause occult or gross bleeding and abdominal pain attributable to obstruction. Treatment will vary according to the size and type of polyp and its location in the colon. Biopsy and subsequent removal of polyps may be performed during sigmoidoscopy or colonoscopy. Fig. 36.15 shows various colonic polyps. Several benign colon polyps, including the common hyperplastic variety, may be found on endoscopic biopsy and carry no cancer risk.

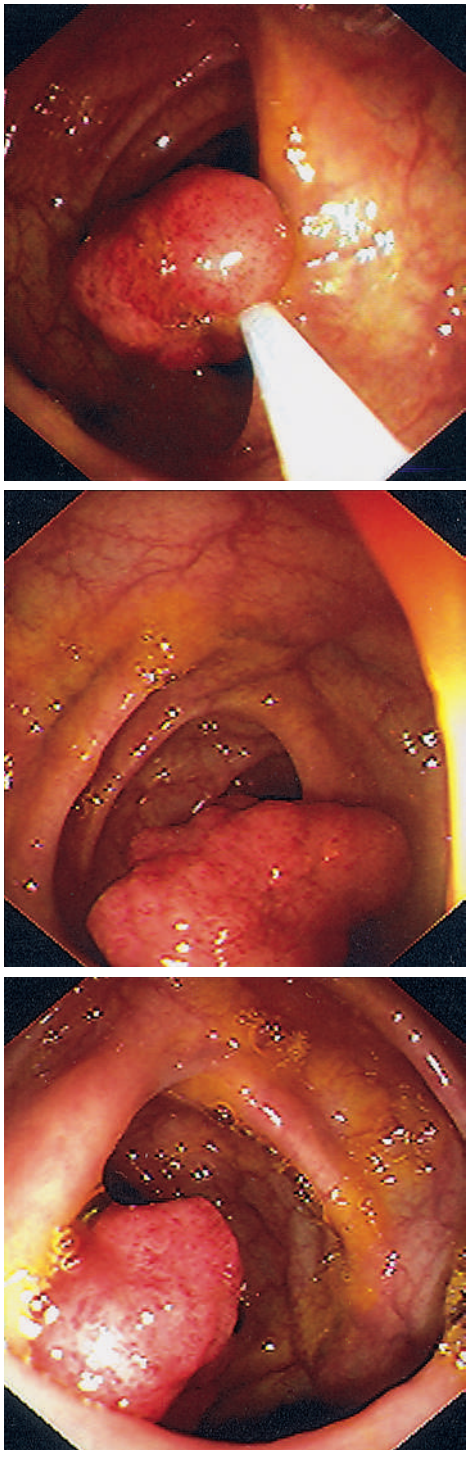


FIG 36.15 Photographs of colonic polyps. (Courtesy LE Copstead.)

Colon Cancer

Etiology and risk factors. A number of risk factors have been identified for the development of colon cancer. The risk increases with advancing age. After age 40 the annual incidence of colon cancer accelerates, doubling every decade until age 80. Dietary factors also seem to increase the risk; a high-fat, low-fiber diet has been proposed as an associated factor. Certain bowel conditions may predispose an individual to colon cancer, including UC and Crohn disease of the colon (see previous discussions). A hereditary predisposition may also be present; the probability of colorectal cancer in a person who has a first-degree relative with the disease is greater than 15%, compared with a 5% risk in the general population.

An important hereditary condition is familial adenomatous polyposis, one form of which is Gardner syndrome. More common are the so-called *hereditary nonpolyposis colorectal cancer syndromes.* Clues to a familial syndrome include at least three close relatives with colorectal cancer, colorectal cancer involving at least two generations, and one or more cases of colorectal cancer occurring before age 50 years. Colon cancer

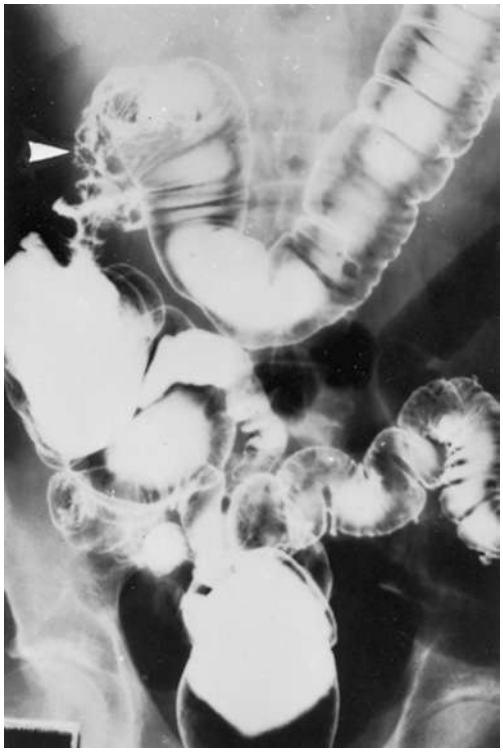


FIG 36.16 Barium enema demonstrating extensive mucosal destruction from a primary lymphoma of the right colon *(arrowhead).* (From Sleisenger MH, Fordtran JS, editors: *Gastrointestinal disease,* ed 5, Philadelphia, 1993, Saunders, p 1484.)

screening guidelines have recently been updated; for the individual at average risk, colonoscopy every 10 years is recommended beginning at age 50. Persons with an increased risk for colon cancer based on family history or IBD should undergo more aggressive colon cancer screening.

Clinical manifestations. The manifestations of colon cancer depend on the anatomic location and function of the bowel segment containing the tumor. On the right side of the colon, the site of water and electrolyte absorption, tumor growth tends to extend along the bowel rather than surround the lumen (Fig. 36.16). Although no signs of obstruction are present, black, tarry stools, which signify bleeding into the intestinal lumen, are a significant finding. On the left side of the colon, a tumor may cause manifestations of obstruction in the early stages of its growth. Feelings of intermittent abdominal cramping and fullness may be present, and "ribbon" or pencil-shaped stools may occur. Typically, the individual may note that the passage of stool or flatus relieves the abdominal pain. As tumor growth progresses, blood or mucus may be present in the stool. When the tumor is located in the rectum, early manifestations may include a change in bowel habits, often beginning with an urgent need to defecate upon awakening in the morning or alternating constipation and diarrhea. Later in the progression of tumor growth, a sensation of rectal fullness and a dull ache may be felt in the rectum or sacral region.

Treatment. The treatment and prognosis for colon cancer depend on several factors, including the extent of tumor invasion through the colon wall, cell type and degree of dysplasia, tumor genetics, and the presence or absence of local or distant tumor metastases. The 5-year survival rate is directly related to the extent of tissue invasion. The traditional Dukes classification scheme (Table 36.1) has been modified and expanded over the last several decades; currently the tumor-node-metastasis (TNM) classification is accepted as the standard scheme (see Chapter 7). The most effective treatment is surgery to remove the malignant tumor and adjacent tissue and lymph nodes that may contain cancer cells. The surgical formation of a colostomy, or an artificial opening of the colon on the abdominal wall, may be performed after removal of the affected bowel segment. Chemotherapy and radiation therapy are used as supportive measures in addition to surgical intervention. Chemotherapy in particular has advanced rapidly over the last 2 decades and has significantly improved the prognosis for moderately advanced (i.e., with nodal metastases) colon cancer. Thus about 90% of patients with localized cancer survive 5 years and 70% of those with regional spread survive this long. Patients with distal metastases have a 14% long-term survival.

TABLE 36.1 Modified Dukes Classification for Colorectal Cancer

Dukes Category	Definition	5-Yr Survival (%) After Treatment
A	Cancer limited to mucosa or submucosa	90
B1	Cancer penetrates into but not through muscularis propria	80
B2	Cancer penetrates through muscularis	70
C1	Same as B1, plus lymph node metastases	50
C2	Same as B2, plus lymph node metastases	50
D	Distant metastases are present	<30

KEY POINTS

- Warning signs for cancer of the GI tract include black, tarry, bloody, or pencil-shaped stools and a change in bowel habits. Risk factors for GI cancer include a low-fiber, high-fat diet; polyps; and chronic irritation or inflammation.
- The prognosis for GI cancer is related to the extent of spread in the body. Surgical removal of tumors followed by chemotherapy, radiation therapy, or both is the usual treatment. Early detection is associated with a better prognosis.

PSYCHOSOCIAL ASPECTS OF GASTROINTESTINAL DISORDERS

Stress of Lifestyle Changes

GI disorders may have profound effects on the psychosocial functioning of the affected individual. Moreover, these disorders may place great stress on the family attempting to cope with the demands of that person's illness. Nutrition and bowel elimination are behaviors that are dependent on cultural norms; changes in these basic areas of human activity caused by a GI disorder may have a variety of meanings to different individuals. A teenager affected by a chronic GI disorder such as Crohn disease may be unable to participate in social activities that revolve around eating and thus may feel isolated from peers. IBS in a young adult who is beginning the most productive years of life may curtail the ability to function fully in the roles of spouse, parent, and wage earner. Finally, the onset of GI disorders, particularly a neoplastic process, in a middle-aged or older individual may not only limit that person's ability to perform activities of daily living, but also may cause depressive symptoms in the face of aging and mortality.

In the past, much of the health care literature, including nursing texts, has tended to stereotype individuals with chronic disorders of the GI tract as having behavioral disorders. Aberrant psychological characteristics were purported to be associated with or even responsible for certain diseases of the GI tract, such as Crohn disease and UC (e.g., the so-called *Crohn personality*). It is now recognized that the pathogenic process is almost never the result of primarily psychological causes, although some chronic diseases of the GI tract may be aggravated by emotional factors. The stress of coping with a chronic, disabling illness may result in psychological trauma; in addition, any type of illness represents a threat to the integrity of the person. Individuals experiencing a chronic GI disorder may exhibit the psychological effects of such threats and will benefit from a sensitive approach to meeting their needs.

SUMMARY

This chapter has described the major alterations in the GI tract that may occur across the human life span. Because of the strong links between cultural and psychological functioning and activities associated with the GI tract, an in-depth understanding of these alterations is essential for health care professionals.

Disorders of the GI tract may have many manifestations, including dysphagia, pain, vomiting, gas, and alterations in bowel elimination patterns. Disorders may occur in any portion of the GI tract, from the mouth to the anus, and may be the result of alterations in the integrity of the GI tract wall (as in UC) or alterations in motility (as in IBS). Disorders of malabsorption, such as celiac disease, may seriously limit the individual's ability to utilize dietary nutrients and are therefore potentially life threatening. Patients who have undergone surgery on the GI tract may also be at risk for malabsorption. Neoplasms of the GI tract are prevalent in the U.S. population, and the reader will want to review the associated risk factors for these neoplasms very carefully. Finally, readers anticipating a career in health care should carefully consider the psychosocial aspects of GI disorders and identify ways to provide optimal care for patients with these conditions.

RESOURCES

Manifestations of GI Disorders

Avunduk C: *Manual of gastroenterology*, ed 4, Philadelphia, 2008, Lippincott Williams & Wilkins.

Feldman M, Friedman LS, Brandt LJ: *Sleisenger and Fordtran's gastrointestinal and liver disease*, ed 10, Philadelphia, 2015, Saunders Elsevier.

Kneepkens CM, Hoekstra JH: Chronic nonspecific diarrhea of childhood: pathophysiology and management. *Pediatr Clin North Am* 43(2):375–390, 1996.

Podolski DK, et al, editors: *Yamada's textbook of gastroenterology*, ed 6, Oxford, 2015, Wiley-Blackwell.

Disorders of Mouth and Esophagus

Bharucha AE, Gostout CJ, Balm RK: Clinical and endoscopic risk factors in the Mallory-Weiss syndrome. *Am J Gastroenterol* 92(5):805–808, 1997.

Bruce A, Rogers RS: New and old therapeutics for oral ulcerations. *Arch Dermatol* 143(4):519–523, 2007.

Coron E, Hatlebakk JG, Galmiche JP: Medical therapy of gastroesophageal reflux disease. *Curr Opin Gastroenterol* 23(4):434–439, 2007.

Hamada H, et al: High incidence of reflux oesophagitis after eradication therapy for Helicobacter pylori: impacts of hiatal hernia and corpus gastritis. *Aliment Pharmacol Ther* 14(6):729–739, 2000.

Khalili H, et al: Use of proton pump inhibitors and risk of hip fracture in relation to dietary and lifestyle factors: a prospective cohort study. *BMJ* 344:e372, 2012.

Liu JY, et al: Determining an appropriate threshold for referral to surgery for gastroesophageal reflux disease. *Surgery* 133(1):5–12, 2003.

Morales P, Baum AE: Therapeutic alternatives for the Mallory-Weiss tear. *Curr Treat Options Gastroenterol* 6(1):75–83, 2003.

Navarro VJ, Garcia-Tsao G: Variceal hemorrhage. *Crit Care Clin* 11(2): 391–414, 1995.

Sampliner RE: Endoscopic ablative therapy for Barrett's esophagus: current status. *Gastrointest Endosc* 59(1):66, 2004.

Santoro MJ, Chen YK, Collen MJ: Polyethylene glycol electrolyte lavage solution-induced Mallory-Weiss tears. *Am J Gastroenterol* 88(8):1292, 1993.

Scully C, Gorsky M, Lozada-Nur F: The diagnosis and management of recurrent aphthous stomatitis: a consensus approach. *J Am Dent Assoc* 134(2):200–207, 2003.

Spechler SJ: Barrett's esophagus and esophageal adenocarcinoma: pathogenesis, diagnosis, and therapy. *Med Clin North Am* 86(6): 1423–1445, vii, 2002.

Vakil N: Gastroesophageal reflux disease and Helicobacter pylori infection. *Rev Gastroenterol Disord* 3(1):1–7, 2003.

Weston AP: Hiatal hernia with Cameron ulcers and erosions. *Gastrointest Endosc Clin N Am* 6(4):671–679, 1996.

Disorders of Stomach and Intestinal Wall

Atherton JC: CagA, the cag pathogenicity island and Helicobacter pylori virulence. *Gut* 44:307–308, 1999.

Carpenter HA, Talley NJ: Gastroscopy is incomplete without biopsy: clinical relevance of distinguishing gastropathy from gastritis. *Gastroenterology* 108(3):917–924, 1995.

Cohen RD, Stein R, Hanauer SB: Intravenous cyclosporin in ulcerative colitis: a five-year experience. *Am J Gastroenterol* 94(6):1587–1592, 1999.

Cosnes J, et al: Smoking cessation and the course of Crohn's disease: an intervention study. *Gastroenterology* 120:1093–1099, 2001.

Crowe S: Indications and diagnostic tests for Helicobacter pylori infection. In Chopra S, Lamont JT, GI editors: *UpToDate*. Available at www.utdol.com. (Accessed 17 November 2015).

Farrell RJ, Peppercorn MA: Overview of the medical management of severe or refractory Crohn disease in adults. In Chopra S, Lamont JT, GI editors: *UpToDate*. Available at www.utdol.com. (Accessed 17 November 2015).

Hogan DL, et al: Duodenal bicarbonate secretion: eradication of Helicobacter pylori and duodenal structure and function in humans. *Gastroenterology* 110(3):705–716, 1996.

Howden CW, Hunt RH: Guidelines for the management of Helicobacter pylori infection. Ad Hoc Committee on Practice Parameters of the American College of Gastroenterology. *Am J Gastroenterol* 93(12): 2330–2338, 1998.

Hurley BW, Nguyen CC: The spectrum of pseudomembranous enterocolitis and antibiotic-associated diarrhea. *Arch Intern Med* 162(19):2177–2184, 2002.

Kurata JH, Nogawa AN: Meta-analysis of risk factors for peptic ulcers: nonsteroidal anti-inflammatory drugs, Helicobacter pylori, and smoking. *J Clin Gastroenterol* 24:2–17, 1997.

Lewis JD, Deren JJ, Lichtenstein GR: Cancer risk in patients with inflammatory bowel disease. *Gastroenterol Clin North Am* 28(2):459–477, 1999.

Loftus EV, Jr: Clinical epidemiology of inflammatory bowel disease: incidence, prevalence, and environmental influences. *Gastroenterology* 126(6):2004, 1504.

Malaty HM, et al: Transmission of Helicobacter pylori infection: studies in families of healthy individuals. *Scand J Gastroenterol* 26:927–932, 1991.

Marshall B: Unidentified curved bacilli on gastric epithelium in active chronic gastritis. *Lancet* 1:1273–1274, 1983. letter.

Oliak D, et al: Initial nonoperative management for periappendiceal abscess. *Dis Colon Rectum* 44(7):936–941, 2001.

Peppercorn MA, Cheifetz AS: Definition, epidemiology, and risk factors in inflammatory bowel disease. In Chopra S, Lamont JT, GI editors: *UpToDate*. Available at www.utdol.com. (Accessed 17 November 2015).

Peyrin-Biroulet L, Deltenre P, et al: Efficacy and safety of tumor necrosis factor antagonists in Crohn's disease: meta-analysis of placebo-controlled trials. *Clin Gastroenterol Hepatol* 6(6):644–653, 2008.

Podolsky DK: Medical progress: inflammatory bowel disease. *N Engl J Med* 347:417–429, 2002.

Rutgeerts P, et al: Infliximab for induction and maintenance therapy for ulcerative colitis. *N Engl J Med* 353(23):2462–2476, 2005.

Sandborn WJ, et al: Adalimumab induces and maintains clinical remission in patients with moderate-to-severe ulcerative colitis. *Gastroenterology* 146(1):96–109, 2014.

Shiotani A, Graham DY: Pathogenesis and therapy of gastric and duodenal ulcer disease. *Med Clin North Am* 86(6):1447–1466, 2002.

Suerbaum S, Michetti P: Medical progress: Helicobacter pylori infection. *N Engl J Med* 347:1175–1186, 2002.

Walsh JH, Peterson WL: The treatment of Helicobacter pylori infection in the management of peptic ulcer disease. *N Engl J Med* 333:984–991, 1995.

Disorders of Bowel Motility and Absorption

Buettcher M, et al: Three-year surveillance of intussusception in children in Switzerland. *Pediatrics* 120(3):473–480, 2007.

Loening-Baucke V: Encopresis. *Curr Opin Pediatr* 14(5):570–575, 2002.

Parisi MA, Kapur RP: Genetics of Hirschsprung disease. *Curr Opin Pediatr* 12(6):610–617, 2000.

Schuppan D: Current concepts of celiac disease pathogenesis. *Gastroenterology* 119(1):234–242, 2000.

Wald A: Treatment of irritable bowel syndrome in adults. In Chopra S, Lamont JT, GI editors: *UpToDate*. Available at www.utdol.com. (Accessed 17 November 2015).

Cancers of the GI Tract

Askling J, et al: Cancer incidence in a population-based cohort of individuals hospitalized with celiac disease or dermatitis herpetiformis. *Gastroenterology* 123(5):1428–1435, 2002.

Chow WH, et al: An inverse relation between cagA+ strains of Helicobacter pylori infection and risk of esophageal and gastric cardia adenocarcinoma. *Cancer Res* 58:589–590, 1998.

Colorectal cancer: Statistics. In cancer.net. Available at http://www.cancer.net/cancer-types/colorectal-cancer/statistics. (Accessed 17 November 2015).

Engel LS, et al: Population attributable risks of esophageal and gastric cancers. *J Natl Cancer Inst* 95(18):1404–1413, 2003.

Esophageal cancer: Statistics. In cancer.net. Available at http://www.cancer.net/cancer-types/esophageal-cancer/statistics. (Accessed 17 November 2015).

Kelsen DP, et al: *Principles and practice of gastrointestinal oncology*, ed 2, Philadelphia, 2008, Lippincott Williams & Wilkins.

National Comprehensive Cancer Network (NCCN): *Clinical practice guidelines in oncology*. Available at http://www.nccn.org/professionals/physician_gls/f_guidelines.asp. (Accessed 12 November 2015).

Stomach cancer: Statistics. In cancer.net. Available at http://www.cancer.net/cancer-types/stomach-cancer/statistics. (Accessed 17 November 2015).

Takada K: Epstein-Barr virus and gastric carcinoma. *Mol Path* 53(5):255–261, 2000.

Trowbridge B: Colorectal cancer screening. *Surg Clin North Am* 82(5):943–945, 2002.

37

Alterations in Function of the Gallbladder and Exocrine Pancreas

Jeffrey S. Sartin

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- How is bile produced, stored, and secreted?
- How is pancreatic enzyme secretion regulated?
- What factors predispose to formation of cholesterol gallstones?
- What is the relationship between cholecystitis and cholelithiasis?
- What clinical and laboratory findings are indicative of acute pancreatitis?
- What serious complications may result from acute pancreatitis?
- How do the etiologic factors, clinical presentation, and management of chronic pancreatitis differ from those of acute pancreatitis?
- What are the signs, symptoms, and treatment for pancreatic cancer?

CHAPTER OUTLINE

Structure and Function of the Pancreaticobiliary System, 742
Embryology of the Pancreaticobiliary System, 743
Physiology of Bile, 743
Functional Anatomy of the Pancreas, 744
Disorders of the Gallbladder, 744
Pathophysiology of Cholesterol Gallstone Formation, 744
Cholelithiasis and Cholecystitis, 745
Chronic Cholelithiasis, 746
Acute Cholecystitis, 747
Chronic Cholecystitis, 748
Biliary Malignancy, 748
Disorders of the Pancreas, 748
Pancreatitis, 748
Acute Pancreatitis, 748
Chronic Pancreatitis, 750
Pancreatic Cancer, 752

The pancreaticobiliary system does not have the prominence of the stomach or intestine or liver in popular culture, but it is certainly a major component of both health and illness. Stones have been discovered in the gallbladders of Egyptian and Chinese mummies, suggesting that gallbladder disease has been present in humans for thousands of years. More than 700,000 operations to remove the gallbladder (cholecystectomy) are performed annually in the United States, and the incidence of new gallstones is 1 million to 2 million cases per year. The incidence of acute pancreatitis varies according to geographic area, ranging from 5 to 24 per 100,000 population in the United States and Europe. Many health care professionals will encounter diseases of the pancreaticobiliary system frequently and need an in-depth understanding of the mechanisms that promote them. This chapter describes the pathophysiology of diseases of the gallbladder, the biliary tree, and the exocrine pancreas and summarizes current treatments. Cancers of the pancreas and biliary tree will be discussed briefly.

STRUCTURE AND FUNCTION OF THE PANCREATICOBILIARY SYSTEM

The pancreaticobiliary system is composed of the gallbladder and cystic duct; the intrahepatic, hepatic, and common bile ducts; and the endocrine and exocrine pancreas. The extrahepatic biliary tree and the gallbladder form a controlled system for delivering bile to the intestinal tract. (For additional background, refer to Chapter 35, which describes the physiology of digestion in the gastrointestinal lumen.)

The gallbladder is a distensible sack of about 30- to 50-mL capacity that connects via the cystic duct with the common hepatic duct to form the common bile duct. The common bile duct, which is about 3 inches long, extends behind the duodenum to terminate at the ampulla of Vater, a complex structure that also forms the terminating point of the main pancreatic duct. The pancreatic duct travels proximally in the pancreas and branches off dorsally to drain the tail of the pancreas.

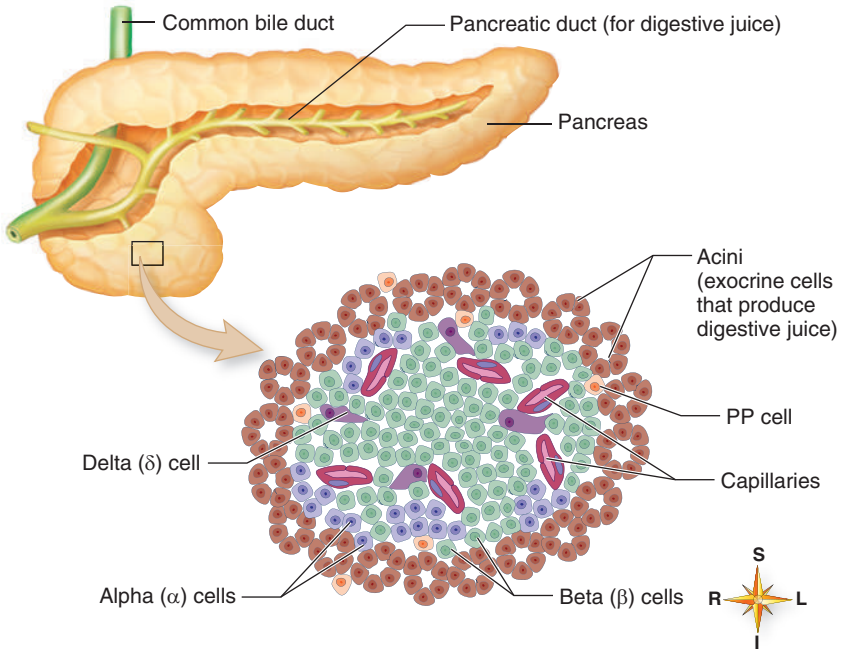


FIG 37.1 Pancreas. A pancreatic islet, or hormone-producing area, is shown among the pancreatic cells that produce the pancreatic digestive juice. The pancreatic islets are more abundant in the tail of the pancreas than in the body or head. (From Patton KT, Thibodeau GA: *Essentials of anatomy & physiology,* St. Louis, 2012, Mosby.)

Thus the ampulla of Vater forms the major aqueduct through which important digestive secretions enter the intestinal tract (Fig. 37.1).

EMBRYOLOGY OF THE PANCREATICOBILIARY SYSTEM

In about the third or fourth week of gestation, the hepatic diverticulum forms from the primitive foregut. It is composed of specialized progenitor liver cells that will eventually develop into the entire liver, biliary tree, and ventral pancreas. The dorsal pancreas forms from a separate outcropping of cells lying on the opposite side of the primitive foregut.

At 5 weeks of gestation, three buds can be seen in the hepatic diverticulum. The cranial bud contains specialized liver cells (hepatoblasts), which will form the liver. The liver sinusoids develop and feed into the developing bile canaliculi, which drain into intralobular ductules and then into interlobular ducts. The caudal bud develops into the gallbladder, which joins the common hepatic duct via the developing cystic duct to form the common bile duct. The gallbladder and the hepatic ducts are initially hollow, but they solidify with development and become solid cords. With further differentiation, they become hollow once again to become the tubes and reservoir for bile flow. During the second trimester, the fetus produces bile, which gives color to the fetal meconium. Finally, the basal bud transforms into the ventral pancreas (Fig. 37.2).

PHYSIOLOGY OF BILE

Normal bile is composed primarily of water, electrolytes, and organic solutes. It has a low protein content, containing mainly bile acids, pigment, cholesterol, and phospholipids. Bile acids consist mostly of primary bile salts (cholic and chenodeoxycholic acids) and secondary bile salts (deoxycholic, ursodeoxycholic, and lithocholic acids). Bile is formed in the liver and then modified and stored in the gallbladder and bile ducts before secretion into the intestinal tract. The major functions of bile are to aid in the digestion of lipids in the diet and to transport waste products (particularly bilirubin), immunoglobulins (IgA), toxins, and cholesterol into the intestine for eventual disposal or reabsorption. After secretion into the bile canaliculi, bile flows through the canals of Hering, ductules, interlobular ducts, septal ducts, right and left lobar ducts, and into the common bile duct (see Fig. 37.1).

The principal function of the gallbladder is the concentration and storage of bile. During the fasting state, the muscular sphincter at the ampulla of Vater is contracted, promoting flow of bile into the gallbladder. Only about half of the bile is stored during this time; the remainder flows into the duodenum. At the same time that bile is moving into the gallbladder, absorption is also occurring so that within 4 hours up to 90% of water in bile can be removed, leaving a very concentrated mixture of sodium, bile salts, and other electrolytes.

With the first morning meal, a hormonally and neurally regulated contraction of the gallbladder occurs, releasing the concentrated bile into the duodenum. Bile acids eventually will be absorbed again in the terminal ileum and travel by the portal circulation to be secreted again into the bile. (Bile acids are reabsorbed on average two or three times daily.) A small amount of the bile acid pool (less than 5%) enters the colon, where primary bile salts undergo bacterial transformation into secondary bile salts (Fig. 37.3).

After secretion into the bile, bile salts have dual properties, being hydrophilic (soluble in water) at one end and hydrophobic (insoluble in water) at the other. Thus these molecules tend to aggregate into clusters called *micelles,* which surround lipids such as cholesterol and allow them to solubilize (Fig. 37.4). Micelles are not good stabilizers of cholesterol alone, but another molecule, lecithin, also secreted in large amounts in the bile, is readily incorporated in the core of the micelle to greatly enhance the solubility of cholesterol. In this way bile keeps cholesterol partly solubilized. Precipitation of cholesterol from bile occurs at high concentrations, predisposing to the formation of gallstones.

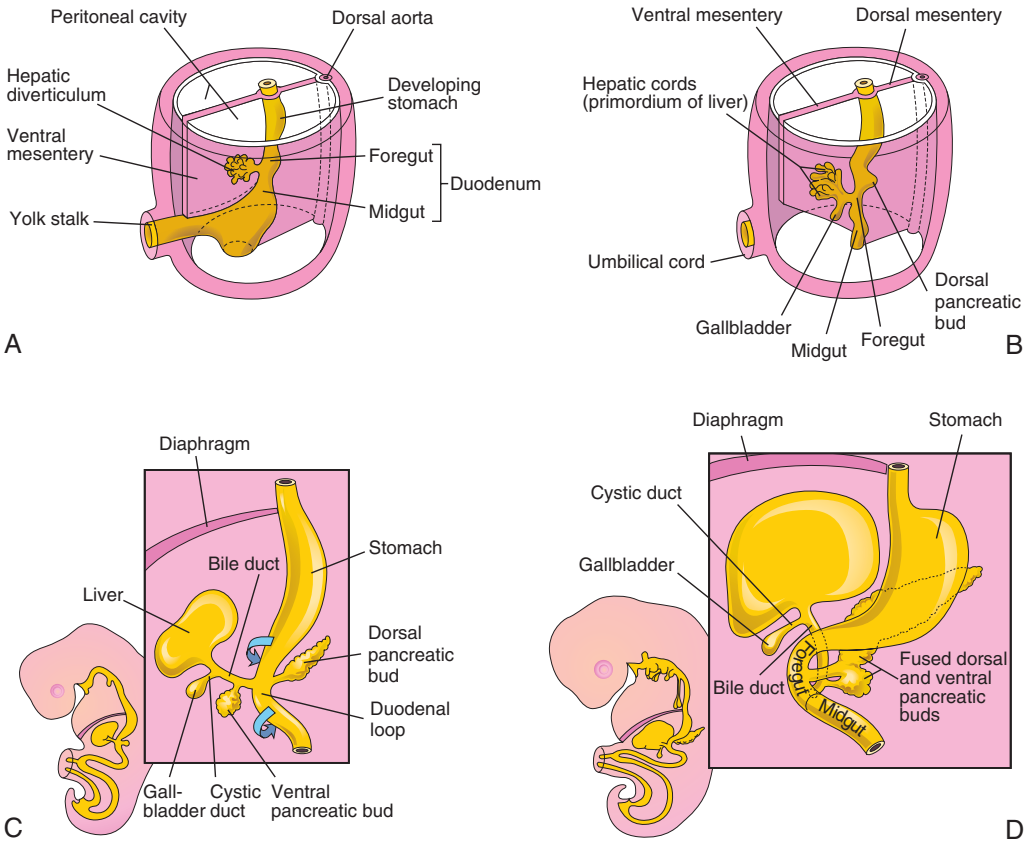


FIG 37.2 Stages in the embryonic development of the liver, pancreas, and duodenum at 4 weeks **(A),** 5 weeks **(B** and **C),** and 6 weeks **(D).** (From Moore KL, Persaud TVN, Torchia MG: *The developing human: clinically oriented embryology,* ed 9, Philadelphia, 2013, Saunders.)

FUNCTIONAL ANATOMY OF THE PANCREAS

The pancreas is really two organs in one: it functions as both an endocrine and an exocrine organ. On the one hand, hormones such as insulin, glucagon, and somatostatin are produced and secreted into the vascular system (characteristic of an endocrine organ). On the other hand, every 24 hours the pancreas secretes more than 1 L of digestive juice into the digestive tract (characteristic of an exocrine organ).

Embryologically, the pancreas is composed of two fused organs: a dorsal and a ventral pancreas (Fig. 37.5). Microscopically, the pancreas is somewhat lobular and arranged into exocrine glands. The pancreatic juices are secreted into the glandular acini, which eventually drain into the main pancreatic duct and then enter the intestinal tract. The juices themselves are composed of both active digestive enzymes (e.g., amylase, lipase) and precursor or proenzymes (e.g., trypsinogen). Their release during a meal is controlled by hormones secreted from the small intestinal mucosa: cholecystokinin (CCK) and secretin. When this regulation is deranged, enzymes may be released within the gland and produce acute pancreatitis, as discussed later.

KEY POINTS

- Bile is produced by hepatocytes in the liver and stored in the gallbladder. The main components of bile are bile acids, pigment, cholesterol, and phospholipids. Bile salts are important for digestion and absorption of fats from the small bowel. Bile is an important route for excretion of waste products, particularly bilirubin. The gallbladder receives bile from the liver, concentrates bile by absorbing water, and then contracts to expel stored bile into the common bile duct, which terminates in the duodenum.
- The pancreas is both an endocrine organ (secreting insulin, glucagon, and somatostatin into the bloodstream) and an exocrine gland (secreting digestive juice into the duodenum). Some pancreatic enzymes are secreted in active form (amylase, lipase), whereas others are proenzymes that are activated in the duodenum (trypsinogen). Release of pancreatic enzymes is stimulated by cholecystokinin and secretin.

DISORDERS OF THE GALLBLADDER

Pathophysiology of Cholesterol Gallstone Formation

The majority of gallstones among patients in the United States are cholesterol stones. In general, the formation of cholesterol stones in the gallbladder (**cholelithiasis**) depends on three factors: (1) supersaturation of bile with cholesterol; (2) nucleation of crystals; and (3) hypomotility, allowing stone growth (Fig. 37.6).

As described previously, cholesterol eventually precipitates from supersaturated bile. If conditions are right, nucleation occurs in which the cholesterol crystals aggregate together. Continued growth of the crystals then depends on the balance between cholesterol growth–promoting factors and factors that tend to cause stone dissolution. A significant factor that promotes the continued growth of stones is hypomotility, or stasis of bile within the gallbladder. Many types of patients are at particular risk for development of cholesterol gallstones, including persons with high spinal cord injuries, patients receiving total parenteral nutrition, and persons who undergo prolonged fasting or

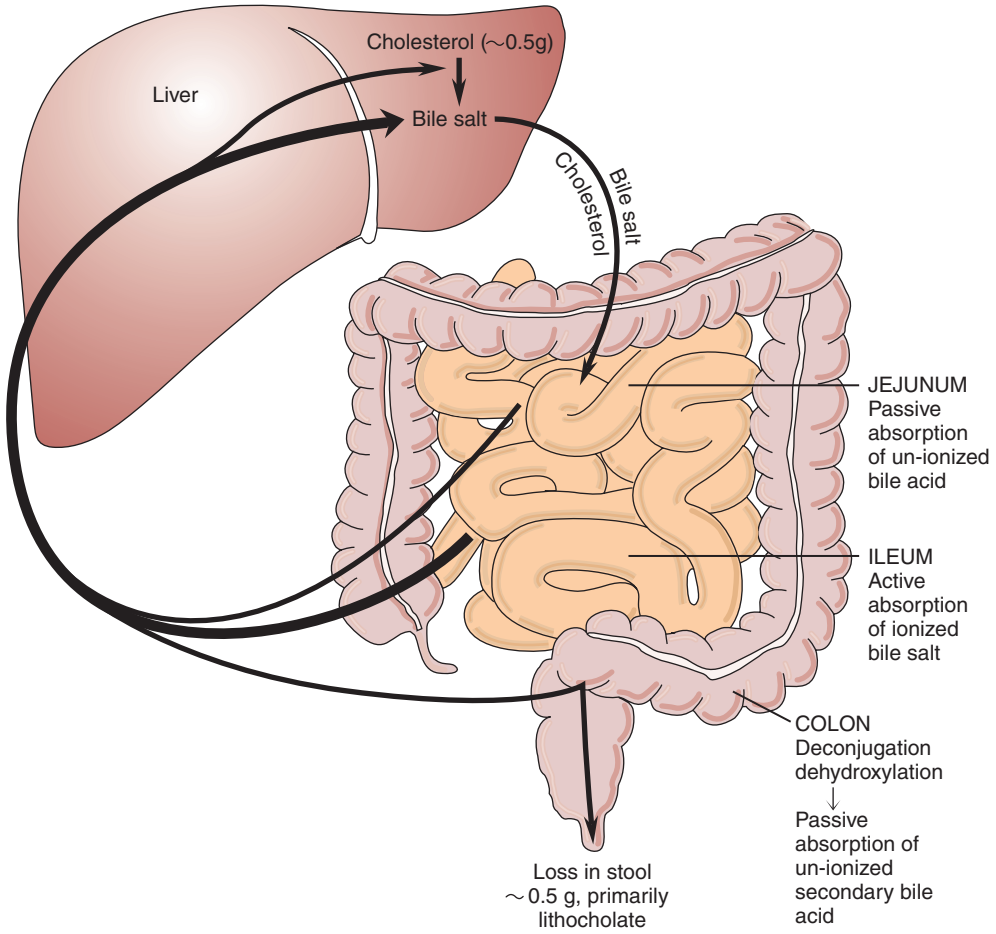


FIG 37.3 Enterohepatic bile salt recirculation is maintained by passive jejunal absorption of un-ionized bile salts, active ileal absorption of ionized bile salts, and colonic deconjugation and dehydroxylation of bile salts followed by passive absorption of lipid-soluble un-ionized secondary bile salt. The loss of unabsorbable bile salt is balanced by the de novo hepatic synthesis of bile salt from cholesterol. (From Cooper AD: Metabolic basis of cholesterol gallstone disease, *Gastroenterol Clin North Am* 1991;20:34.)

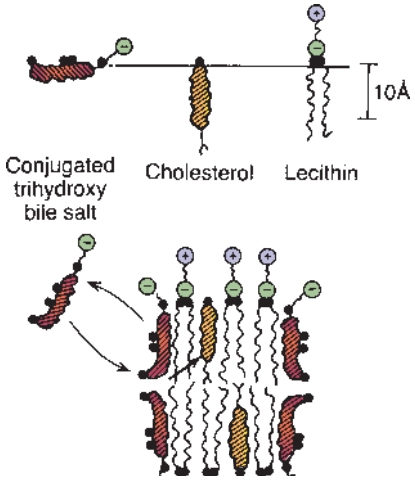


FIG 37.4 Bile acid–lecithin–cholesterol mixed micelle. Polar ends of bile acid and lecithin are oriented outward, whereas hydrophobic, nonpolar portions make up the interior. Cholesterol is solubilized within the hydrophobic, nonpolar center. (From Saunders KD et al: Pathogenesis of gallstones. *Surg Clin North Am* 1990;70:1197–1216.)

rapid weight loss. Other risk factors for cholelithiasis include pregnancy, oral contraceptives, obesity, diabetes mellitus, and octreotide (somatostatin analog) therapy.

About 25% of gallstones in Western countries are due to pigment stones, which contain a mixture of pigment polymers and calcium salts. "Black" pigment stones are most common and may be idiopathic or associated with cirrhosis or hemolysis. "Brown" pigment stones differ in their composition and are much more common in developing countries, where they are associated with biliary parasitosis and bacterial colonization and infection.

Cholelithiasis and Cholecystitis

Currently, about 20 million people in the United States have gallstones (cholelithiasis). The incidence of gallstones is related to age, gender, and a variety of medical factors. Gallstones are twice as common in women as in men. Native Americans, particularly the Pima Indians of North America, are markedly susceptible to gallstones, and American Caucasians somewhat less so. European Caucasians are intermediate in prevalence, and persons of Asian descent are at the lowest risk. Obesity correlates with the development of gallstones, but so does rapid weight loss in an obese individual. However, the presence of gallstones does not necessarily mean that an individual will have any symptoms. In many individuals, gallstones are completely asymptomatic

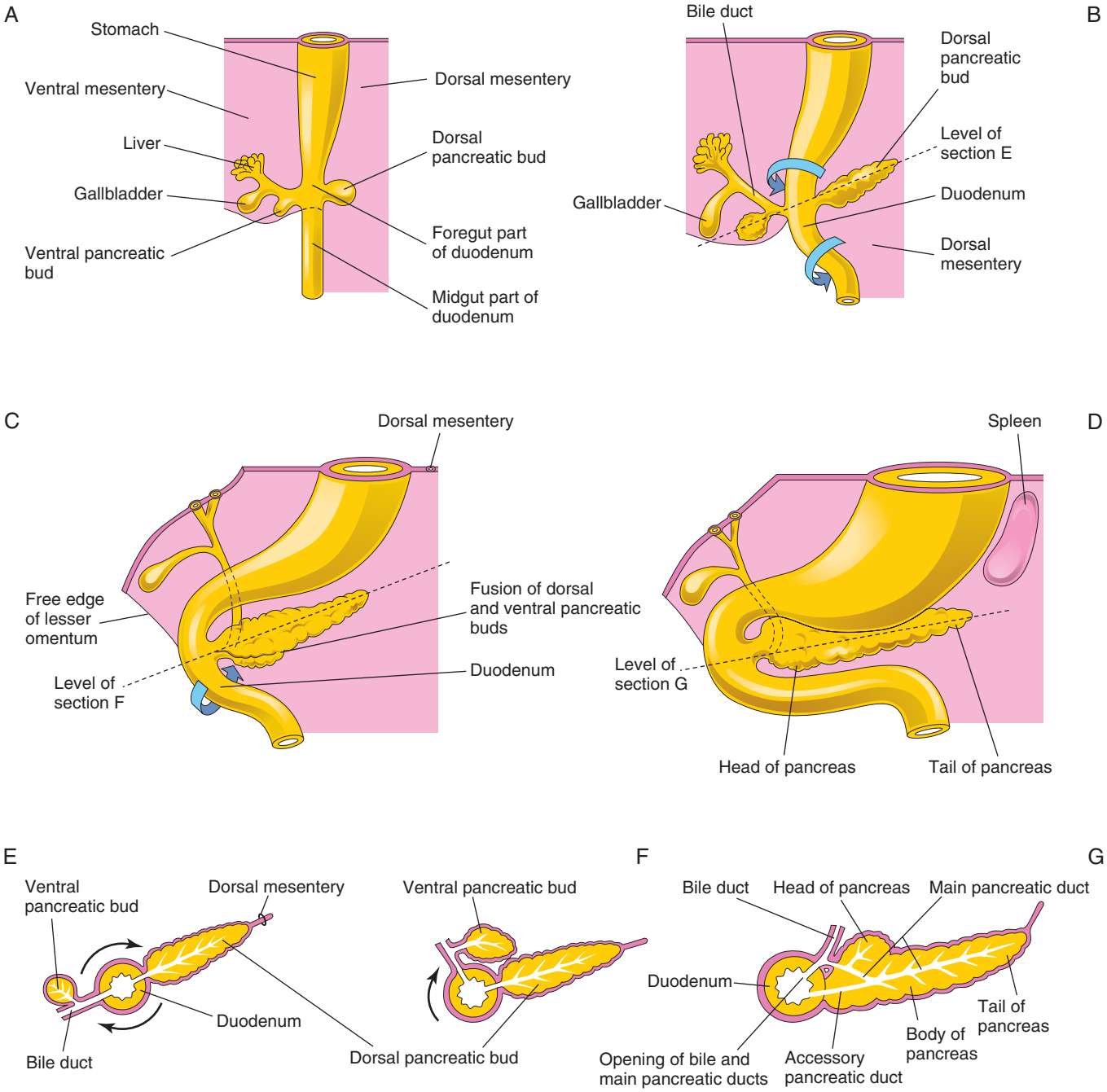


FIG 37.5 Stages in the embryonic development of the pancreas from the fifth to eighth week **(A–D)** and a diagram of the progressive development of the bile duct and main pancreatic duct **(E–G).** (From Moore KL, Persaud TVN, Torchia MG: *The developing human: clinically oriented embryology,* ed 9, Philadelphia, 2013, Saunders.)

and require no treatment. Cholelithiasis in children is usually associated with an underlying systemic disease such as cystic fibrosis or sickle cell disease. It is often symptomatic and requires treatment by gallbladder removal.

The term *cholecystitis* refers to inflammation of the gallbladder wall. The continued presence of gallstones in the gallbladder ultimately promotes inflammatory changes in the gallbladder wall, with fibrosis and thickening (Fig. 37.7). Cholecystitis is classified as acute or chronic, according to its clinical manifestations. The clinical manifestations of each are described, followed by diagnostic and treatment methods.

Chronic Cholelithiasis

Clinical manifestations. The chief complication of chronic cholelithiasis is intermittent biliary colic, a persistent epigastric or right upper abdominal pain. Often the pain radiates to the back and is accompanied by nausea, vomiting, sweating, and flatus. A typical episode lasts several hours. Biliary colic is typically caused by periodic obstruction of the cystic duct by a gallstone, although rarely spasm of the sphincter of Oddi may be etiologic.

Symptoms may be precipitated by a meal, but they often occur spontaneously and may manifest at night. The pain often increases

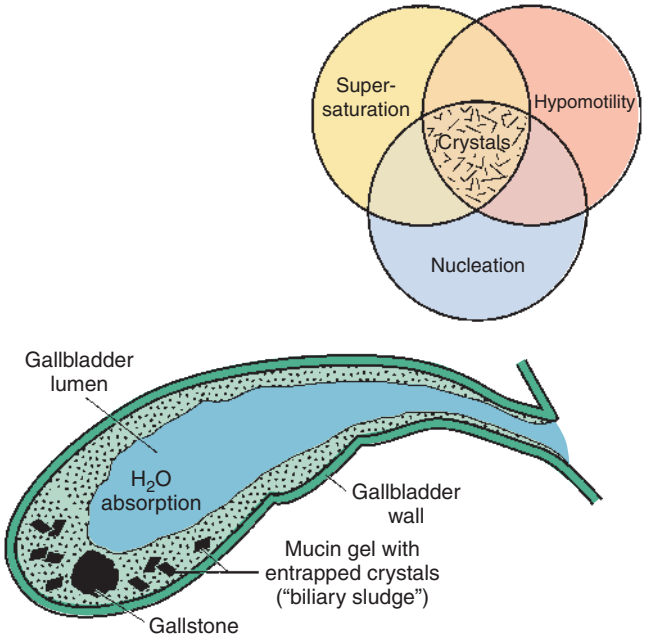


FIG 37.6 Three principal phases responsible for the formation of cholesterol gallstones illustrated with a Venn diagram. Absorption of water concentrates the bile resulting in sludging and increasing the risk of crystal formation and stones. (From Feldman M et al: *Sleisenger and Fordtran's gastrointestinal and liver disease: pathophysiology, diagnosis, management,* ed 10, Philadelphia, 2015, Saunders.)

FIG 37.7 Chronic cholecystitis demonstrated by a thickened gallbladder wall and luminal cholesterol gallstones. (From Cotran RS et al: *Robbins pathologic basis of disease,* ed 6, Philadelphia, 1999, Saunders, p 897.)

steadily for 15 minutes and persists for 1 hour or more and then slowly decreases. Attacks may recur on a frequent or infrequent schedule. Patients often describe additional symptoms, including fatty food intolerance, belching, flatus, bloating, and epigastric burning.

Diagnosis and treatment. The diagnosis of chronic cholelithiasis depends on imaging studies, particularly ultrasonography, and is essentially the same as that described for acute cholecystitis.

In cases of chronic cholelithiasis with no or very few recurrent symptoms, "watchful waiting" may be considered. However, patients with significant recurrences of biliary colic are at risk for potentially life-threatening complications and would be candidates for one of the three current modes of treatment: cholecystectomy (surgical removal of the gallbladder), chemical dissolution of gallstones, or lithotripsy (mechanical breaking up of gallstones within the gallbladder). These will be discussed in more detail in the Acute Cholecystitis section.

Acute Cholecystitis

Clinical manifestations. Acute cholecystitis is defined as acute inflammation of the gallbladder wall. It is characterized by severe right upper abdominal pain that often radiates to the back. Abdominal tenderness and fever are often present. Cholelithiasis is present in about 90% of patients. Obstruction of the cystic duct occurs in almost all cases, suggesting that stasis of bile in the gallbladder is important in the pathogenesis of the disease. Bacterial infection commonly accompanies acute cholecystitis, although it is not thought to be the direct cause of the inflammatory process. Laboratory evaluation may reveal leukocytosis, mild elevations in the levels of bilirubin and serum transaminases, and, less often, elevated amylase levels.

If left untreated, the inflammatory process often escalates, and gangrene of the gallbladder wall with rupture may occur. This can lead to peritonitis and septic shock, a localized abscess (empyema), or a cholecystoenteric fistula (fistula between the gallbladder and gastrointestinal tract).

Acalculous cholecystitis is an important subgroup of acute cholecystitis. As the name implies, it occurs in patients without preexisting gallstones. It tends to occur in the setting of major surgery, critical illness, trauma, or burn-related injury. Patients tend to be predominantly male and older than 50 years, and total parenteral nutrition (TPN) is a common cofactor. It carries a somewhat more serious implication than stone-associated cholecystitis, in part because of the associated medical conditions. In fact, gangrene of the gallbladder wall and perforation, emphysematous cholecystitis, and empyema all develop more rapidly than in calculous cholecystitis.

Diagnosis. Evaluation for possible cholecystitis includes an appropriate history and physical examination, laboratory studies, and imaging studies designed to evaluate the gallbladder and biliary tree. Ultrasound of the abdomen is the procedure of choice early in the diagnostic evaluation. Typically, the ultrasound scan reveals the presence of stones, occasionally in the cystic duct, as well as thickening of the gallbladder wall and distention of the lumen. It may also point to another diagnosis of right upper quadrant pain, such as a liver neoplasm or renal lesions. The sensitivity and specificity of ultrasound for stones larger than 2 mm in diameter approach 95%. Sensitivity rates for acalculous cholecystitis are somewhat lower.

However, ultrasound may not be diagnostic, and other diagnostic tests are used on occasion. Hepatobiliary nuclear scintigraphy by hydroxyl-iminodiacetic acid scanning provides a good functional assessment of gallbladder excretion, which is markedly impaired with cholecystitis. Computed tomography (CT), magnetic resonance cholangiography, and endoscopic retrograde cholangiopancreatography (ERCP) are useful for selected cases. The latter carries a risk of

perforation and pancreatitis and so is usually reserved for cases requiring biopsy or intervention (stent placement, biopsy, or special contrast studies).

Treatment. Treatment for cholecystitis depends on the severity of symptoms and the patient's clinical status. Acute cholecystitis may necessitate urgent intervention, but surgeons generally prefer to allow a "hot" gallbladder to "cool down" before performing surgery, particularly if the patient is considered a high surgical risk. The benefit of antibiotics for uncomplicated cholecystitis remains unclear, but broad-spectrum antibiotics are generally administered in the setting of acute cholecystitis. Percutaneous catheter drainage or endoscopic drainage with stent placement via ERCP may be performed to relieve obstruction, particularly if infection is involved. If the patient's condition precludes surgery, these may be the main treatments, and drains can be left in place indefinitely. Advanced acute cholecystitis complicated by empyema, gangrene, or emphysematous change is considered a surgical emergency.

Laparoscopic cholecystectomy was first performed in 1987 and has benefited from rapid improvements in video laparoscopy and instrumentation. Laparoscopic cholecystectomy is now the treatment of choice for symptomatic gallstones. The procedure is usually performed with four small incisions through which instruments are inserted. The gallbladder is freed either by electrosurgical or laser excision and is then withdrawn through one of the small incisions. Advantages of the laparoscopic technique include minimal scarring, less postoperative pain than after laparotomy, shortened hospital stays, and a rapid return to daily activities.

Open cholecystectomy was first performed in the nineteenth century and remains an extremely safe operation with low morbidity and mortality. However, the length of the incision and accompanying postoperative pain render many patients immobile after the procedure. Altered anatomy or scarring from previous surgery or the presence of common bile duct stones may necessitate a traditional cholecystectomy. A laparoscopic procedure may be converted to an open one intraoperatively if necessary (e.g., upon finding a malignant tumor in the vicinity of the gallbladder). Complications of cholecystectomy (either laparoscopic or open) can include infection, inadvertent transection of the common bile duct, and the rare but debilitating syndrome of recurrent sclerosing cholangitis.

Some patients, such as elderly or debilitated persons, may be poor surgical risks and cannot undergo the stress of surgery. Nonoperative methods to manage gallstones, such as chemodissolution with a variety of bile acids or organic solvents and lithotripsy, have been tried as alternatives to surgery. Chemodissolution is the use of chemical substances, such as bile acids or organic solvents, to dissolve gallstones. Such agents as ursodeoxycholic acid (UDCA) and chenodeoxycholic acid (CDCA) are administered orally. However, this approach has several major drawbacks, including diarrhea in about 50% of patients and a low overall efficacy. CDCA is also quite expensive. Extracorporeal shock wave lithotripsy (ESWL), which involves the breaking up of gallstones using shock waves, is another nonsurgical approach. The objective is to fragment stones into pieces small enough to be passed through the cystic duct, or small enough to allow dissolving agents to function. ESWL is safe and relatively effective under proper circumstances. The disadvantages include strict selection criteria (e.g., stones less than 2 cm in diameter), resulting in a low percentage of eligible patients. The gallbladder also is left in place, allowing for possible recurrence of gallstones and necessitating the concurrent use of dissolving agents such as UDCA or CDCA to prevent new stone formation.

For the most part, in the near future the management of gallstone disease is likely to remain surgical, with traditional or laparoscopic cholecystectomy being the major forms of intervention.

Chronic Cholecystitis

Clinical manifestations. Chronic cholecystitis is defined as chronic inflammation of the gallbladder wall attributable to persistent low-grade irritation from gallstones or to recurrent attacks of acute cholecystitis. Diabetes mellitus and obesity are important predisposing factors. Although many patients suffer from intermittent biliary colic or have symptomatic acute attacks, a surprising number of patients experience no symptoms.

Chronic cholecystitis may lead to many of the complications described earlier for acute cholecystitis, including biliary sepsis, as well as a specific type of scarring known as a *calcified* or *porcelain gallbladder,* which is associated with a higher risk of cancer.

Biliary Malignancy

Cancers of the biliary system are relatively rare; in the United States they account for only 1 to 2 cases per 100,000 per year. Unfortunately, they tend to be asymptomatic and progress insidiously until well advanced. Chronic cholecystitis predisposes sufferers to the very uncommon gallbladder cancer, but it is only a minor risk factor for cancer of the biliary tract. Primary sclerosing cholangitis is a significant risk factor, and among immigrants from endemic countries infestation with liver flukes such as *Clonorchis*/Opisthorchis species can lead to biliary cancer. Typical symptoms at presentation include right upper quadrant pain and jaundice associated with biliary obstruction.

Surgery for cure is a treatment option in fewer than 10% of cases. Chemotherapy and radiation therapy are indicted for palliation, and certain patients may benefit from stenting and other procedures to alleviate biliary obstruction.

KEY POINTS

- Lecithin is an important component of bile that helps keep cholesterol from precipitating into crystals. Crystals of cholesterol may initiate gallstone formation. The relative concentrations of cholesterol, lecithin, and bile acids appear to determine the likelihood of cholesterol gallstone formation. Bile hypomotility or stasis contributes to growth of cholesterol stones.
- Gallstones occur more frequently in women than in men. Ethnicity, obesity, and rapid weight loss are predisposing factors. Gallstones may be asymptomatic or associated with symptomatic cholecystitis. Colicky pain attributable to intermittent obstruction of the cystic duct by a stone is the chief complaint. Symptoms of chronic cholecystitis include epigastric or right upper quadrant pain radiating to the back, nausea, vomiting, sweating, fat intolerance, bloating, and flatus.
- Acute cholecystitis is acute inflammation of the gallbladder associated with abdominal pain, leukocytosis, and fever. Cholelithiasis is present in about 90% of patients; obstruction of the cystic duct is present in nearly all patients.
- Treatment for cholecystitis includes surgical removal of the gallbladder (cholecystectomy), chemodissolution, ESWL (lithotripsy) for stones, antibiotics if indicated, and management of pain. Cholecystectomy is the mainstay of therapy.
- Biliary cancer is a rare malignancy with a poor prognosis. Surgery is the only curative therapy, but is only available for a small minority of sufferers.

DISORDERS OF THE PANCREAS

Pancreatitis

Acute Pancreatitis

Etiology and pathogenesis. Acute pancreatitis is an inflammatory process involving the pancreas that may range from mild and

BOX 37.1 Conditions Predisposing to Acute Pancreatitis

Gallstones
Biliary sludge and microlithiasis
Other causes of mechanical ampullary obstruction
Alcohol
Hypertriglyceridemia
Hypercalcemia
Drugs
Infections and toxins
Trauma
Pancreas divisum
Vascular disease
Pregnancy
Post-ERCP
Postoperative pancreatitis
Hereditary pancreatitis
Structural abnormalities
- Duodenum/ampullary region
- Bile duct
- Sphincter of Oddi dysfunction
- Main pancreatic duct

From Feldman M et al: *Sleisenger and Fordtran's gastrointestinal and liver disease: pathophysiology, diagnosis, management,* ed 10, Philadelphia, 2015, Saunders.
ERCP, Endoscopic retrograde cholangiopancreatography.

inconveniencing to severe and life threatening. After an attack, the exocrine and endocrine functions of the pancreas may remain impaired for a variable period. Pancreatitis affects between 5 and 35 per 100,000 individuals in the United States annually. Predisposing factors for pancreatitis have been well known for more than 100 years (Box 37.1); in the United States the most common causes are biliary tract disease, hypertriglyceridemia, and ethanol-associated pancreatitis. Although the exact mechanisms leading to pancreatitis are not fully understood, three possible pathways are known (Fig. 37.8). The most prominent factor is obstruction of the pancreatic duct by a stone or other cause (usually unknown), with release of digestive enzymes within the parenchyma, followed by enzyme activation and then autodigestion of the pancreas. Edema leading to vascular insufficiency and ischemic injury is a contributing factor. Other possible mechanisms include acinar cell injury from alcohol or drugs, trauma, or viral infection and defective intracellular transport of proenzymes within acinar cells.

Up to 66% of first cases of pancreatitis are associated with alcoholism. Although there is clearly an association between alcohol and pancreatitis, the causal mechanism has not been determined. Transient increases in pancreatic exocrine secretion, contraction of the sphincter of Oddi, and direct toxic effects on acinar cells have all been postulated from experimental studies. Accumulating evidence suggests that many cases of alcoholic pancreatitis are sudden exacerbations of chronic pancreatitis, presenting as apparent de novo acute pancreatitis. According to this view, chronic alcohol ingestion causes secretion of protein-rich pancreatic fluid, leading to deposition of inspissated protein plugs and obstruction of small pancreatic ducts, followed by the train of events described previously. However, other pathologic studies show no evidence of chronic pancreatitis in up to 40% of acute alcoholic pancreatitis patients.

Clinical manifestations. The presentation of acute pancreatitis usually begins with steady, boring pain in the epigastrium or left upper quadrant, which gradually increases in intensity. It often radiates or penetrates through to the back and is accompanied by nausea and vomiting. Tenderness on palpation may be exquisite. Bowel sounds are reduced but not absent. Abdominal distention may be present. Fever is common but is usually low grade initially. In more severe pancreatitis, this clinical picture is accompanied by signs of circulatory instability, respiratory insufficiency, and shock.

Diagnosis. The laboratory evaluation of acute pancreatitis begins with measurements of serum pancreatic enzymes. Serum lipase and amylase levels rise more or less in tandem during the first 12 hours and remain elevated for several days. Lipase is more specific and persists longer, and therefore has become the preferred test for most clinicians. Levels of serum hepatic aminotransferases (aspartate aminotransferase, alanine aminotransferase) may also be elevated. Marked elevation of the alkaline phosphatase and bilirubin levels suggests the possibility of biliary disease or obstruction, particularly by gallstones. Associated laboratory findings include leukocytosis, hyperlipidemia (which may be marked), and hypocalcemia.

The diagnosis of acute pancreatitis is based on the signs and symptoms, laboratory data, and imaging studies of the pancreas and surrounding organs. Plain radiographs of the abdomen may reveal an ileus pattern, or a "sentinel loop" (a distended loop of small bowel in the area of the pancreas). Ultrasound can provide a convenient bedside technique to visualize the pancreas, gallbladder, common bile duct, and other abdominal structures, but is limited by poor image resolution attributable to bowel gas. CT of the abdomen is the gold standard for evaluation of the pancreas and allows depiction of the pancreas in remarkable detail, including edema and abscess or cyst formation. Magnetic resonance cholangiopancreatography (MRCP) may be used on select cases to delineate the ductal system in greater detail and to identify causative factors. The differential diagnosis of acute pancreatitis includes perforated peptic ulcer, acute cholecystitis, mesenteric vascular disease, and a variety of other illnesses (Box 37.2), most of which may be differentiated on the basis of biochemical and radiographic tests.

Grading systems allow prediction of the clinical course of acute pancreatitis. Ranson criteria are a widely used benchmark for prognostic assessment and are particularly useful with modifications based on CT scoring. Early monitoring in the intensive care unit is indicated for patients with a large number of risk factors. One particularly important finding on contrast CT is the presence of significant pancreatic necrosis. Acute necrotizing pancreatitis (ANP) carries a high risk for progression to infected pancreatic necrosis, a devastating complication with a high morbidity and mortality.

Treatment. Conservative management is indicated for mild to moderate cases of acute pancreatitis. In general, withholding oral feedings, providing nasogastric suction for significant adynamic ileus, and providing careful volume replacement with IV fluids are indicated. Analgesics are administered parenterally. All narcotics should be used carefully because of the potential of sphincter of Oddi dysfunction, although recent studies show that no single agent is contraindicated. This treatment is often sufficient when carried out for 3 to 7 days, after which the acute episode subsides and oral intake may gradually be resumed.

Severe pancreatitis, particularly in the setting of acute necrotizing pancreatitis, may result in multisystem organ dysfunction, requiring aggressive support in the intensive care unit setting. Nutritional deficits develop rapidly with extensive catabolism (tissue breakdown) and lack of caloric intake; TPN is usually indicated with pancreatitis of more than a few days' duration or if complications arise. Additional supportive measures include calcium administration to reverse severe hypocalcemia, correction of magnesium deficiency, and control of hyperglycemia. Causes of death from severe pancreatitis include respiratory failure (usually associated with acute respiratory distress syndrome), acute

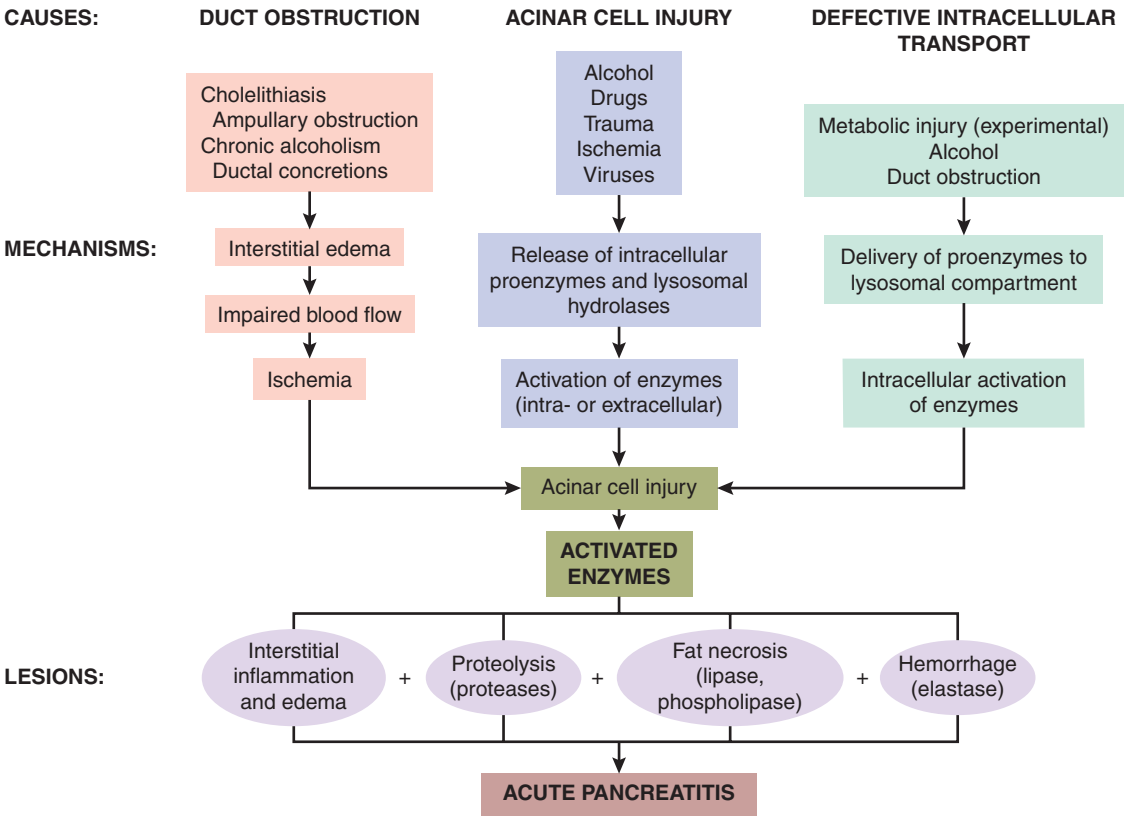


FIG 37.8 Three proposed pathways in the pathogenesis of acute pancreatitis. Obstruction of the duct, extrinsic injury, and intrinsic metabolic mechanisms lead to injury of pancreatic cells. Injured cells release activated pancreatic enzymes that cause autodigestion and inflammation of the pancreas.

renal failure, and acute intraabdominal sepsis. Mechanical ventilation and hemodialysis may be required in complicated cases.

Bacterial infection is a critical determinant of poor outcome in acute necrotizing pancreatitis. In the patient with fever and other signs of sepsis, empiric broad-spectrum antibiotics (e.g., carbapenems) should be administered, and any significant fluid collections found on CT scan should be aspirated for culture and sensitivity. The indiscriminate administration of antibiotics for acute necrotizing pancreatitis has not been shown to be helpful and carries the risk of significant side effects.

Abscess or hemorrhage may complicate pancreatitis, with or without significant pancreatic necrosis, and may necessitate percutaneous drainage or surgical intervention. Open laparotomy with debridement of devitalized tissue (necrosectomy) and major pancreatic resection (pancreatectomy) are the main surgical options. Drains are typically left in place postoperatively, and repeated debridement may be needed to remove infective debris and necrotic tissue. Pancreatic surgery is technically challenging and risky, and generally is only considered as a last resort.

Localized complications of acute pancreatitis may result in prolonged morbidity for the patient. The most common localized complication is pancreatic pseudocyst. This is a collection of fluid within or adjacent to the pancreas that often has a direct communication to the pancreatic duct. Unlike a true cyst, a pseudocyst contains no epithelial lining. It can develop rather acutely or more subacutely as the patient recuperates from the acute illness. The presentation generally includes an abdominal fullness or mass and tenderness. Complications of pseudocysts include infection (usually termed an *infected pseudocyst,* as opposed to the pancreatic abscess described earlier), spontaneous rupture, or hemorrhage. Management of pseudocyst includes percutaneous, endoscopic, or surgical drainage of the cyst, either externally or internally, usually into the stomach or bowel.

Pancreatic ascites may occur as a result of a persistent leak in the main pancreatic duct. It is usually painless and often massive. The fluid may find its way into unusual places, including the pleural space and mediastinum. Pancreatic ascites may be detected by ultrasonography or CT, and diagnosis is confirmed by analysis of fluid obtained by aspiration, in particular, the amylase level. Management is often conservative, with prolonged parenteral nutrition. Improvement may occur after the endoscopic placement of a stent (a thin-walled tube) into the main pancreatic duct.

Other complications of acute pancreatitis include common bile duct obstruction, portal or splenic vein thrombosis, peptic ulcer disease, and chronic fistula formation.

Endoscopic treatment may be carried out for gallstone pancreatitis in selected cases. Indications for urgent ERCP with ampullotomy (incision of the ampulla) include biliary sepsis, recalcitrant severe pancreatitis, and jaundice. In milder cases, traditional conservative therapy followed by elective ERCP is acceptable. The risks of ERCP include exacerbation of pancreatitis, and therefore the need for this procedure must be carefully considered.

Chronic Pancreatitis

Etiology and pathogenesis. Chronic pancreatitis is defined histologically as the presence of chronic inflammatory lesions in the pancreas, and on a clinical basis one sees the persistence of symptoms secondary to pancreatic dysfunction over weeks and months. Destruction of exocrine parenchyma and fibrosis precede the destruction of endocrine

BOX 37.2 Causes of Increased Serum Amylase Activity

Pancreatic diseases
- Acute pancreatitis
- Complications of pancreatitis
- Acute exacerbation of chronic pancreatitis
- Pancreatic tumors, cysts

Other serious intraabdominal diseases
- Acute cholecystitis
- Common bile duct obstruction
- Perforation of esophagus, stomach, small bowel, or colon
- Intestinal ischemia or infarction
- Intestinal obstruction
- Acute appendicitis
- Acute gynecologic conditions such as ruptured ectopic pregnancy and acute salpingitis

Diseases of salivary glands
- Mumps
- Effects of alcohol

Tumors
- Ovarian cysts
- Papillary cystadenocarcinoma of ovary
- Carcinoma of lung

Renal insufficiency

Macroamylasemia

Miscellaneous
- Morphine
- Endoscopy
- Sphincter of Oddi stenosis or spasm
- Anorexia nervosa
- Head trauma with intracranial bleeding
- Diabetic ketoacidosis
- Human immunodeficiency virus

From Feldman M et al: *Sleisenger and Fordtran's gastrointestinal and liver disease: pathophysiology, diagnosis, management,* ed 10, Philadelphia, 2015, Saunders.

parenchyma. After a variable time, most patients with chronic pancreatitis develop calcifications that become visible on radiologic films of the abdomen or CT. Chronic pancreatitis is most often associated with alcohol consumption, although a small percentage of cases are idiopathic, hereditary, or associated with hyperparathyroidism (hypercalcemia), trauma, or various other factors.

The association of alcohol ingestion with chronic pancreatitis is profound. Autopsy studies have shown that the changes of chronic pancreatitis are present in 45% of alcoholics, even those without symptoms, and that this rate is 40 to 50 times higher than that in nondrinkers. Exactly how alcohol causes chronic pancreatitis is not known (see the previous discussion of alcohol and acute pancreatitis). One recent theory suggests that the initial factors include an increase in the protein concentration in pancreatic juice coupled with reduction in a specific "pancreatic stone protein" that inhibits the formation of pancreatic protein plugs. This biochemical situation allows the formation of protein plugs that can later calcify, in addition to causing obstruction to the flow of pancreatic juice. A key element seems to be necrosis, followed by fibrosis, perhaps analogous to cirrhosis of the liver. Ischemic and antioxidant damage may occur as well. Another facet of alcohol-associated chronic pancreatitis is its tendency to progress after alcohol consumption is stopped.

Clinical manifestations. The presentation of chronic pancreatitis may consist of bouts of acute pancreatitis with progressive signs of persistent pancreatic dysfunction after the acute attack subsides. Alternatively, an insidious onset of pain in the epigastrium that radiates to the back may be the first symptom. About 10% to 15% of patients will not present with pain, but rather with the sequelae of chronic pancreatitis, including diabetes mellitus, malabsorption, and weight loss. The mortality is 3% to 4% per year. Interestingly, the incidence of pancreatic carcinoma does not appear to be substantially increased in patients with chronic pancreatitis.

The pain of chronic pancreatitis is often the major form of debility. Nerve fibers from the pancreas pass to the celiac plexus and then to spinal sympathetic ganglia. The events that actually trigger the pain are not well understood. There may be a relation to ductal pressures or possibly to ischemia in the pancreas. The pain is often accompanied by nausea and is steady and boring in nature. The pain is usually located in the upper abdomen, particularly in the epigastrium, and radiates to the back in more than half of cases. In alcoholic pancreatitis, continued drinking affords temporary anesthesia but may foster recurrences of pain. Cessation of drinking may allow for a better long-term prognosis. After about 5 years of continual pain, many patients note a decrease in the symptoms (i.e., the pain "burns out").

Endocrine and exocrine pancreatic insufficiency lead to diabetes mellitus, malabsorption, and weight loss. Diabetes mellitus arises from progressive loss of endocrine cells in the pancreatic islets and usually requires exogenous insulin administration; diabetic ketoacidosis is an unusual finding. Weight loss may be aggravated by poor intake as a result of pancreatic pain. Malabsorption of fat does not occur until pancreatic enzyme output drops to 10% of normal. Along with the malabsorption of fat, the absorption of fat-soluble vitamins (A, D, E, and K) may be impaired, leading to such problems as coagulopathy and night vision impairment.

Further complications of chronic pancreatitis are similar to those of acute pancreatitis and include pseudocyst, pancreatic ascites, and obstruction of the common bile duct. Obstruction of the bile duct may lead to elevated values on liver function tests and the need to intervene either surgically or endoscopically. Alkaline phosphatase and bilirubin levels may become markedly elevated if obstruction is severe. Unusual complications include thrombosis of the portal and splenic veins. This may lead to gastrointestinal hemorrhage from gastric varices. Peptic ulcer disease is also increased in patients with chronic pancreatitis, although a definite causal relationship has not been established.

Diagnosis. The diagnosis of chronic pancreatitis is usually suggested by the clinical history, physical examination findings, and routine blood chemical analyses. Biochemical studies of pancreatic function may be helpful. Confirmation of the diagnosis is aided by plain radiographs showing calcifications in the area of the pancreas. Abdominal ultrasound, CT, or MRCP is usually performed with reasonable sensitivities and specificities. ERCP is reserved for suspicious cases that cannot be confirmed by other techniques or for cases in which biopsy or cytologic examination is necessary to rule out malignancy. It shows the pancreatic duct to range from almost normal in early cases to markedly dilated or beaded—the "chain of lakes" appearance (Fig. 37.9). A common finding is truncation of the secondary branches of the pancreatic duct.

Treatment. The treatment for chronic pancreatitis is directed toward controlling pain, addressing exocrine and endocrine insufficiency, and managing complications. By far the most challenging is the management of pain. Absolute abstention from alcohol is paramount to prevent worsening of symptoms. For almost 40 years, analgesics and surgical intervention have been the mainstay of pain control, and celiac plexus block is helpful for some patients. With the advent of ERCP, less drastic forms of intervention are now possible. Pancreatic sphincterotomy is

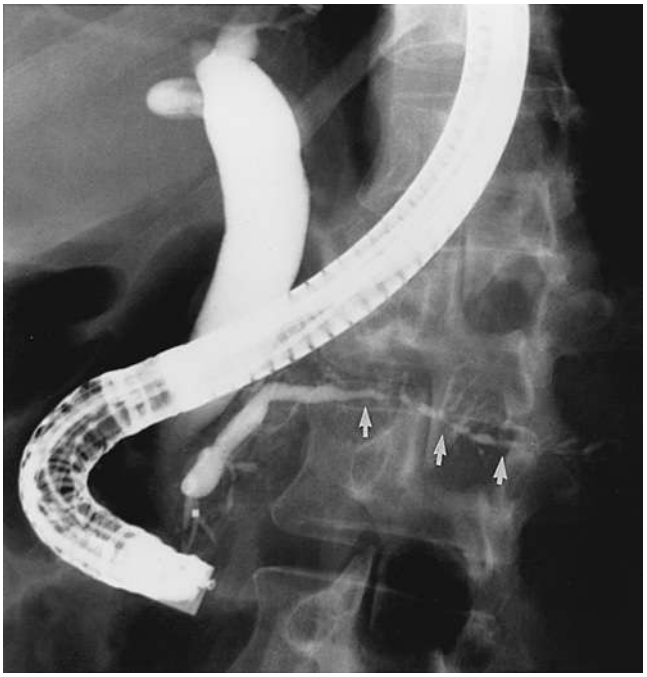


FIG 37.9 Endoscopic retrograde cholangiopancreatogram in a patient with chronic pancreatitis shows marked narrowing and irregularity of the main pancreatic duct body and tail *(arrows).* (From Feldman M et al: *Sleisenger and Fordtran's gastrointestinal and liver disease: pathophysiology, diagnosis, management,* ed 6, Philadelphia, 1998, Saunders, p 952.)

indicated for the management of single or multiple stones. Endoscopic drains may be placed for pseudocysts of the pancreas if they are adjacent to the stomach or duodenum. Obstruction of the common bile duct can be managed with endoscopically placed biliary stents. Strictures of the main pancreatic duct can be managed with indwelling pancreatic stents.

If endoscopic management fails or is not appropriate in a given patient, surgery may be indicated. Various procedures, such as the pylorus-preserving Whipple procedure (pancreaticoduodenectomy), may be used, generally with favorable results regarding relief of pain (though with significant morbidity and mortality). Complete and distal pancreatic resections are rarely used for this condition. As with management of acute pancreatitis, the judgment of the surgeon is paramount to prevent the catastrophic complications that can result from surgery on the pancreas.

Pancreatic enzyme replacement is a standard therapy for chronic pancreatitis, both for its effects on steatorrhea and for the management of chronic pain. Because proteases (e.g., trypsin, chymotrypsin, and elastase) exert a controlling influence on pancreatic secretion, feedback regulation should result in relief of pain after oral administration of pancreatic enzymes. Unfortunately, only 20% to 30% of patients with the typical alcohol-induced type of disease respond to such therapy. Responses seem to be higher in patients with small-duct disease. Acid suppression with an H_2-blocker or proton pump inhibitor is indicated to reduce inactivation of the enzymes from gastric acid. Medium-chain triglycerides can provide extra calories in patients with weight loss. Research on octreotide, a synthetic long-acting analog of somatostatin that has been shown to inhibit CCK release and both basal and neural-stimulated pancreatic secretion, has generally been disappointing.

Management of exocrine insufficiency can usually be accomplished with low-fat diets and pancreatic enzyme supplementation. Likewise, endocrine insufficiency in the form of diabetes mellitus is managed with diet and either oral hypoglycemic agents or insulin.

Pancreatic Cancer

Pancreatic cancer is a challenging malignancy with a generally dismal prognosis, which has been unfortunately increasing in incidence, particularly among nonwhite populations. About 50,000 cases are diagnosed annually in the United States with a slight male predominance. The current incidence comprises about 2% of all cancers, but this cancer ranks fourth among deaths from all malignancies. Cigarette smoking and obesity are risk factors for pancreatic cancer, both increasing the risk by about 50% compared with the general population. About 5% to 10% of cases have a hereditary component. The vast majority of pancreatic tumors are adenocarcinomas, but an important subset of neoplasms is neuroendocrine tumors.

Lesions of the head of the pancreas outnumber those in the body and tail 3 to 1. Symptoms of pancreatic head tumors include jaundice, malabsorption, and weight loss, whereas those of the tail generally include abdominal pain and nausea. The distinction between chronic pancreatitis, autoimmune pancreatitis, or benign cystic adenomas and pancreatic cancer may be difficult, and even a negative needle biopsy does not preclude cancer. Most cancers are advanced at the time of presentation, with only 15% to 20% of cases being candidates for surgical resection. Patients presenting with painless jaundice attributable to an obstructing pancreatic head lesion have the best prognosis. Even with surgery, most patients die of their disease. Chemotherapy plays a role in attempts at both surgical cure and palliation. The median survival among all patients is only 1 year, though there is some evidence that the cure rate may be improving somewhat because of new therapeutic modalities being employed at major medical centers.

KEY POINTS

- Acute pancreatitis is commonly associated with biliary tract disease and excessive ethanol ingestion. Activation of pancreatic proenzymes to active forms within the pancreas leads to autodigestion and inflammation of the gland. The manifestations of acute pancreatitis may be mild or severe and include a steady, boring pain in the epigastrium or left upper quadrant, nausea, vomiting, a tender abdomen, reduced bowel sounds, and fever. In severe cases, circulatory shock may occur. Elevated serum amylase and lipase levels are indicative of pancreatitis.
- Management of acute pancreatitis is aimed at reducing pancreatic secretion. Because chyme entering the duodenum is the primary stimulus for pancreatic secretion, food is withheld, and nasogastric suctioning may be instituted. Complications of acute pancreatitis include hyperglycemia, nutritional deficit, pancreatic hemorrhage, infection, abscess formation, or necrosis. Antibiotics, fluid management, TPN, and insulin may be indicated to manage complications.
- Chronic pancreatitis is closely associated with alcohol use. Acute pancreatitis attributable to biliary obstruction rarely progresses to chronic pancreatitis. Chronic pancreatitis results in progressive destruction of endocrine and exocrine function. Manifestations of chronic pancreatitis are more insidious than those of acute pancreatitis. Epigastric pain, diabetes mellitus, malabsorption, and weight loss may be the presenting problems.
- The complications of chronic pancreatitis are similar to those of acute pancreatitis. Therapy is directed at controlling pain, ameliorating endocrine and exocrine deficiency, and monitoring and managing complications. Surgery to correct obstruction of the pancreatic duct may be performed. Pancreatic enzyme therapy may be helpful in reducing pain by providing negative feedback, which reduces pancreatic secretion.
- Pancreatic cancer is a highly fatal cancer that can be difficult to diagnose and treat. Median survival is 12 months.

SUMMARY

The pancreaticobiliary system is central to the digestion of food because it provides necessary digestive enzymes and lipid-emulsifying agents that allow the intestine to absorb nutrients. This chapter has considered alterations in the function of the gallbladder and exocrine pancreas. A major disease of the pancreaticobiliary system is the formation of cholesterol gallstones, which can lead to acute and chronic cholecystitis and acute and chronic pancreatitis. New forms of surgical and nonsurgical interventions for the management of gallstone disease have become available in the past few years, with conventional open surgery remaining a useful option. These interventions, as well as interventions for acute and chronic pancreatitis, are currently the focus of much clinical research.

RESOURCES

Development, Structure, and Function

Avunduk C: *Manual of gastroenterology*, ed 4, Philadelphia, 2008, Lippincott Williams & Wilkins.

Hall JE: *Guyton and Hall textbook of medical physiology*, ed 13, Philadelphia, 2015, Saunders Elsevier.

Moore KL, Persaud TVN: *The developing human: Clinically oriented embryology*, ed 10, Philadelphia, 2015, Saunders.

Moore KL, Persaud TVN, Shiota K: *Color atlas of clinical embryology*, ed 2, Philadelphia, 2000, Saunders.

Podolski DK, et al: *Yamada's textbook of gastroenterology*, ed 6, Oxford, 2015, Wiley-Blackwell.

Ross AC, et al: *Modern nutrition in health and disease (Shils)*, ed 11, Philadelphia, 2012, Lippincott Williams & Wilkins.

Gallbladder Disorders

Feldman M, Friedman LS, Brandt LJ: *Sleisenger and Fordtran's gastrointestinal and liver disease*, ed 10, Philadelphia, 2015, Saunders Elsevier.

Kelsen DP, et al: *Principles and practice of gastrointestinal oncology*, ed 2, Philadelphia, 2008, Lippincott Williams & Wilkins.

Kumar V, Abbas AK: *Robbins pathologic basis of disease*, ed 9, Philadelphia, 2014, Saunders.

Townsend CM, Beauchamp RD: *Sabiston textbook of surgery: the biological basis of modern surgical practice*, ed 19, Philadelphia, 2012, Saunders Elsevier.

Pancreatitis

Chatzicostas C, et al: Computed tomography severity index is superior to Ranson criteria and Apache II and III scoring systems in predicting acute pancreatitis outcome. *J Clin Gastroenterol* 36(3):253–260, 2003.

Haubrich WS, Schaffner F, Berk JE: *Bockus gastroenterology*, ed 5, Philadelphia, 1995, Saunders.

Laoser C, Faolsch UR: A concept of treatment in acute pancreatitis: results of controlled trials and future developments. *Hepatogastroenterology* 40:569–573, 1993.

Singh VV, Toskes PP: Medical therapy for chronic pancreatitis pain. *Curr Gastroenterol Rep* 5(2):110–116, 2003.

Sole ML, Klein DG: *Introduction to critical care nursing*, ed 6, Philadelphia, 2012, Saunders Elsevier.

Thompson DR: Narcotic analgesic effects on the sphincter of Oddi: a review of the data and therapeutic implications in treating pancreatitis. *Am J Gastroenterol* 96(4):1266–1272, 2001.

Treacy PJ, Worthley CS: Pancreatic stents in the management of chronic pancreatitis. *Aust N Z J Surg* 66:210–213, 1996.

Vege SS, Yadav D, Chari ST: Pancreatitis. In Talley NJ, Locke GR, Saito YA, editors: *GI epidemiology*, ed 1, Malden, MA, 2007, Blackwell Publishing.

Yamedera K, Moriyama T, Makino I: Identification of immunoreactive pancreatic stone protein in pancreatic stone, pancreatic tissue and pancreatic juice. *Pancreas* 5(3):255–260, 1990.

38

Liver Diseases

Jeffrey S. Sartin

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- What role does the liver play in nutrient metabolism, bile synthesis, storage of vitamins and minerals, urea synthesis, clotting factor synthesis, and detoxification?
- Which manifestations of liver disease are due to hepatocellular failure and which are due to portal hypertension?
- How do the different types of viral hepatitis vary with regard to mode of transmission and severity of symptoms?
- What clinical and laboratory findings would lead to a diagnosis of liver cirrhosis?
- What treatment modalities are available to patients with end-stage liver failure?

CHAPTER OUTLINE

Structure and Function of the Liver, 755
General Manifestations of Liver Disease, 755
Hepatocellular Failure, 755
Jaundice, 755
Portal Hypertension, 759
Gastroesophageal Varices, 760
Portal Systemic Encephalopathy, 763
Hepatic Encephalopathy, 763
Cerebral Edema, 764
Complications of Advanced Liver Disease, 764
Ascites, 764
Spontaneous Bacterial Peritonitis, 765
Hepatorenal Syndrome, 766
Disorders of the Liver, 767
Hepatitis, 767
Acute Viral Hepatitis, 767
Hepatitis A, 767
Hepatitis B, 767
Hepatitis C, 771
Hepatitis D (Delta), 771
Hepatitis E, 771
Chronic Hepatitis, 772
Chronic Persistent Hepatitis, 772
Chronic Active Hepatitis, 772
Autoimmune Hepatitis, 772
Cirrhosis, 773
Biliary Cirrhosis, 773
Primary Sclerosing Cholangitis, 773
Alcoholic Liver Disease, 773
Alcoholic Fatty Liver, 773
Alcoholic Hepatitis, 773
Toxic Liver Disorders, 774
Metal Storage Diseases, 774
Hereditary Hemochromatosis, 774
Wilson Disease (Hepatolenticular Degeneration), 774
Toxic Metabolic Agents, 775
Acetaminophen Poisoning, 775
Other Structural Liver Conditions, 775
Liver Abscess, 775
Trauma, 776
Malignancy, 776
Transplantation, 776
Evaluation of the Transplantation Patient, 777
Posttransplantation Management, 777
Age-Related Liver Disorders, 778
Liver Diseases and Pediatric Considerations, 778
Abnormal Bilirubin Metabolism in the Neonatal Period, 778
Infectious and Acquired Hepatitides in Children, 778
Congenital Liver Disease, 779
Multisystem Enzyme Deficiencies, 779
Disorders of Bilirubin Metabolism, 779
Inborn Errors of Metabolism, 780
Intrahepatic Cholestatic Conditions, 780
Extrahepatic Cholestatic Conditions (Biliary Atresia), 780
Liver Diseases and Geriatric Considerations, 780

The liver is a vital but vulnerable organ. Its role in digestion of fats, storage of carbohydrates, detoxification of blood, and production of proteins makes it indispensable; in contrast to the kidney and heart, there are no "artificial livers." Nevertheless, the liver is susceptible to a wide variety of metabolic, circulatory, toxic, microbial, and neoplastic insults. In some instances, the disease is primary to the liver, as in viral hepatitis and hepatocellular carcinoma (HCC). More often the hepatic involvement is secondary, a consequence of some of the more common diseases of humans, such as cardiac decompensation, metastatic cancer, alcoholism, and infections. This chapter will focus on primary diseases of the liver.

STRUCTURE AND FUNCTION OF THE LIVER

The liver, the largest parenchymal organ of the body, averages 1500 grams in weight. It is located in the right upper quadrant of the abdomen, beneath the diaphragm, and is anatomically divided into right and left lobes and then further subdivided according to the pattern of its blood supply and biliary drainage. It is covered by a connective tissue capsule, the Glisson capsule, which in turn is covered by visceral peritoneum, reflections of which form the various suspensory hepatic ligaments. These structures demarcate the bare area of the liver directly in contact with the diaphragm (Fig. 38.1).

The liver has a dual blood supply. Arterial inflow from the aorta via the celiac trunk and hepatic artery provides 25% of the organ's blood supply, with the remainder from the portal vein, which drains the capillary bed of the alimentary canal and pancreas (Fig. 38.2). This oxygen-depleted venous blood is rich in substances absorbed and secreted by the gut. The afferent blood vessels then branch throughout the liver in association with the bile ducts and form the portal triads (consisting of the portal veins, hepatic arteries, and bile ducts). Eventually, blood from both the hepatic artery and the portal vein drains into the hepatic sinusoids, which surround sheets of liver cells, or hepatic plates (Fig. 38.3). The sinusoids are lined by endothelial cells and Kupffer cells (a type of phagocytic macrophage). This blood drains into the central veins, which finally coalesce into the hepatic vein and empty into the inferior vena cava. Any obstruction to the flow of blood may result in a rise in portal venous pressure proximal to the level of blockage. This condition is called *portal hypertension* and is a central pathophysiologic event in many liver diseases. The liver also has a rich and complex lymphatic drainage system.

The liver is one of the most metabolically active organs in the body and functions simultaneously as a digestive organ, an endocrine organ, a hematologic organ, and an excretory organ (Box 38.1). All of these functions are elegantly interwoven with such redundancy that more than 80% of the liver may be destroyed before life is threatened.

BOX 38.1 Summary of Normal Liver Function

The Liver as a Digestive Organ
Bile salt secretion for fat digestion
Processing and storage of fats, carbohydrates, and proteins absorbed by the intestines
Processing and storage of vitamins and minerals

The Liver as an Endocrine Organ
Metabolism of glucocorticoids, mineralocorticoids, and sex hormones
Regulation of carbohydrate, fat, and protein metabolism

The Liver as a Hematologic Organ
Temporary storage of blood
Removal of bilirubin from the bloodstream
Hematopoiesis in certain disease states
Synthesis of blood clotting factors

The Liver as an Excretory Organ
Excretion of bile pigment
Excretion of cholesterol via bile
Synthesis of urea
Detoxification of drugs and other foreign substances

GENERAL MANIFESTATIONS OF LIVER DISEASE

Whether primary or secondary, all hepatic derangements tend to cause similar signs and symptoms that are directly attributable to loss of hepatocellular function or disruption of blood flow through the liver. Because of the liver's considerable reserve, however, manifestations appear only when the injury is significant and diffuse or so strategically located that it obstructs biliary outflow.

Hepatocellular Failure

Hepatocellular failure results in a number of typical manifestations, including jaundice, muscle wasting, ascites, excessive bleeding, deficiencies of important blood proteins and vitamins, glucose imbalance, and impaired hormone production (Table 38.1). At its most basic the liver is a sophisticated biochemical factory, and these conditions all derive from problems with processing the essential molecules of the body. Inadequate protein metabolism leads to decreased production of clotting factors and hypoalbuminemia. Decreased serum albumin level in turn leads to generalized edema as a result of low serum oncotic pressure. Abnormal storage of glycogen and release of glucose may result in bouts of either hyperglycemia or hypoglycemia. Reduced production of bile salts by the liver impairs absorption of the fat-soluble vitamins A, D, E, and K from the gastrointestinal (GI) tract. Lack of vitamin D may lead to osteomalacia (impaired bone homeostasis); lack of vitamin K contributes to poor blood-clotting factor production. Altered lipoprotein processing leads to dyslipidemias, particularly hypertriglyceridemia.

Hepatocellular failure results in impaired processing of endogenous steroid hormones and the by-products of protein metabolism, as well as decreased clearance of exogenous drugs and toxins. Impaired metabolism of estrogen leads to feminization in men (gynecomastia, erectile dysfunction, testicular atrophy, female hair distribution), irregular menses in women, palmar erythema, and spider telangiectasia. Impaired conversion of ammonia to urea is associated with hepatic encephalopathy, which will be discussed later in this chapter.

Jaundice

Etiology and pathogenesis. Jaundice, the green-yellow staining of tissues by bilirubin, is the consequence of impaired bilirubin metabolism and is one of the most characteristic signs of liver disease. A study of the mechanisms of bilirubin metabolism is essential to an understanding of liver disease and may serve as a paradigm for other hepatic processes (Fig. 38.4).

As red blood cells age or are damaged by disease, they lyse and release oxygen-carrying hemoglobin molecules. These are taken up by the reticuloendothelial system, which separates heme from globin, and through the action of heme oxygenase opens the heme ring to release the central iron atom. This process yields biliverdin, which in turn is converted by the enzyme bilirubin reductase to bilirubin. (A small percentage of bilirubin is derived from the premature destruction of immature cells in the bone marrow and spleen and from heme proteins such as myoglobin and the cytochromes in the liver.) Bilirubin is released

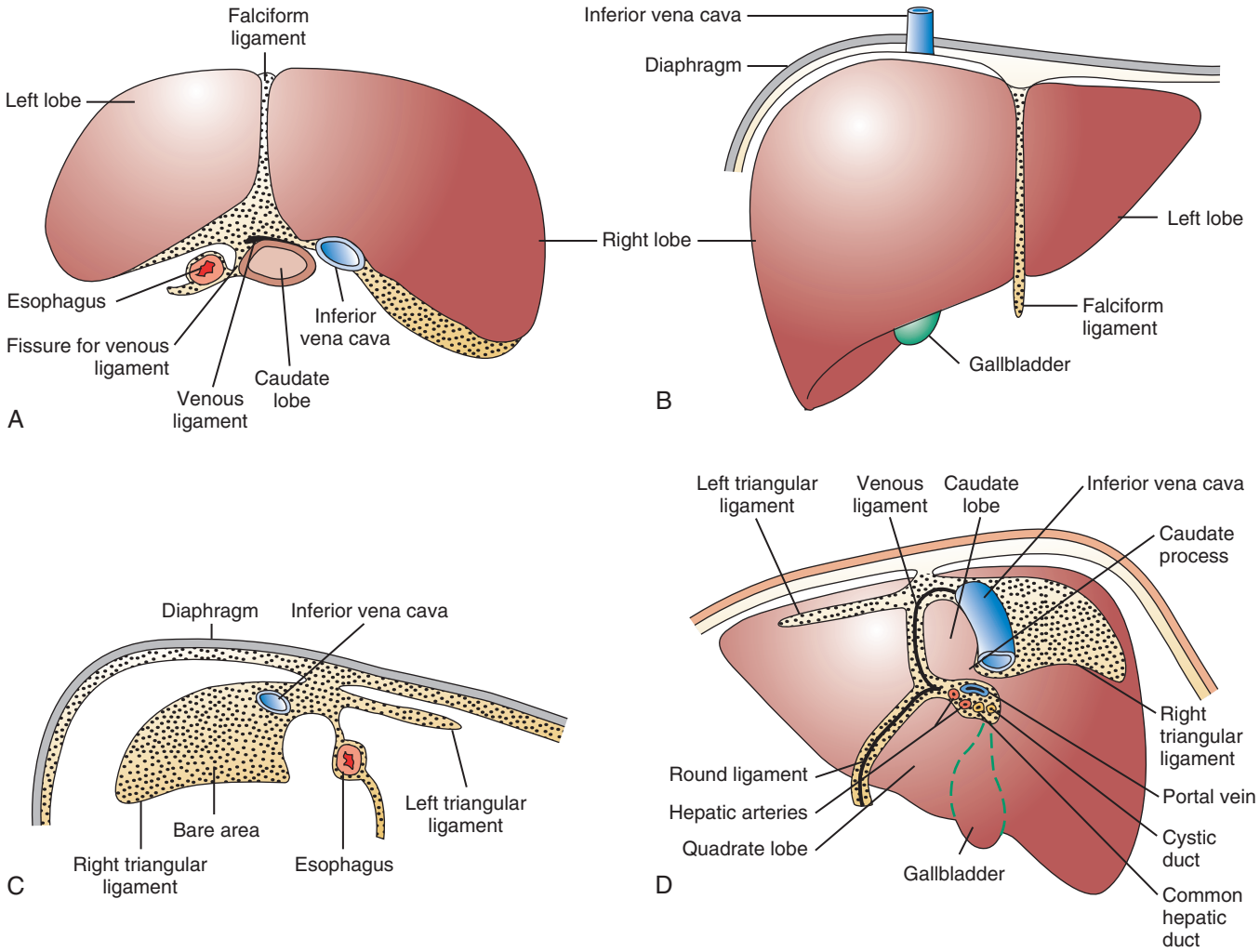


FIG 38.1 The liver and its peritoneal relations. Stippled areas represent surfaces not covered with peritoneum. **A,** Superior view. **B,** Anterior view. **C,** The diaphragm, viewed from the front, showing the position of the bare area of the liver. **D,** Visceral surface of the liver, viewed from behind. (Redrawn from Gardner E et al: *Anatomy: a regional study of human structure,* ed 3, Philadelphia, 1969, Saunders, p 414.)

into the plasma and transported to the liver tightly bound to the plasma protein albumin. The free unconjugated bilirubin is lipid soluble and can be displaced from albumin by fatty acids and certain organic anions (e.g., sulfonamides, salicylates). The neonate is particularly sensitive to free unconjugated bilirubin, which can diffuse into the brain and cause a type of encephalopathy known as *kernicterus* (see the Liver Diseases and Pediatric Considerations section).

Liver cells are able to extract unconjugated bilirubin from the plasma with special transport proteins. In the cytosol, bilirubin is quickly bound, or conjugated, to water-soluble derivatives of glucuronic acid by the action of the enzyme uridine diphosphate glucuronosyltransferase (UDPGT) located in the endoplasmic reticulum. This process yields water-soluble bilirubin monoglucuronide and diglucuronide, which is then actively excreted into microscopic bile ducts (canaliculi). Bilirubin is then transported through the biliary system as a component of bile to the small intestine. Because it cannot be absorbed in the small intestine, it passes to the colon where bacterial β-glucuronidase enzymes convert it to urobilinogen. A small fraction of urobilinogen is absorbed from the colon and reexcreted by the kidneys and the liver. In the presence of liver disease, the hepatic fraction decreases and the urinary fraction increases, thus accounting for the rise in urinary urobilinogen concentration seen with liver dysfunction. With complete obstruction to bile flow or with intestinal obstruction above the colonic level, urinary urobilinogen level falls to zero, because no bilirubin reaches the colon. (The function of bile salts and the other components of bile flow are discussed in Chapter 37.)

Therefore jaundice may result from dysfunction anywhere along this complex pathway. Classically, it is divided into prehepatic, hepatic, and posthepatic or cholestatic, but much overlap occurs.

Prehepatic. The most common causes of prehepatic jaundice are hemolysis and ineffective erythropoiesis. The resorption of large hematomas in patients with mild liver disease is a frequent and harmless cause of mild jaundice attributable to unconjugated hyperbilirubinemia.

Hepatic. Dysfunction of any of the hepatic steps in bilirubin metabolism may cause jaundice. In the neonate, immature UDPGT levels may result in physiologic jaundice of the newborn. Various genetic disorders of UDPGT synthesis are characterized by high levels of unconjugated bilirubin in the blood. Mutant UDPGT enzymes can produce the common and benign Gilbert syndrome, in which low levels of unconjugated bilirubin may be increased by fasting or illness (e.g., viral gastroenteritis). Other UDPGT mutations cause the Crigler–Najjar type I and II syndromes with severe neonatal unconjugated hyperbilirubinemia (see the Liver Diseases and Pediatric Considerations section). Most of the liver diseases to be discussed later, such as viral hepatitis,

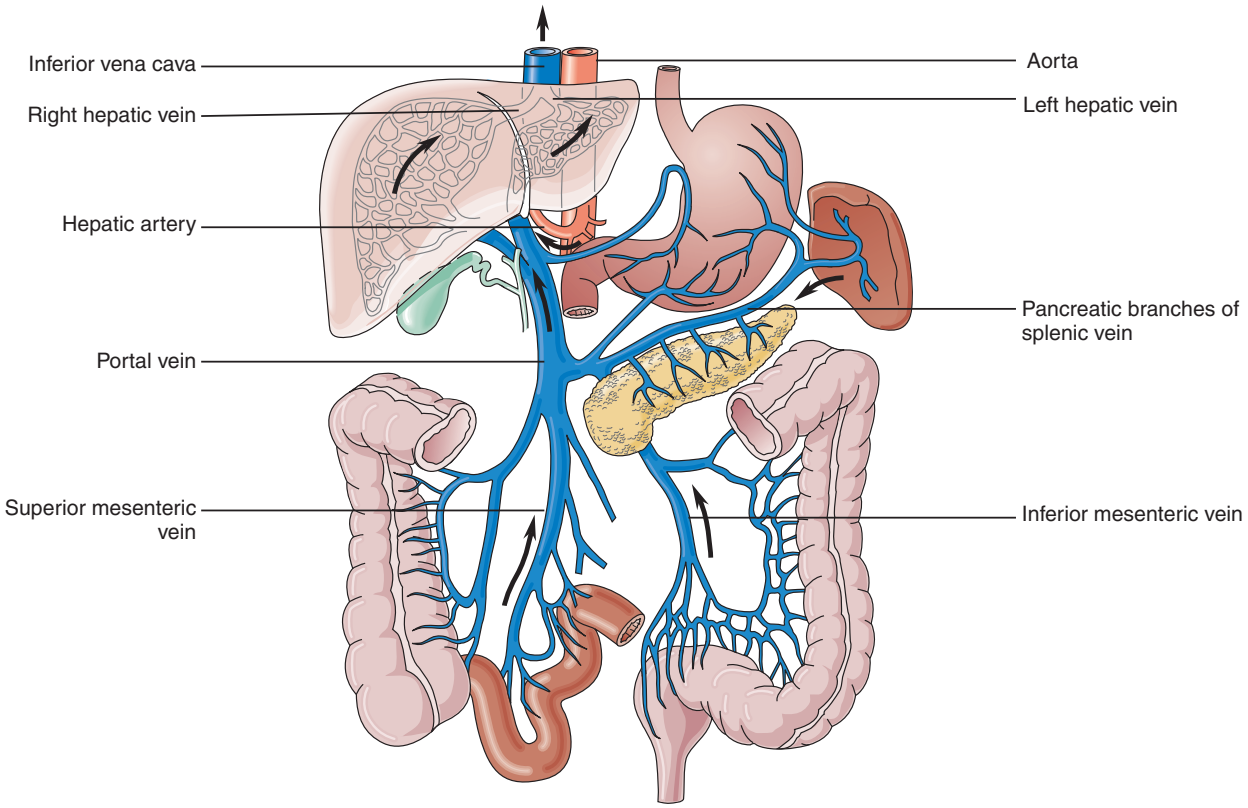


FIG 38.2 Schematic diagram of the portal circulation. Blood from the aorta supplies the alimentary canal. Venous blood from the intestine reaches the sinusoids of the liver by way of the portal vein. Venous blood from the liver reaches the inferior vena cava by way of the hepatic veins.

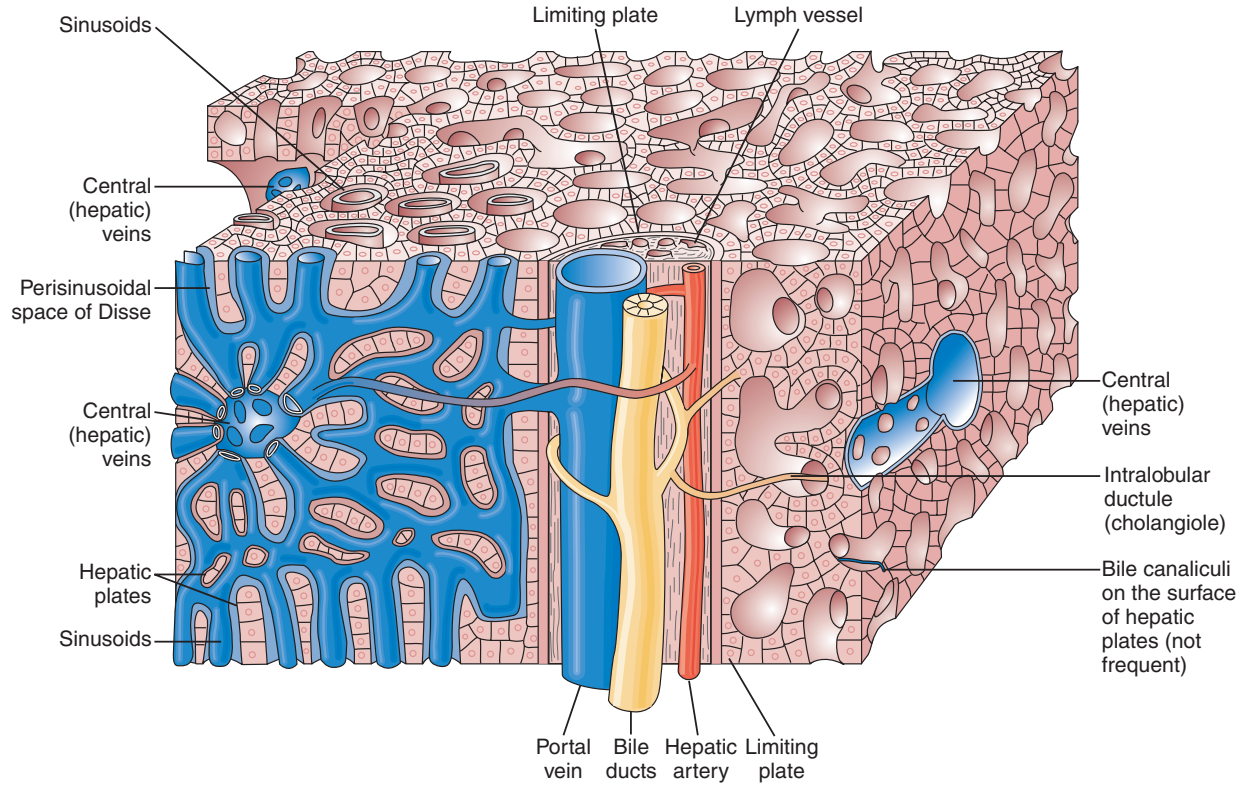


FIG 38.3 Liver lobule showing portal vein blood mixing with hepatic arterial blood to form the sinusoids that supply blood to the hepatocytes.

TABLE 38.1 Pathophysiology Underlying the Symptoms and Signs of Liver Disease

Symptoms/Signs	Pathophysiologic Mechanism
Weakness, fatigue, anorexia, weight loss, muscle wasting	Failure of multiple metabolic functions
Fever	Liver inflammation, decreased reticuloendothelial function with increased risk of infection
Bruising, increased bleeding	Thrombocytopenia secondary to splenic enlargement, decreased synthesis of clotting factors I, II, V, VII, VIII, IX, and X
Palmar erythema, cutaneous spider telangiectases, irregular menses, gynecomastia, erectile dysfunction, female body hair distribution in men, testicular atrophy	Altered metabolism of sex hormones, chronic debilitation
Hepatic encephalopathy	Abnormal protein metabolism
Fetor hepaticus	Decreased detoxification
Pruritus	Decreased bile salt excretion
Cyanosis	Arteriovenous shunts in lungs, liver
Jaundice	Biliary obstruction, decreased bilirubin synthesis, decreased bilirubin excretion
Hyperdynamic circulation, wide pulse pressure, tachycardia	Generalized vasodilation (hormonally mediated?)
Ascites, peripheral edema	Portal hypertension, sodium and water retention, low serum albumin secondary to decreased hepatic synthesis
Splenomegaly	Portal hypertension
Hepatomegaly	Cirrhosis (liver may be small), hepatitis, vascular congestion, bile duct obstruction, infection, benign infiltrative disease (e.g., fatty liver, amyloidosis, hemochromatosis), malignant infiltrative disease (e.g., metastatic cancer, lymphoma, large space-occupying lesions such as neoplasm, abscess)
Varices (esophageal, gastric, rectal, ectopic) or abnormal abdominal vascular pattern (caput medusae, umbilical bruit)	Portal hypertension with collateral blood flow around hepatic blockage
Osteomalacia, hypocalcemia, night blindness, coagulopathy	Fat-soluble vitamin malabsorption and loss of fat-soluble vitamin reserves A, D, and K; loss of vitamin K metabolism (a cofactor for I, II, VII, VIII, IX, and X)
Anemia	Multifactorial: blood loss, chronic disease, vitamin B_{12} deficiency, splenic sequestration
Leukopenia	Hypersplenism secondary to portal hypertension
Hypoglycemia	Altered glycogenolysis, gluconeogenesis
Hyperglycemia	Portosystemic shunting with delayed hepatic uptake of absorbed glucose
Hypercholesterolemia	Obstructive jaundice with decreased cholesterol excretion

alcoholic liver disease, and autoimmune hepatitis, result in jaundice because dysfunction within the liver cells result in elevated levels of conjugated bilirubin.

Posthepatic. At the level of canalicular bilirubin transport, the rare inherited Dubin–Johnson and Rotor syndromes cause conjugated hyperbilirubinemia. Both conditions have an excellent prognosis. At the canalicular posthepatocytic level, many drugs such as the phenothiazines and the sex hormones may cause jaundice. In susceptible women, the high sex hormone levels of normal pregnancy can cause benign cholestasis of pregnancy. This condition is also associated with defective transport of bile salts and is characterized by jaundice and intense pruritus (i.e., itching).

Mechanical obstruction of the bile ducts from obstructing tumors, strictures, or gallstones is the most common cause of cholestatic jaundice. Experts differentiate between obstructive jaundice caused by a gross mechanical blockage to bile flow in the biliary tract and intrahepatic cholestatic jaundice, the latter implying a defect at the microscopic level.

Evaluation. After a complete history and physical examination are obtained and routine laboratory data reviewed, specific liver-related tests may be performed (Table 38.2). As with many disease processes, the underlying cause (such as alcoholic liver disease, a drug reaction, or malignancy) is often suggested by the history and physical examination. The physical stigmata of chronic liver disease include telangiectasia, ascites, palmar erythema, gynecomastia, testicular atrophy, hair loss (in men), and central obesity with peripheral muscle wasting.

Diagnostic tests. Biochemical test abnormalities usually fall into one of several categories. A significant elevation in the levels of transaminases out of proportion to the other liver enzymes indicates a primarily hepatocellular disorder (i.e., hepatitis). Alcoholic and other toxic hepatitides virtually always show the aspartate aminotransferase (AST) level markedly elevated in relation to the alanine aminotransferase (ALT), whereas in viral hepatitis the reverse is usually true. Predominant elevation of alkaline phosphatase (ALP) level indicates intrahepatic cholestasis and is generally due to an infiltrative process (e.g., metastatic carcinoma, sarcoidosis) or drug reaction.

A predominant elevation in conjugated bilirubin points to extrahepatic cholestasis from biliary obstruction. Elevated bilirubin can result from either direct (conjugated) or indirect (unconjugated) causes. Although an overwhelming liver process in an adult will produce elevations of both, as a practical matter unconjugated hyperbilirubinemia is due to significant hemolysis. It is noteworthy that jaundice in patients with cirrhosis often shows elevations in all parameters, reflecting the widespread liver dysfunction and obstruction of the bile canals and small vessels caused by scarring.

Evaluation of hepatocellular and cholestatic disorders may include tests for viral hepatitis, various biochemical assays (see the specific disorders discussed later), or a needle biopsy of the liver. Needle biopsies are generally directed by ultrasound or computed tomography (CT), allowing focus on a specific target such as a mass lesion.

Ultrasonography is an important imaging study for significant liver disease. This is particularly true given the fact that structural liver abnormalities such as tumors, cysts, or abscesses may present with any of the aforementioned enzyme patterns. CT provides more detailed information of the parenchyma, but has less utility for the biliary tree itself. Specific visualization of the bile ducts traditionally has been carried

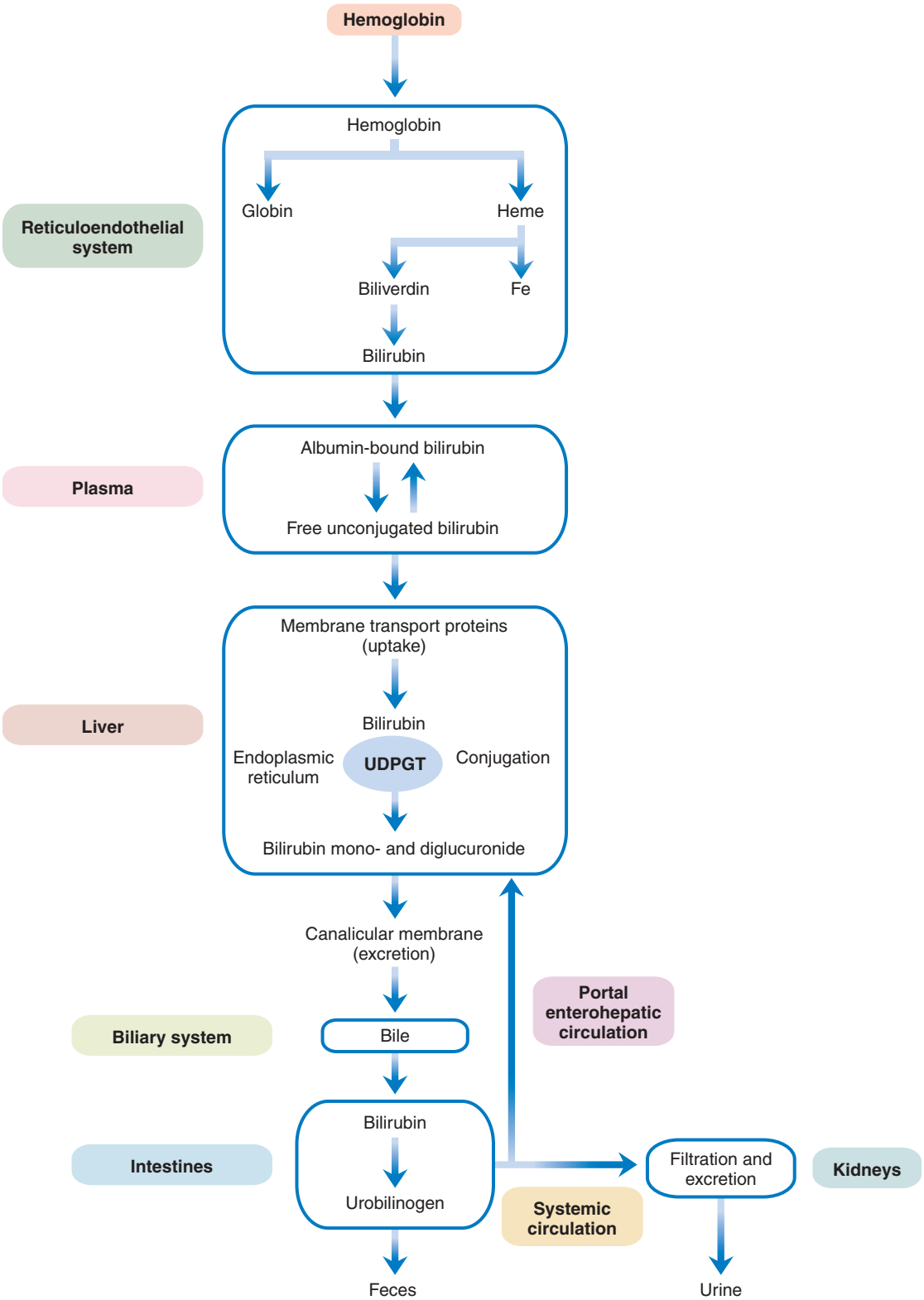


FIG 38.4 Summary of bilirubin metabolism (see text for explanation). *UDPGT,* Uridine diphosphate glucuronosyltransferase.

out with percutaneous transhepatic cholangiography or endoscopic retrograde cholangiopancreatography (ERCP; see Chapter 37), but magnetic resonance cholangiopancreatography (MRCP) has rapidly gained acceptance as a noninvasive way of evaluating the liver and biliary tree in great detail.

Portal Hypertension

Manifestations of liver disease not attributable to hepatocellular failure are mainly due to portal hypertension, a condition resulting from impaired blood flow through the liver as a result of increased resistance from fibrosis and degeneration of liver tissue (Fig. 38.5). In this condition,

TABLE 38.2 Common Laboratory Tests in Liver Disease

Test	Normal Range	Significance
AST/SGOT	5–40 units/mL	Elevated levels indicate hepatocellular inflammation or necrosis
ALT/SGPT	5–35 units/mL	AST much greater than ALT in alcoholic liver disease
		AST less specific; may be of skeletal muscle, myocardial, kidney, or liver origin
		ALT more specific for liver disease
Alkaline phosphatase	35–150 units/mL	Elevated in cholestasis, infiltrative liver disease (e.g., cancer, granulomas)
		May be of bone origin
γ-Glutamyltranspeptidase	10–48 units/mL	Elevated in cholestasis and hepatocellular disease
		Used to confirm that elevated alkaline phosphatase is of hepatic origin
		Disproportionately elevated in alcoholic liver disease
		May be induced by many drugs (e.g., phenobarbital)
5′-Nucleotidase	2–11 units/mL	Elevated in cholestasis
		Very specific to liver
Total bilirubin	<1.0 mg/dL	Elevated levels diagnose jaundice
Indirect bilirubin	<0.8 mg/dL	Elevated in hemolysis, Gilbert disease
Prothrombin time	11.5–14 sec	Prolongation suggests decreased hepatic synthetic function
Serum albumin	3.5–5.5 gm/dL	Decreased level suggests decreased hepatic synthesis
Serum globulin	2.5–3.5 gm/dL	Elevated in autoimmune hepatitis
Urine bilirubin	0	Increased with elevation of serum conjugated (direct) bilirubin, zero in unconjugated (indirect) hyperbilirubinemia
Urinary urobilinogen	0–4 mg/24 hr; spot test ± on urine dipstick	Zero in complete biliary or proximal bowel obstruction
		Increase may suggest liver disease
		Nonspecifically insensitive
		Primary utility because its presence on urine dipsticks allows simple office/bedside testing or screening with one-time urinalysis

venous drainage of much of the GI tract is congested. Symptoms are surprisingly few early in the course, but as abnormal vascular patterns progress, anorexia may result. The central consequence of chronic portal hypertension is varices, particularly esophageal, but also to a lesser extent gastric and hemorrhoidal. A pathognomic feature of advanced liver disease is superficial periumbilical varices, known as *caput medusae,* or the head of Medusa. Portal hypertension may present as abdominal swelling due to an accumulation of peritoneal fluid, or ascites. A catastrophic consequence of portal hypertension is uncontrolled bleeding from esophageal varices, which are thin walled and prone to rupture.

Gastroesophageal Varices

Etiology. Esophageal varices result from portal hypertension, which in Western society is generally the result of cirrhosis due to chronic alcoholism or viral hepatitis. In developing tropical countries, chronic infection with the *Schistosoma* species of liver fluke is a major cause of portal hypertension. Recently it has been recognized that vasoactive hormones, as well as increased splanchnic blood flow and increased vascular resistance in the liver, have a prominent role in contributing to the formation of variceal esophageal veins. Gastric varices may occur in conjunction with or independent from esophageal varices, the latter if splenic vein obstruction or thrombosis is the precipitant.

Pathogenesis. Gastroesophageal varices are one of a number of collateral venous pathways that dilate in response to elevated portal pressure in an attempt to transport blood from the splanchnic bed back to the heart. Other common collateral pathways include a variety of spontaneous deep and usually entirely asymptomatic portosystemic shunts, such as splenorenal shunts; dilated veins in the small intestine, colon, and rectum are also not uncommon. Unfortunately, part of the very complex venous network that surrounds the proximal part of the stomach and esophagus lies just beneath the mucosa, rendering it especially liable to rupture when portal pressures reach a critical level (Fig. 38.6). Rupture often results in massive, life-threatening upper GI bleeding.

Clinical features. Varices will affect more than half of cirrhotic patients, and approximately 30% of them will experience an episode of variceal hemorrhage within 2 years of the diagnosis. Variceal size is the main determinant of risk for bleeding, which is one of the main causes of death (20% to 33%) in persons with long-standing cirrhosis. The mortality after an episode of significant variceal bleeding is as high as 50%. The diagnosis is made mainly endoscopically, but varices may be seen on CT scans of the abdomen, as well as on upper GI barium examinations.

The initial symptoms and signs of bleeding from gastroesophageal varices include hematemesis, melena, and potentially even bright red rectal bleeding. These characteristics may be associated with profound anemia and symptoms and signs of shock. In most cases, concomitant evidence of chronic liver disease and portal hypertension is seen on physical and laboratory examination. There are two distinct phases of the variceal hemorrhage process: one, coincident with and shortly after the bleeding; and two, a period of 6 to 8 weeks after the initial bleed, when there is a high risk of rebleeding. The greatest risk of rebleeding occurs in the first 72 hours.

Treatment. Initial treatment is directed at performing fluid resuscitation, correcting the coagulopathy, and stopping further bleeding. Large-bore intravenous lines are placed, and fluid resuscitation is carried out with normal saline. Blood components and clotting factors are replaced as needed. Any significant coagulopathy would necessitate administration of parenteral vitamin K and fresh frozen plasma, and platelet infusions are indicated if profound thrombocytopenia is present. Recombinant factor VIIa is helpful in reversing the coagulopathy associated with advanced liver disease and may be considered for patients whose prothrombin time (PT) fails to normalize after the previously stated measures have been performed.

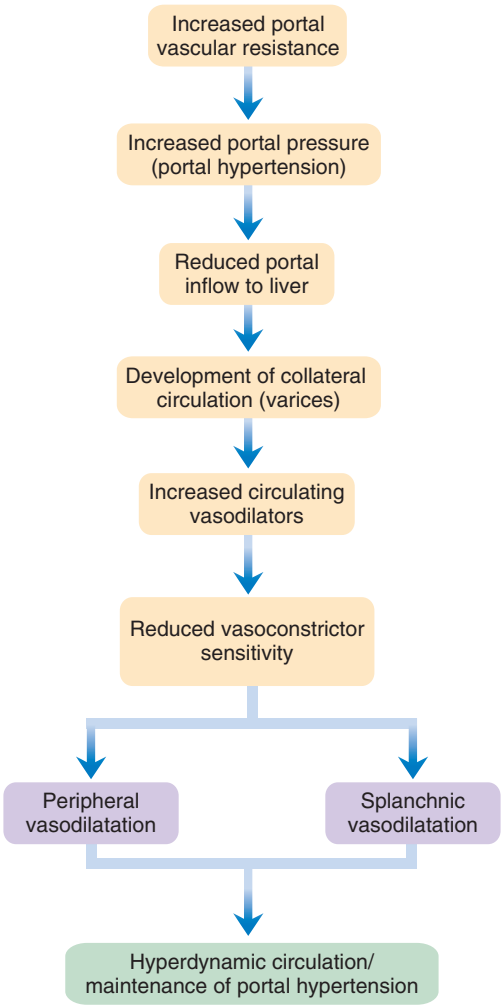


FIG 38.5 Pathophysiologic process of portal hypertension. (Redrawn from MacMathuna P: The pathogenesis of variceal rupture. *Gastrointest Endosc Clin North Am* 1992;2(1):1–8.)

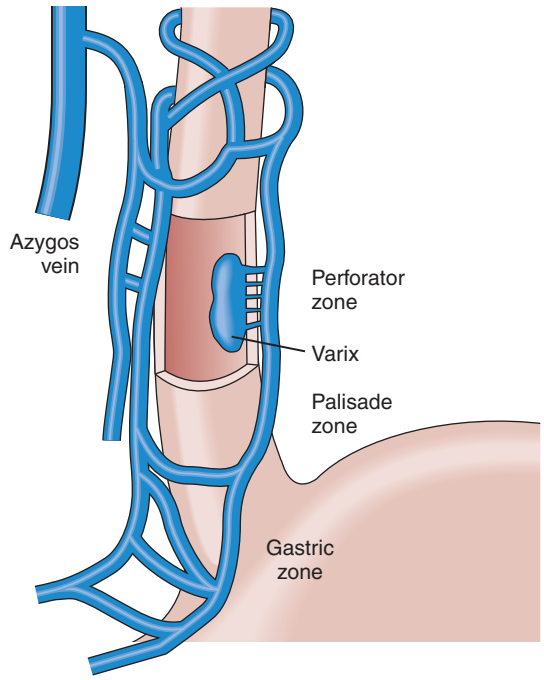


FIG 38.6 Gastroesophageal venous anatomy. (Redrawn from MacMathuna P: The pathogenesis of variceal rupture. *Gastrointest Endosc Clin North Am* 1992;2(1):1–8.)

Primary acute pharmacologic management rests on drugs that can effectively lower portal pressure by dilating alternative collateral pathways, reducing splanchnic blood flow, or both. Until recently, the agent of choice in the United States was vasopressin, an analog of antidiuretic hormone, administered by continuous intravenous infusion along with nitroglycerin. Although effective in controlling variceal bleeding, vasopressin use may be associated with angina pectoris, severe abdominal cramping, and hyponatremia, side effects that limit its usefulness.

In recent years, octreotide acetate, a synthetic analog of the naturally occurring hormone somatostatin, has been effectively used as a replacement for vasopressin. It is administered as an initial intravenous bolus followed by continuous infusion, which may be administered for as long as 3 to 5 days. In the doses used, the drug is remarkably free of side effects and more effective than vasopressin. However, it should be noted that no drug treatment has shown a mortality benefit for this condition.

Metoclopramide and β-blockers have been used as ancillary treatments in the past and may be considered for selected patients. Intravenous H_2-blockers or proton pump inhibitors are also often administered. The use of prophylactic antibiotics is recommended by the American Association for the Study of Liver Diseases; intravenous ceftriaxone or an oral quinolone for a 1-week course is the preferred regimen.

Emergency esophagogastroduodenoscopy (EGD) is crucial in determining the site of bleeding, as well as excluding other causes of upper GI bleeding. In addition to its diagnostic role, EGD actively addresses bleeding varices. *Endoscopic sclerosis* of esophageal varices is accomplished by passing a flexible needle through the gastroscope and injecting various sclerosant solutions into and around the bleeding varix (Fig. 38.7). Such treatment results in initial thrombosis of the vein with hemostasis. Repeated injections cause fibrosis and obliteration of the varix and fibrosis of the overlying mucosa. This process can effectively obliterate all of the varices at risk of bleeding. Though generally well tolerated, this treatment may be associated with a variety of acute and chronic complications, including drug reactions to the sclerosing solutions, exacerbation of bleeding, perforation, ulceration, infection, and stricture formation.

An alternative treatment method is endoscopic ligation of esophageal varices (Fig. 38.8). In this technique, a special apparatus is preloaded onto the gastroscope so that the endoscopist can suction a varix into a special chamber at the end of the gastroscope and then ligate the varix with a small rubber band. This technique also results in immediate loss of flow in the varix and eventually leads to thrombosis and fibrosis. The area ligated simply sloughs off over the next few weeks without significant residual ulceration or scarring. This method requires fewer sessions than endoscopic sclerosis to completely obliterate the varices, seems to be associated with a lower complication rate, and may be more effective in the management of bleeding gastric varices. However, it is technically more challenging and is generally reserved for elective use.

Endoscopic techniques have shown a mortality benefit relative to treatment with drugs alone but fail to control acute bleeding in 10% to 20% of patients. Unfortunately, both of these methods may result in an increase in venous pressure proximal to the area treated, perhaps resulting in bleeding from congestive gastroenteropathy, a diffuse venous congestion associated with portal hypertension.

Balloon tamponade of varices was widely used before the availability of endoscopic treatment. This treatment consists of a gastric balloon that is passed orally or transnasally into the stomach, inflated, and held in gentle traction against the gastric varices in the fundus, thus tamponading bleeding vessels and restricting blood flow from the fundus up into the esophageal varices (Fig. 38.9). Suction of oropharyngeal and gastric secretions is accomplished with integral or separate drainage tubes. Balloon tamponade carries major risks, including aspiration of stomach contents, migration of the tube with airway compression, pressure necrosis of the esophagus and stomach, rupture of the balloon, and rebleeding after the maximal inflation period of 24 hours. These limit the usefulness of balloon tamponade to a temporizing role until definitive treatment can be provided.

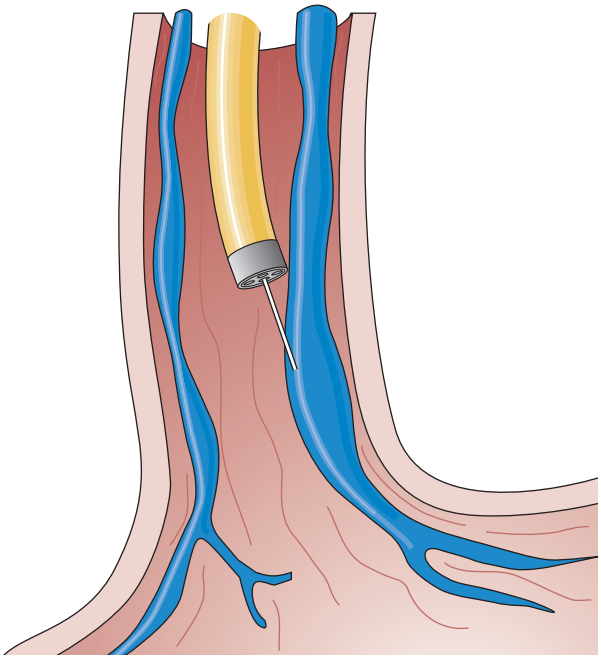


FIG 38.7 Endoscopic sclerosis of varices.

Chronic pharmacologic management of portal hypertension is frequently successful with nonselective β-blockers such as propranolol or nadolol. The drug is carefully titrated to reduce the initial resting heart rate by 25%. These drugs may be used prophylactically in patients with known portal hypertension and are often prescribed after endoscopic therapy as part of a combined approach to variceal bleeding prevention. It should be noted that studies have not yet shown a consistent survival benefit for this intervention, however. In addition to β-blockers, oral long-acting nitrates such as isosorbide mononitrate act synergistically with β-blockers to reduce portal pressure and have been studied for prevention of variceal hemorrhage. Side effects limit the use of nitrates as primary prophylaxis, but these agents could be considered for secondary prophylaxis in patients who do not respond to β-blockers alone.

If the aforementioned measures are ineffective, a number of surgical and radiologic procedures are possible. Although rarely used in the United States, esophageal transection and reanastomosis with ligation of other collateral channels have been used in other countries. Surgery to reduce portal pressure is effective in decreasing the rate of rebleeding, but it may not alter overall survival. A variety of surgical techniques are used, all of which create an alternative connection between the splanchnic and systemic circulations. These techniques include portacaval, mesocaval, splenorenal, and distal splenorenal shunts (Fig. 38.10). A discussion of the specific indications and technical aspects of these shunts is beyond the scope of this text, but each has certain specific indications, advantages, and disadvantages.

In recent years a radiographic procedure called *transjugular intrahepatic portosystemic shunting (TIPS)* has been developed that combines angiographic and ultrasonographic techniques. The hepatic vein is cannulated by the transjugular route. A needle is then passed into a main portal vein branch. Catheters are passed over this guidewire along with balloon dilation of the tract just created. This step is then followed by placement of an expandable metallic stent, thus creating a portosystemic shunt (portal vein to hepatic vein) within the liver itself. This procedure is technically demanding and may be complicated by hemorrhage, infection, stent migration, stent stenosis, and occlusion,

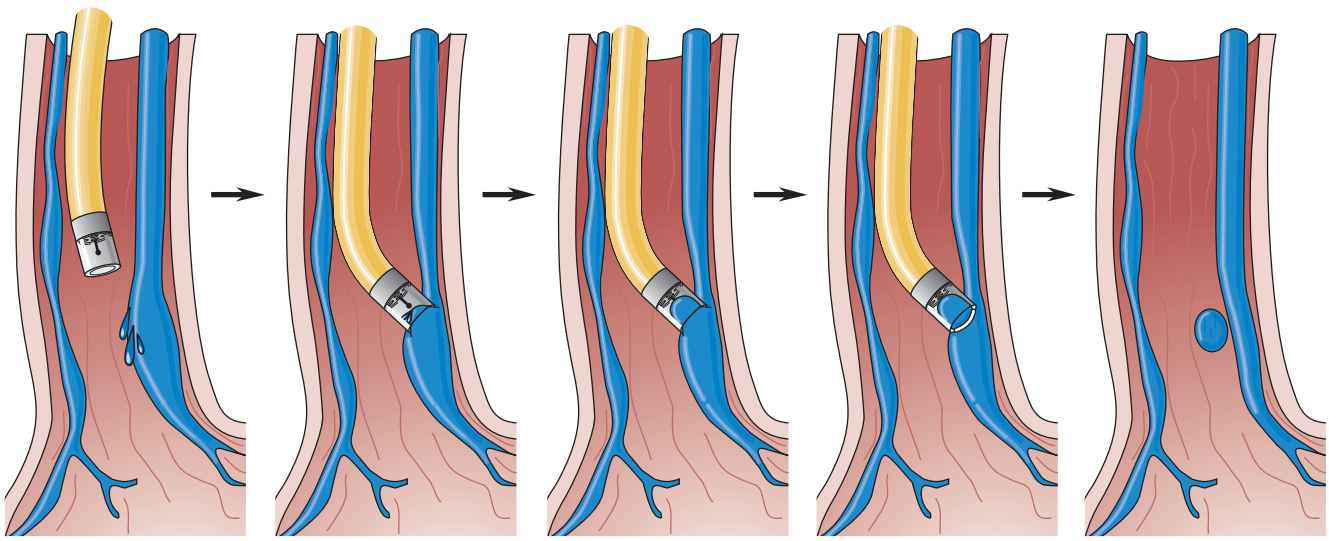


FIG 38.8 Endoscopic band ligation of varices.

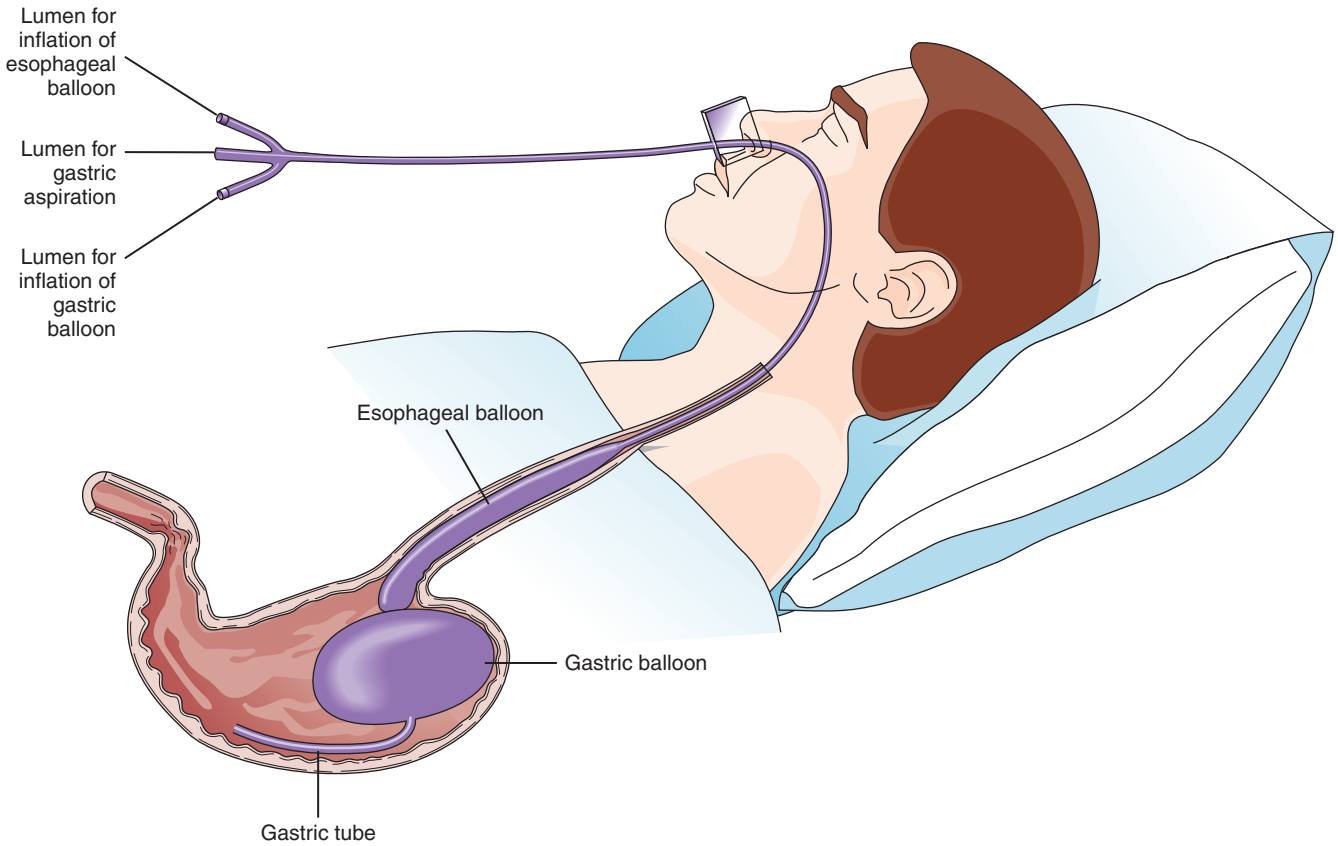


FIG 38.9 Sengstaken–Blakemore tube.

both acute and chronic. In addition, hepatic encephalopathy and congestive heart failure may result. The primary use of this modality is as a bridge to allow stabilization of patients who are candidates for liver transplantation.

Treatment of esophageal varices is often unsatisfactory. Ideally, the underlying condition for varices (i.e., portal hypertension) should be reversed. The only consistently effective way to accomplish this goal is by liver transplantation (see the Transplantation section), which is limited in its application to a select group of patients.

Portal Systemic Encephalopathy

Hepatic Encephalopathy

Pathogenesis. Hepatic encephalopathy is a complex neuropsychiatric syndrome characterized by symptoms ranging from mild confusion and lethargy to stupor and coma. Some patients exhibit dementia, psychotic symptoms, spastic myelopathy, and cerebellar or extrapyramidal signs. The classic physical finding is asterixis, or "liver flap," a spastic jerking of the hands held in forced extension. Hepatic encephalopathy is associated with fulminant hepatic failure and severe chronic liver disease, conditions in which liver function is severely depressed and blood is shunted around the liver. The arterial ammonia level correlates positively with the level of encephalopathy in most patients, consistent with its central role in the pathogenesis of hepatic encephalopathy as one of the primary causes of neuronal dysfunction. The exact cause is unclear, and other contributing factors such as elevated mercaptan levels, enhanced activation of certain neurotransmitter receptors (including γ-aminobutyric acid and benzodiazepine receptors), and elevated levels of aromatic amino acids (false neurotransmitters) remain under investigation.

Clinical manifestations. Hepatic encephalopathy is usually precipitated by certain well-defined clinical developments, including hypokalemia, hyponatremia, alkalosis, hypoxia, hypercarbia, infection, use of sedatives, GI hemorrhage, protein meal gorging, renal failure, and constipation. In some patients, progressive liver failure leads to chronic encephalopathy without other exacerbating factors.

Hepatic encephalopathy is graded 1 to 4:

Grade 1: Confusion, subtle behavioral changes, no flap
Grade 2: Drowsy, clear behavioral changes, flap present
Grade 3: Stuporous but can follow commands, marked confusion, slurred speech, flap present
Grade 4: Coma, no flap

Treatment. The first step in treatment of hepatic encephalopathy consists of correcting any identifiable precipitating factors, such as GI bleeding. Enhanced elimination of the toxic nitrogenous substances produced by intestinal digestion is indicated for all patients with chronic encephalopathy. Although intake of dietary proteins should be generally targeted at 1.2 to 1.5 g/kg daily, extreme protein restriction is not recommended due to the generally malnourished state of persons with chronic liver disease. Critically ill patients should receive peripheral or central glucose infusions along with vitamins, especially thiamine. As the patient's ammonia levels drop, protein may be reintroduced into the diet. The initial amount of 20 g/day is increased by 10 or 20 g/day every few days to the ultimate target level. Observation for worsening encephalopathy is crucial at this time. When protein is restricted, it is essential to provide at least 400 g of carbohydrate daily. Vegetable protein may be better tolerated than animal protein. High dietary fiber intake may help by decreasing constipation. If dietary measures fail, oral defined-formula feedings containing essential amino acids and enriched with branched-chain amino acids may be indicated.

Osmotic diuretics or antibiotics are used to enhance elimination of nitrogenous wastes. Lactulose and lactitol are standard nonabsorbable

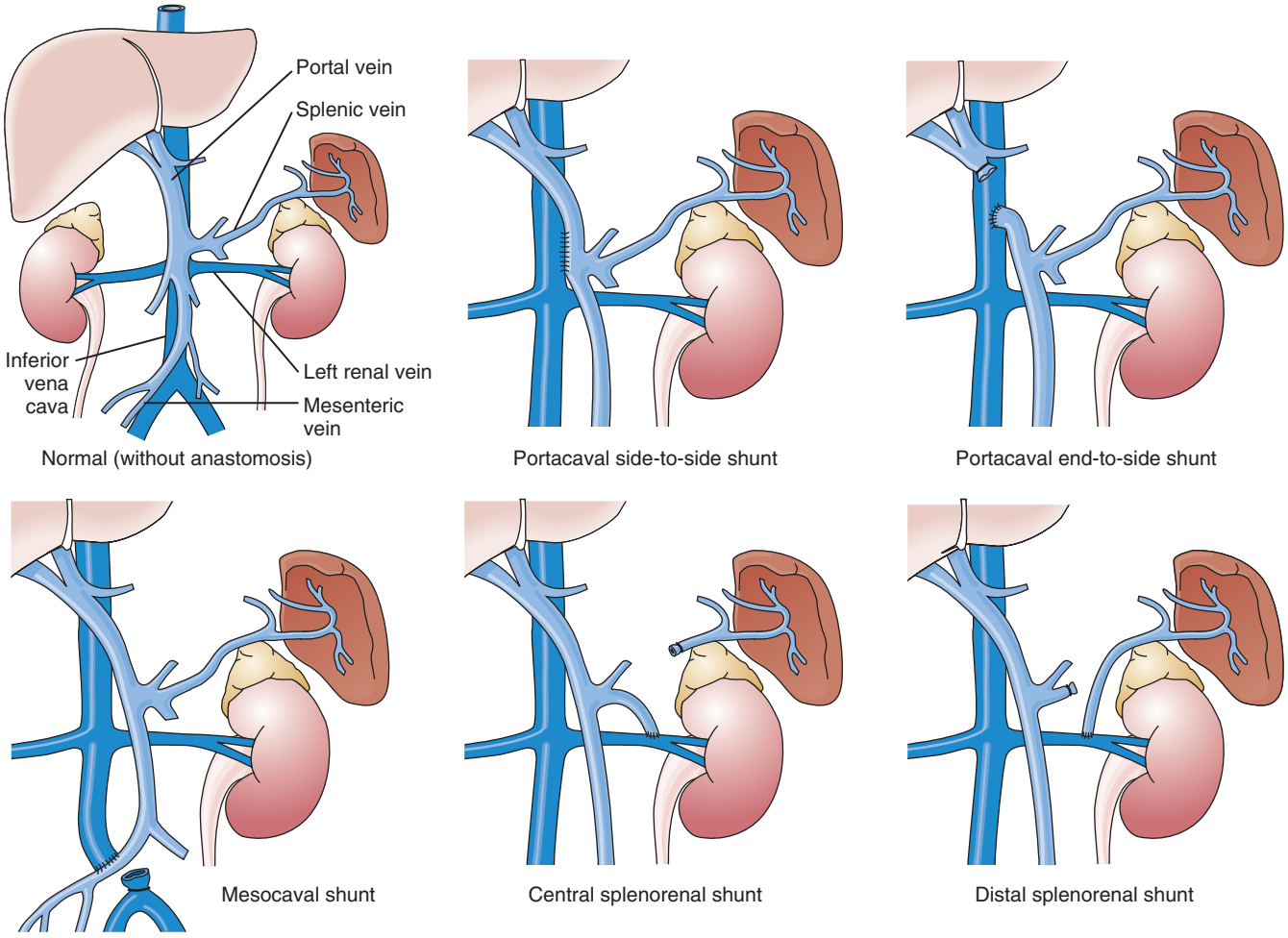


FIG 38.10 Portosystemic shunt operations.

osmotic cathartics and may be given orally or rectally by enema. (Standard precautions must be taken before any cathartic is administered, including ruling out bowel obstruction and monitoring electrolyte levels, particularly in patients with renal insufficiency.) Although the exact mechanism is unclear, evidence suggests that a cathartic-related change in pH inhibits ammonia production by gut flora, possibly by selecting for bacterial populations that are less ammoniagenic. Serious adverse reactions are not usually reported with lactulose therapy, although flatulence and abdominal cramping may occur. The dosage should be individually titrated so that two soft, acidic stools are passed daily.

Oral antibiotics have been used for many years to suppress the intestinal flora that break down dietary protein and release ammonia. Rifaximin is the current antibiotic of choice used for this purpose. The cost of this medication may be prohibitive for many patients. Neomycin, the first antibiotic widely used in hepatic encephalopathy, is no longer recommended because of side effects.

Cerebral Edema

Pathogenesis. Swelling of the brain (cerebral edema) is a serious complication in patients with grade 3 or 4 hepatic encephalopathy, which results in an increase in intracranial pressure. Both vascular and toxic mechanisms have been implicated as etiologic factors. With increasing intracranial pressure, blood perfusion of the brain is decreased (cerebral perfusion pressure = carotid artery pressure − intracranial pressure), with resulting cerebral hypoxia. Cerebral edema is a major cause of death in patients with acute hepatic failure.

Clinical manifestations. Clinically, cerebral edema is characterized by deepening coma, systolic hypertension, and extensor rigidity (decerebrate posture), followed by pupillary dilation and eventually respiratory arrest as the brainstem herniates. Some highly specialized referral centers monitor patients with advanced hepatic encephalopathy using extradural pressure monitors to permit early detection. Unfortunately, complications of extradural monitors occur up to 20% of the time and include infection and intracranial bleeding.

Treatment. Cerebral edema is managed primarily by the intravenous infusion of mannitol, which by increasing serum osmolarity draws water from the brain and thus reduces the swelling. Patients should be kept in the semi-Fowler position (head and trunk elevated 30 degrees). Barbiturate-induced coma utilizing sodium pentothal is used as a second-line treatment for patients with recalcitrant intracranial hypertension and those who cannot tolerate the fluid volume component of mannitol therapy (e.g., those in heart or kidney failure). Moderate hypothermia with the use of cooling blankets has benefit in patients failing conventional therapies as a bridge to transplantation. Aggressive treatment allows patient survival in 60% of cases, until liver failure resolves or liver transplantation can be accomplished.

Complications of Advanced Liver Disease

Ascites

Etiology, pathogenesis, and clinical manifestations. Ascites, or the pathologic accumulation of fluid in the peritoneal cavity, is a typical concomitant of advanced liver disease complicated by portal hypertension

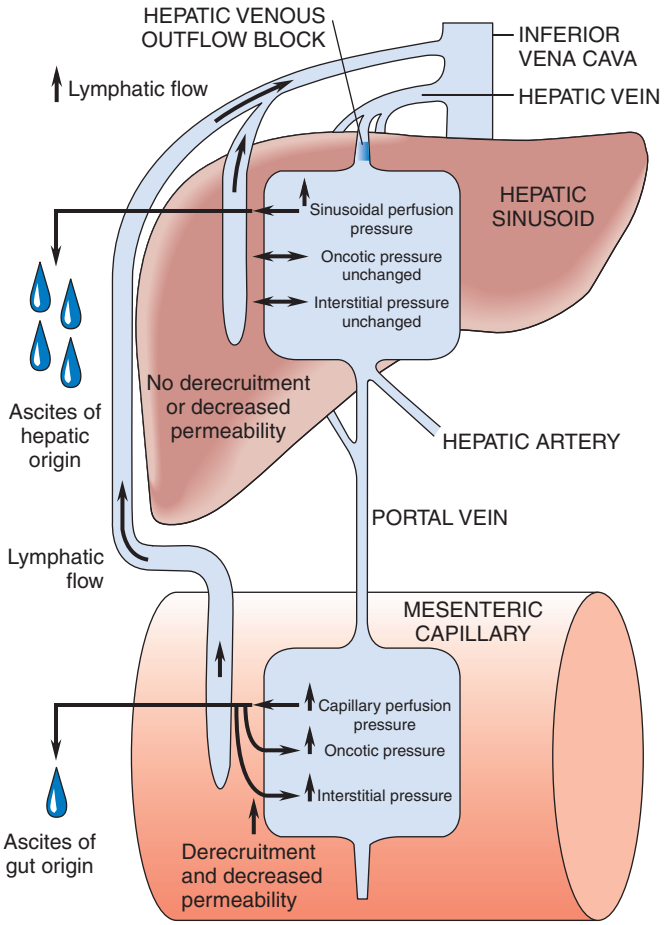


FIG 38.11 Pathophysiologic process of ascites. (Redrawn from Dudley FJ: Pathophysiology of ascites formation. *Gastroenterol Clin North Am* 1992;21:215–235.)

and plasma hypoalbuminemia (Fig. 38.11). Abdominal distention results from an inappropriate pressure gradient across the pleura, with the intraabdominal accumulation of sodium, water, and protein. Other less common causes of ascites are malignancy, infection, pancreatitis, hypothyroidism, vasculitis, nephrosis, cardiac failure, constrictive pericarditis, Budd–Chiari syndrome, and portal vein thrombosis. The specific chemical and cellular composition of the ascites varies with its cause.

Abdominal paracentesis should be performed in all patients with new ascites and in those with known ascites who have experienced significant worsening of their condition to identify treatable causes and to rule out malignancy. Appropriate fluid studies include total protein level, albumin level, cell count, and cytology. Optional tests include culture for bacteria, fungi, and mycobacteria and measurement of amylase, glucose, and lactate dehydrogenase levels. A small amount of ascites may not require specific therapy. However, with increasing volumes, abdominal discomfort, abdominal or umbilical herniation, respiratory embarrassment, or infection may occur.

Treatment. Dietary sodium should be restricted to 88 mEq (2000 mg) per day in patients with ascites. In motivated patients, this is perhaps the most helpful intervention that can be undertaken. Bed rest is useful, although strict bed rest can result in deconditioning and decubitus ulcers, among other problems. Diuretics are necessary for the majority of patients. The aldosterone antagonist spironolactone works in the distal nephron as a weak diuretic that also inhibits potassium secretion, thus sparing serum potassium. A delay of 2 to 3 days may occur before the full effect of the drug is seen, and the dosage should not be increased more frequently. Many authorities suggest adding a loop diuretic such as furosemide from the beginning and increasing the dose at a ratio of 4:10 with spironolactone. It is helpful to monitor both urinary sodium and potassium levels periodically. When the urinary sodium level exceeds the urinary potassium level, spironolactone is exerting its maximal effect. Serum potassium levels must be carefully controlled.

The goal is the loss of approximately 0.5 kg of body weight daily, although in patients with peripheral edema, slightly more rapid weight loss is tolerable. More rapid losses may result in intravascular volume depletion and severe electrolyte abnormalities, as well as precipitation of hepatorenal syndrome or hepatic encephalopathy. Diuresis should be continued until the ascites is barely detectable. Free water restriction is prescribed if hyponatremia is present or develops during treatment, although compliance with a strict regimen is challenging.

In patients who do not respond to both diuretic therapy and sodium restriction, the use of 25% albumin infusions may help initiate diuresis. However, the effect of this treatment is often short lived and carries a risk of intravascular volume overexpansion, with the possibility of pulmonary edema and precipitation of variceal hemorrhage. Alternatively, large volumes of ascitic fluid can be removed from the peritoneal space through paracentesis. This "large-volume therapeutic paracentesis" is a rapid and effective treatment that can be safely instituted if the intravascular volume is maintained by appropriate measures. Diuretic and dietary treatment should be continued, but in severe cases, large-volume paracentesis may be repeated as needed.

Shunting procedures such as the LeVeen (Fig. 38.12) and Denver shunts have been used. These one-way valves connect the peritoneal space with the venous system via the jugular vein. Such shunting procedures, although useful, carry significant risks and are best reserved for a small subset of patients who have refractory ascites and are not candidates for liver transplantation. The use of TIPS has largely replaced traditional surgical procedures for treatment of refractory ascites (see earlier discussion). Diuretic-resistant ascites is a key indication for liver transplantation in the patient who meets other criteria.

Spontaneous Bacterial Peritonitis

Pathogenesis and clinical manifestations. Patients with cirrhosis and ascites suffer from a variety of defects in host defense predisposing to infection in the peritoneal cavity. This observation is especially true in patients with ascites with a low protein concentration. These defects include diminished opsonic activity of the ascitic fluid, diminished reticuloendothelial function, and transmigration of gut bacteria across the intestinal wall and into the ascites. Chronic alcoholics demonstrate abnormal white blood cell function as well (Fig. 38.13).

Typically, patients with spontaneous bacterial peritonitis have a single infecting organism of gut origin in the fluid. *Streptococcus pneumoniae* is most common. This pattern is in marked distinction to patients with secondary infection of the ascites, as may occur after traumatic gut perforation, for instance, in which polymicrobial infection is typical. The onset of spontaneous bacterial peritonitis may be subtle, with only mild abdominal discomfort or general clinical deterioration, including worsening hepatic encephalopathy, renal failure, altered mental status, or a nonspecific "septic" picture. Significant fever may be lacking. The most important diagnostic study is abdominal paracentesis. An ascitic polymorphonuclear leukocyte count greater than 250/mm^3 suggests the diagnosis, although it should be noted that a low count does not rule out the diagnosis and bacterial culture remains the gold standard.

Treatment. Antimicrobial therapy should be initiated promptly in suspected cases of spontaneous bacterial peritonitis pending culture results. Third-generation cephalosporins and quinolone antibiotics are effective empirical therapy. Use of long-term oral antibiotics for

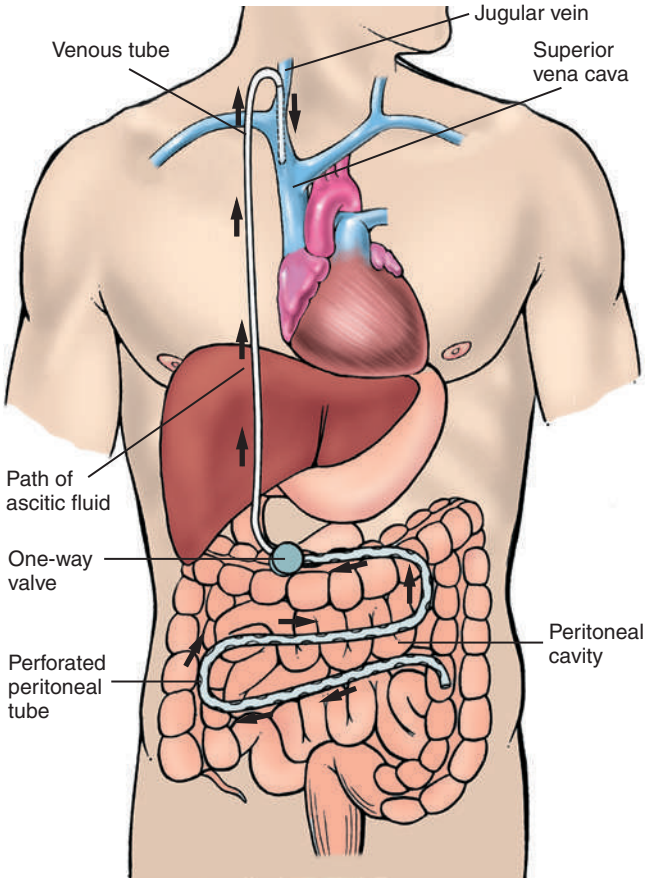


FIG 38.12 LeVeen shunt. Arrows show direction of flow of ascitic fluid out of the peritoneal cavity, through the shunt, into the superior vena cava. (From Monahan FD, Neighbors M: *Medical-surgical nursing: foundations for clinical practice,* ed 2, Philadelphia, 1998, Saunders, p 1146.)

prophylaxis in at-risk patients with cirrhosis may be considered for selected patients. Use of a narrow-spectrum agent such as trimethoprim-sulfamethoxazole (TMP-SMX) therapy is preferable to broader-spectrum agents such as the quinolone class. Overall, the occurrence of spontaneous bacterial peritonitis is a poor prognostic sign, as it is typically associated with end-stage liver disease.

Hepatorenal Syndrome

Etiology and pathogenesis. Hepatorenal syndrome is a dire complication of patients with liver failure. Such patients may experience acute kidney failure with rising serum creatinine levels and oliguria. The kidney itself remains histologically normal, but intrarenal blood flow is seriously disturbed. This disturbance in blood flow appears to be due to an imbalance between vasoconstricting and vasodilating mechanisms related to the liver disease. The onset is usually acute and progressive, but chronic cases are occasionally seen. Differentiation from reversible causes of renal failure (e.g., prerenal azotemia) is essential. Hepatorenal syndrome may be precipitated by overly vigorous diuretic therapy or paracentesis, severe diarrhea, nonsteroidal antiinflammatory drugs, variceal bleeding, and sepsis, among other conditions.

Prognosis and treatment. The prognosis of hepatorenal syndrome depends on the severity of the underlying liver disease and is generally poor. Treatment is primarily preventive and supportive, and hemodialysis should be considered only as a bridge to definitive therapy. Fortunately, among transplant patients, liver transplantation generally results in return of normal renal function, providing no undue delay has occurred.

KEY POINTS

- The liver is a vital, multifunctional organ located in the right upper quadrant beneath the diaphragm. Blood is supplied to the liver by the hepatic artery and the portal vein. The portal vein drains the capillaries of the alimentary canal and pancreas. Arterial and portal blood flows into the hepatic sinusoids, which have direct contact with hepatic cells.

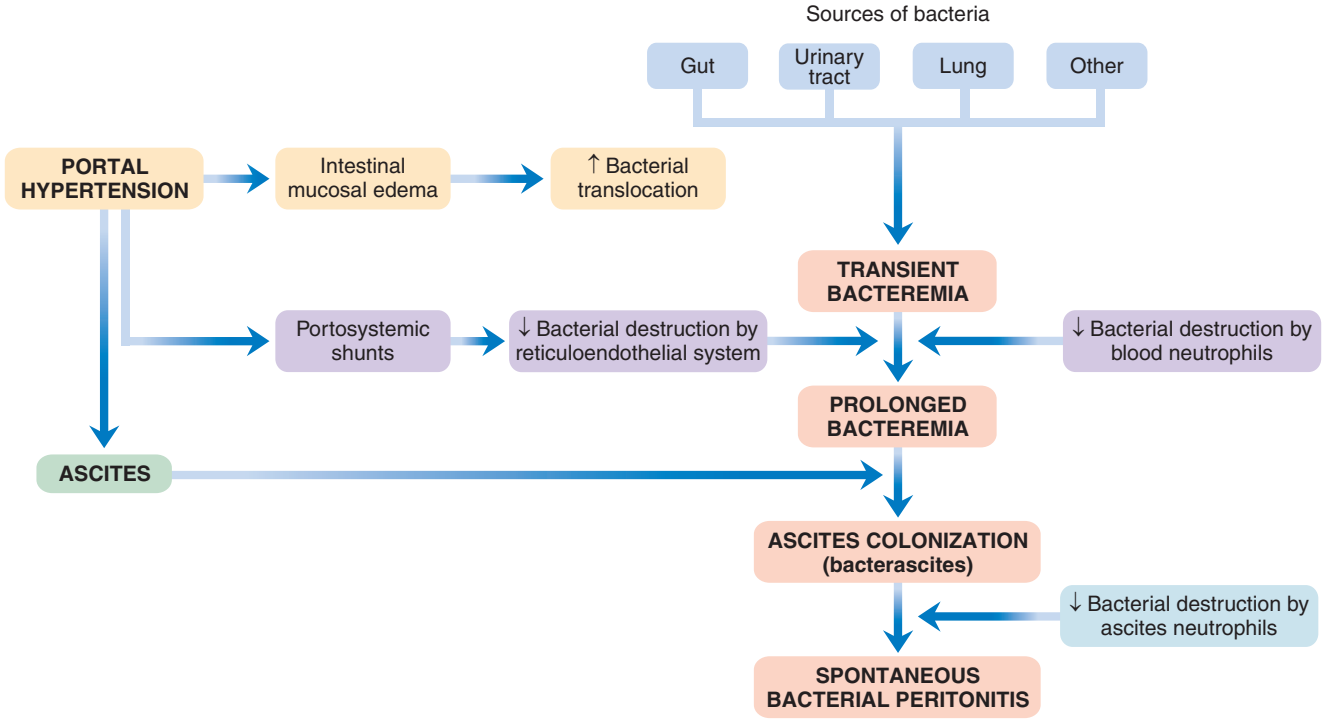


FIG 38.13 Pathophysiologic process of spontaneous bacterial peritonitis. (Redrawn from Garcia-Tsao G: Spontaneous bacterial peritonitis. *Gastroenterol Clin North Am* 1992;21:257–275.)

- The functions of the liver are multiple and include metabolism of fats, proteins, and glucose; synthesis and secretion of bile salts; storage of vitamins and minerals; metabolism and detoxification of endogenous and exogenous substances; and synthesis of urea.
- Manifestations of liver disease are attributable to hepatocellular failure and portal hypertension. Jaundice, decreased levels of clotting factors, hypoalbuminemia, decreased levels of vitamins D and K, and feminization are attributed to hepatocellular failure. Portal hypertension may result in GI congestion with the development of esophageal or gastric varices, hemorrhoids, splenomegaly, and ascites. Bleeding from varices is often massive and life threatening; various pharmacologic, endoscopic, and surgical treatments are now available.
- Symptoms of hepatic encephalopathy range from confusion and lethargy to coma. A spastic flapping tremor of the hands, called *asterixis,* is a classic finding. The severity of the encephalopathy correlates positively with serum ammonia levels. Encephalopathy may be precipitated by conditions that increase protein metabolism, such as GI hemorrhage and increased protein consumption, and by conditions that further impair hepatocyte function. Treatment may include restriction of protein intake, administration of antibiotics to reduce ammonia production by intestinal organisms, and utilization of lactulose to enhance ammonia excretion in the stool. Cerebral edema is a common cause of death in patients in deep hepatic coma.
- Ascites is a pathologic accumulation of fluid in the peritoneal cavity. It occurs commonly in liver disease because of the increased fluid transudation that occurs with portal hypertension and hypoalbuminemia. In severe ascites, treatment may be instituted to ameliorate pain and respiratory difficulty. Sodium restriction, diuretics, and intermittent paracentesis are commonly prescribed. Surgical shunting procedures that allow accumulated peritoneal fluid to flow back into the circulation through a one-way valve may be effective.
- Spontaneous bacterial peritonitis is infection of ascites by a single organism unrelated to bowel perforation or surgical procedures. Antibiotic therapy alone is usually curative.
- Hepatorenal syndrome is a type of functional renal failure caused by severe liver disease. The prognosis is poor and is contingent on the outcome of the liver disease.

DISORDERS OF THE LIVER

Hepatitis

Acute Viral Hepatitis

Hepatitis is inflammation of the liver parenchyma, often due to infection. The liver is susceptible to a variety of infectious agents, especially viruses. Serologic testing is necessary to differentiate various types and stages of viral hepatitis (Table 38.3). The term *viral hepatitis* is usually applied to illnesses caused by hepatitis A, hepatitis B, and hepatitis C viruses (Table 38.4). A fourth virus, known as *hepatitis D* or *delta agent,* is a defective RNA virus that requires coinfection with hepatitis B virus to be pathogenic. Hepatitis E virus is a relatively recently described agent common in developing countries, but uncommon in the United States.

Despite variation in the symptoms, signs, and epidemiologic progression of these diseases, it is generally not possible to differentiate them in a given patient without appropriate serologic tests (Fig. 38.14).

Hepatitis A

Pathogenesis and clinical manifestations. Hepatitis A virus (HAV) is an RNA virus that is spread by enteric means (i.e., by the fecal-oral route). The infection has a 2- to 7-week incubation period (see Fig. 38.14). The illness may be asymptomatic or mildly symptomatic without jaundice (anicteric); the latter occurs especially in children, with the patient exhibiting nonspecific GI symptoms. The majority of adults develop hepatitis with jaundice. The prodromal symptoms of icteric hepatitis consist of malaise, anorexia, nausea, low-grade fever, and right upper quadrant pain. This is followed by jaundice lasting 2 weeks on average. The clinical course is generally self-limited, although fulminant and fatal attacks occur rarely, particularly in patients with preexisting chronic active hepatitis B or C infection. Two uncommon prolonged syndromes are recognized: prolonged cholestasis and relapsing hepatitis.

Diagnosis, treatment, and prevention. HAV infection is diagnosed through serologic testing. The presence of anti-HAV immunoglobulin G (IgG) indicates previous infection, and the presence of immunoglobulin M (IgM) indicates acute infection. The test is highly reliable within several weeks of exposure.

Treatment does not change the course of acute HAV infection. Supportive management includes rest and a nutritious diet. Alcohol, acetaminophen, and other potential hepatotoxins should be avoided. HAV is a common infection-control concern, particularly in the community setting. The usual fecal-oral precautions, such as careful handwashing, segregation, and cleaning of laundry and personal items, should be undertaken by patients and contacts.

Active immunization is currently recommended for all children at age 12 to 23 months and is indicated for adults in risk groups (e.g., foreign travelers, persons with chronic active hepatitis B or C, persons in institutions) using an inactivated whole-virus vaccine that is highly immunogenic and effective in preventing acute hepatitis A. An intramuscular dose is followed by a booster 6 to 12 months later, providing lifelong immunity in at least 98% of recipients. The vaccine is ideally administered at least 2 weeks preexposure (e.g., before travel to a developing country), but can also be administered in the setting of a community outbreak. It is effective even among persons with advanced chronic liver disease. Hepatitis A vaccination was added to the routine childhood vaccination schedule by the Advisory Committee on Immunization Practices of the Centers for Disease Control and Prevention (CDC) in 2006.

Patients exposed to HAV who are anti–HAV antibody negative should receive passive immunization with pooled human immunoglobulin in addition to active vaccination. An intramuscular dose should be given within 2 weeks of exposure. Passive immunity lasts 4 to 6 months and does not diminish vaccine effectiveness. However, subclinical or symptomatic but attenuated infections may develop in some exposed patients.

Hepatitis B

Pathogenesis and clinical manifestations. Hepatitis B virus (HBV) is a partially double-stranded DNA virus that is highly prevalent worldwide. Probably 300 million persons, or 5% of the world population, have chronic HBV infection. Chronic infection in the United States affects approximately 1 to 1.25 million people, most of them immigrants from endemic countries. In contrast to HAV, HBV is spread by parenteral contact with infected blood or blood products, including contaminated needles, and by sexual contact. Perinatal infection is a major route in endemic (mainly developing) countries. Other risk factors for HBV infection include working in a health care setting (3% of cases in the United States), undergoing transfusions and dialysis (1% each), having acupuncture treatments, tattooing, traveling for extended time overseas, and residing in an institution.

HBV has an incubation period of 2 to 6 months. The prodrome of HBV infection is often longer and more insidious than that of HAV infection and may involve a variety of immune complex–related phenomena, including urticarial (i.e., hives) and other rashes, arthralgia and arthritis, angioedema, serum sickness, and glomerulonephritis.

TABLE 38.3 Immunologic Markers in Viral Hepatitis

Marker	Description
Hepatitis A	
Anti-HAV IgM	Acute infection with HAV, but may persist for months
Anti-HAV IgG	Past infection with HAV
	Implies immunity to the virus
Hepatitis B	
Hepatitis B surface antigen (HBsAg)	Surface protein coat of HBV
	Implies active infection
	Detectable 2–6 wk after infection
	Remains present as long as infection is active
Hepatitis B surface antibody IgG (HBsAb IgG)	Antibody to surface protein of HBV
	Detectable shortly after or with clearance of HBsAg
	Implies resolution of infection and immunity to HBV
Hepatitis B core antibody IgM (HBcAb IgM)	Antibody to inner core protein of HBV
	Detectable 3–5 wk after infection
	Implies recent infection
Hepatitis B core antibody IgG (HBcAb IgG)	Same as for HBcAb IgM, but implies past infection
Hepatitis Be antigen (HBeAg)	Soluble fraction of HBV
	Detectable 2–6 wk after infection
	Implies ongoing infection with high infectivity
	May resolve independently of HBsAg
Hepatitis Be antibody (HBeAb)	Antibody to soluble fraction of HBV
	Detectable when HBeAg clears
	Implies decreased infectivity
HBV DNA polymerase activity	Same significance as HBV DNA
Hepatitis C	
Anti-HCV	Antibody to HCV antigens
	May not be detectable early in infection
	Does not indicate immunity to the virus
	Rapidly evolving area; many commercially available assays of different sensitivity and specificity
	Many false-positive and false-negative results
HCV RNA by PCR	Assay for level of viremia
	Correlates positively with activity of infection
	Clears with resolution of infection
	Technically difficult; available through reference laboratories
Hepatitis D (Delta)	
Hepatitis delta antigen (HDAg)	Assay for 35-nm RNA virus
	Detectable 2–10 wk after infection
	Implies early infection
Anti-HDV	Implies past or chronic infection
	Does not indicate immunity to the virus
Hepatitis E	
HEV RNA by PCR	Detects presence of virus
HEV antibody IgG and IgM	Not well standardized
Hepatitis G	
HGV by PCR	Measures level of viremia

PCR, Polymerase chain reaction.

Severity of illness ranges from no symptoms to moderate illness to fulminant hepatitis (1% of cases). The jaundice phase for most HBV infections is similar in degree and duration to that of HAV infections, although serious extrahepatic illness occurs more frequently.

Diagnosis. The serologic diagnosis is more complicated than with hepatitis A (see Fig. 38.14). A typical screening panel for HBV infection includes surface antigen (HBsAg), surface antibody (HBsAb), core antigen (HBcAg), and core antibody (HBcAb). In brief, with acute infection HBV core antigen (HBcAg) appears first, followed by seroconversion to core antibody (HBcAb). Presence of HBV surface antigen (HBsAg) shows up early and may persist, indicating active infection; development of surface antibody (HBsAb) points to resolution and immunity. It

TABLE 38.4 Comparison of Hepatitis Viruses

Virus	Hepatitis A	Hepatitis B	Hepatitis C	Hepatitis D	Hepatitis E
Type of virus	ssRNA	Partially dsDNA	ssRNA	Circular defective ssRNA	ssRNA
Viral family	Hepatovirus; related to picornavirus	Hepadnavirus	Flaviviridae	Subviral particle in Deltaviridae family	Calicivirus
Route of transmission	Fecal-oral (contaminated food or water)	Parenteral, sexual contact, perinatal	Parenteral; intranasal cocaine use is a risk factor	Parenteral	Fecal-oral
Mean incubation period	2–4 weeks	1–4 months	7–8 weeks	Same as HBV	4–5 weeks
Frequency of chronic liver disease	Never	10%	≈80%	5% (coinfection); ≤70% for superinfection	Never
Diagnosis	Detection of serum IgM antibodies	Detection of HBsAg or antibody to HBcAg	PCR for HCV RNA; third-generation ELISA for antibody detection	Detection of IgM and IgG antibodies; HDV RNA serum; HDAg in liver	PCR for HEV RNA; detection of serum IgM and IgG antibodies

From Washington K: Inflammatory and infectious diseases of the liver. In Kumar V et al, editors: *Robbins basic pathology,* ed 8, Philadelphia, 2007, Saunders, p 640.

dsDNA, Double-stranded DNA; *ELISA,* enzyme-linked immunosorbent assay; *HBcAg,* hepatitis B core antigen; *HBsAg,* hepatitis B surface antigen; *HBV,* hepatitis B virus; *HCV,* hepatitis C virus; *HDAg,* hepatitis D antigen; *HDV,* hepatitis D virus; *HEV,* hepatitis E virus; *PCR,* polymerase chain reaction; *ssRNA,* single-stranded RNA.

should be noted that conversion from surface antigen to surface antibody positivity can take as long as 1 year after acute infection, so treatment should not be considered immediately after infection.

Chronic infection is indicated by persistent HBsAg positivity. In this setting, the detection of hepatitis B e antigen (HBeAg) in the serum is associated with viral replication and infectivity, whereas seroconversion to hepatitis B envelope antibody (HBeAb) indicates minimal replication and infective potential.

There are two cardinal features of chronic viral liver infection: ongoing liver inflammation and active viral replication. Ongoing liver damage may be deduced from persistently elevated levels of liver enzymes and confirmed by liver biopsy. Viral replication can be measured by molecular testing (HBV DNA by quantitative polymerase chain reaction) and is virtually always associated with a positive HBeAg. Persons who have a detectable virus and are HBeAg positive can readily transmit the virus to their contacts and should be counseled regarding appropriate sexual and blood exposure precautions (including no blood donation). Patients with chronic HBV and cirrhosis are at risk for development of hepatocellular carcinoma (HCC), and periodic screening (e.g., at 6- to 12-month intervals) with ultrasound and α-fetoprotein determinations is recommended. Screening for other bloodborne pathogens such as HIV and hepatitis C virus (HCV) is recommended.

Treatment. Fulminant hepatitis is a life-threatening illness with high mortality. Care for patients with acute hepatitis is largely supportive, including aggressive treatment for coagulopathy, encephalopathy, cerebral edema, and other manifestations. Liver transplantation is the only definitive treatment for progressive liver failure in acute hepatitis.

Most nonfulminant HBV infections resolve spontaneously, although about 5% of acute hepatitis cases progress to chronic infection. Management of acute HBV infection is similar to that of HAV infection in terms of supportive care.

Management of chronic HBV infection has advanced dramatically over the last decade. End points are defined as elimination of hepatitis B DNA from the blood and conversion from e antigen positivity to envelope antibody positivity. Currently available therapies are interferon-α, lamivudine, adefovir, entecavir, tenofovir, and telbivudine. Interferon-α was the first effective treatment available and may be given daily or weekly (in the pegylated form) for 24 to 48 weeks with a response rate of about 33%. A significant proportion of patients experience a flare-up of acute hepatitis, and therefore this treatment cannot be used for patients with advanced liver disease (e.g., cirrhosis) or significant comorbidities. It is currently not used as widely as other therapies and is reserved for select groups such as young patients with well-compensated liver disease who do not wish to be on long-term treatment or are planning to be pregnant in the near future.

Lamivudine, emtricitabine, adefovir, entecavir, tenofovir, and telbivudine are oral nucleoside analogs that are generally given for several years. Lamivudine has the advantage of being significantly cheaper than the other options and is safe during pregnancy, but increasing drug resistance limits its role as primary therapy. Telbivudine has a higher durable response rate, but it is much more expensive and cannot be used for lamivudine-resistant strains of HBV. Entecavir is the treatment of choice for most patients, with a response rate of about 67% and e antibody seroconversion rate of 21%. It can be used in patients with cirrhosis as well. The drawback is the cost, as the treatment is quite expensive. Adefovir has significant nephrotoxicity, and its use is reserved for patients with resistant strains, though tenofovir has greater activity and has generally supplanted adefovir in such cases.

Treatment of patients coinfected with HIV is more complicated. Early treatment of HIV with an appropriate antiretroviral agent should include one or more of the nucleoside analogs noted previously (i.e., tenofovir, lamivudine, emtricitabine) that has activity against both viruses. If HIV treatment is not deemed necessary at the time HBV treatment is begun, adefovir or pegylated interferon can be used, because these will not induce HIV resistance.

Patients with chronic active hepatitis B infection are at risk for fulminant hepatitis with a superimposed viral hepatitis infection and should be vaccinated against hepatitis A using the killed, two-dose vaccine. They are also at risk for fulminant hepatitis or more rapid progression of liver disease if they contract hepatitis D virus (see Hepatitis D [Delta] section), so they should avoid risk factors for this virus such as injection drug use and unprotected sex.

Prevention. The HBV vaccine is a recombinant vaccine that is highly immunogenic with no material of human origin. (Thus there is no risk of transmission of other agents such as HIV.) Children are vaccinated intramuscularly with three doses of HBV vaccine given at birth and at

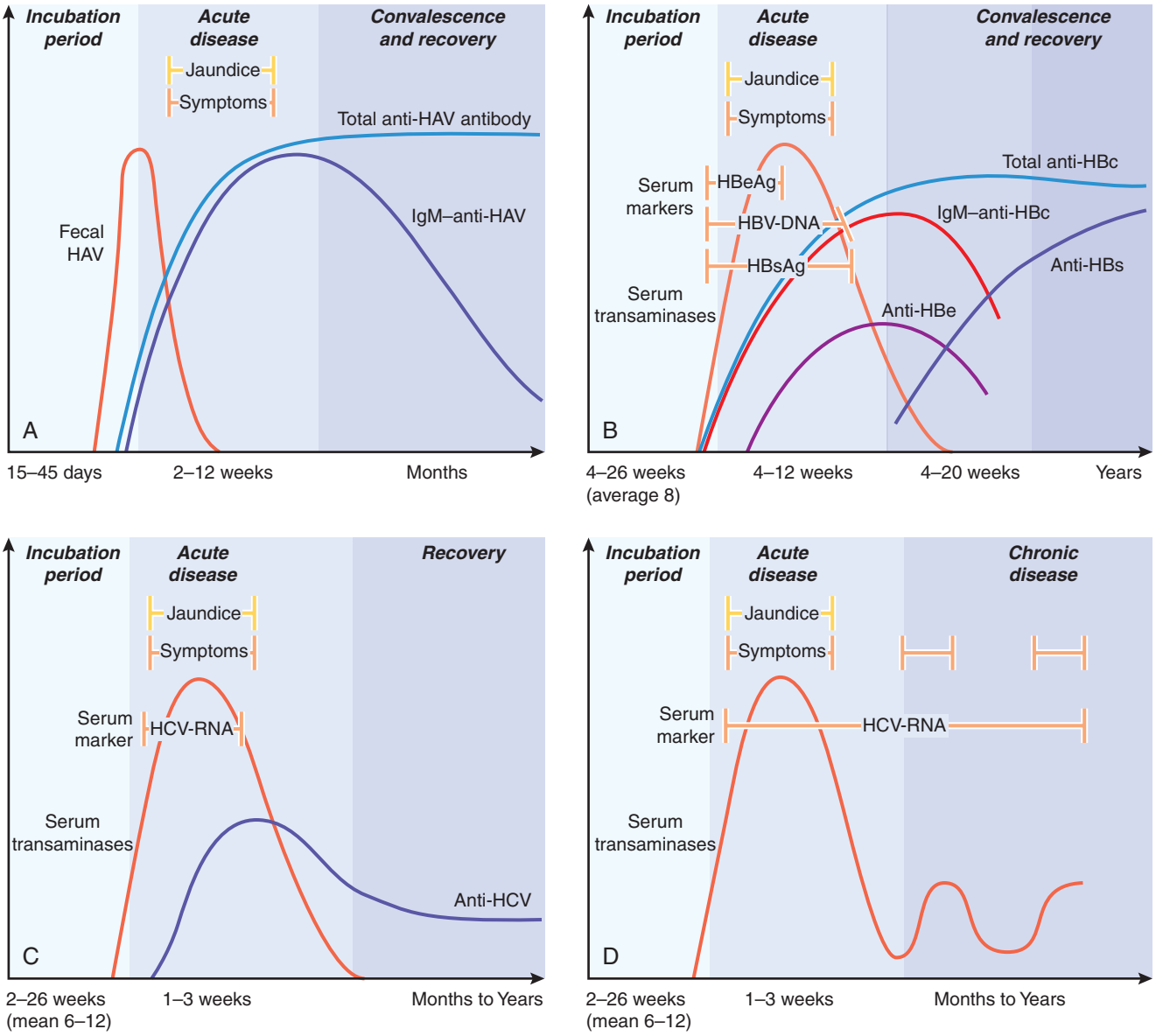


FIG 38.14 Comparison of clinical course and serologic tests in viral hepatitis. **A,** Hepatitis A only occurs in an acute form. **B,** The acute phase of hepatitis B—a small number of patients may not clear the virus during this stage and become chronic carriers. **C,** Acute hepatitis C is often not diagnosed because the symptoms tend to be mild. **D,** Most patients with hepatitis C infection are unable to clear the virus and develop chronic hepatitis C.

1 and 6 months of age. Adults are vaccinated at 0, 1, and 6 months. Simultaneous administration of hepatitis B immunoglobulin (HBIG) with other vaccines has no effect on efficacy. After the full course, the antibody response rate is 95% for normal hosts. The response rate is slightly lower in obese individuals, smokers, and men and is significantly lower in patients with cirrhosis or chronic renal failure, organ transplant recipients, children with celiac disease, and immunosuppressed patients. The effect of vaccination is considered long lasting, and booster shots are not required. Postvaccination testing is not recommended but could be considered for certain high-risk groups (e.g., sexual partners of chronic carriers).

Universal immunization is indicated for neonates and children in this country, but has also been suggested for individuals of any age, especially in the following high-risk groups: sexually active individuals with multiple sex partners, users of illicit drugs, household contacts of HBV carriers, hemodialysis patients, residents of institutions, health care workers, patients needing frequent transfusions, and individuals planning to reside in high-risk areas (e.g., the Far East and sub-Saharan Africa). If exposed to HBV, a susceptible person should receive one dose of HBIG and HBV vaccine as soon as possible after exposure and then complete the vaccination series. The issue of how to deal with persistent nonresponders, especially health care workers, is unsettled, but most authorities recommend a full three-injection course according to the usual schedule.

The administration of immunoglobulin containing high levels of hepatitis B surface antibody (HBIG) affords effective postinoculation

prophylaxis if given within 7 days of exposure. The indications for HBIG are as follows: (1) neonates born to HBsAg-positive mothers; (2) prophylaxis after needle stick or sexual exposure in nonimmune persons; and (3) after liver transplantation in patients who are HBsAg positive before transplantation. The usual dose is 0.05 to 0.07 mL/kg given intramuscularly, with the same dose repeated 25 to 30 days later. In cases of postexposure prophylaxis, the immune status of the recipient may be determined before treatment to avoid unnecessary administration. HBV vaccine should be given concomitantly with HBIG in most cases.

Hepatitis C

Pathogenesis and clinical manifestations. HCV (previously categorized as non-A, non-B hepatitis virus) is a single-stranded RNA virus that belongs to the Flaviviridae family. Our knowledge of HCV is evolving rapidly but has lagged behind that of HAV and HBV for several important reasons: lack of a suitable diagnostic test until the 1990s, lack of a suitable cell line for replication in the laboratory, and an extremely high mutation rate. Worldwide about 3% of the population is chronically infected, with a somewhat lower rate in the United States. The mode of transmission of HCV closely resembles that of HBV, although sexual and perinatal transmission is much less likely. The main pool of infected U.S. individuals acquired HCV through intravenous drug use or blood transfusions before the availability of the screening test in 1990. HCV remains an important occupational risk for health care workers, with the risk after a single needle stick being about 3%, as opposed to 30% for HBV or 0.3% for HIV. A significant number of seropositive persons have no known risk factors for HCV.

Acute HCV infection is usually asymptomatic. Clinical illness, when it occurs, is usually mild, with transaminase levels rarely exceeding 1000 IU/L. Only a minority of acute infections resolve, with the remainder progressing to chronic active infection (see Fig. 38.14). The course is erratic, with wide fluctuations in the concentrations of liver enzymes (primarily ALT). A number of extrahepatic manifestations occur, the most prominent of which are a medium-vessel vasculitis (polyarteritis nodosa), essential mixed cryoglobulinemia, and membranoproliferative glomerulonephritis. Chronic infection seems to progress to significant liver disease about 20% of the time, although there are no reliable noninvasive ways to predict who will progress. HCV infection is currently one of the most common causes of end-stage liver disease with cirrhosis in the United States. If untreated, as with hepatitis B, chronic active HCV infection with cirrhosis predisposes to HCC. Of the six recognized serotypes, type 1 is most common in the United States but has a lower response rate to treatment. Types 2 and 3 are also common in North America, whereas types 4 to 6 predominate overseas.

As with hepatitis B, screening for other bloodborne pathogens, including HIV and HBV, is recommended.

Treatment. Management of acute HCV infection is the same as for other acute viral strains (i.e., supportive and expectant unless complications or fulminant hepatic failure develops). Treatment of acute HCV infection with antiviral agents is not currently recommended, and immunoglobulin is not helpful in preventing infection in the acute exposure setting. Between 20% and 40% of acute seropositive patients will convert to seronegativity and an undetectable viral load during the first 6 months after infection, so early treatment is not recommended. Chronic infections should be assessed by determining viral load and viral genotype, and a liver biopsy to stage disease activity should be considered for those with a type 1 virus.

Few conditions have seen progress as rapid in treatment as hepatitis C. Current treatment regimens, though expensive, have response rates up to 95%. In the case of hepatitis C, a complete response in general indicates a permanent cure.

The current standard of treatment for chronic HCV infection depends on virus type. Treatment of type 1 can be undertaken with several newer protease inhibitors: ledipasvir-sofosbuvir, ombitasvir-paritaprevir-ritonavir plus dasabuvir, or simeprevir plus sofosbuvir. These all have comparable efficacies, with somewhat different side effect and drug interaction profiles. Ledipasvir-sofosbuvir (Harvoni) for 12 weeks is the current generally recommended treatment. It should be noted that high cost and coverage of insurance plans may affect access to one or all of these treatments. Response rates are around 95%, with relatively mild side effects. Treatment is also highly effective for patients with HCV/HIV coinfection.

Sofosbuvir plus ribavirin is the treatment of choice for genotype 2 hepatitis C infection, whereas genotype 3 is treated with daclatasvir plus sofosbuvir or interferon-based regimens. Response rates for types 2 and 3 are also extremely high.

Discussion of special situations with hepatitis C and nonresponders to conventional therapies is beyond the scope of this text, but interested readers can refer to the references for further information.

Patients with chronic active hepatitis C infection should be vaccinated against hepatitis A and hepatitis B and counseled regarding bloodborne precautions. Because the issue of sexual transmission is unsettled, the CDC does not currently recommend barrier methods for patients with long-term sexual partners because of the apparent low risk of infection.

Hepatitis D (Delta)

Pathogenesis and clinical manifestations. Hepatitis D virus (HDV) is an incomplete viral organism that requires the presence of HBV for replication. It may occur coincident with or subsequent to initial infection with hepatitis B. The disease is primarily transmitted by parenteral routes and by intimate personal contact, like hepatitis B. In the United States and northern Europe, HDV infection is most prevalent in persons exposed to blood and blood products (e.g., drug addicts and hemophiliacs). HDV infection tends to accelerate the progress of liver disease associated with HBV infection. In fact, fulminant hepatitis may result from HDV infection superimposed on chronic HBV infection. Because HDV is a deficient virus, its persistence is determined by the duration of the HBV infection. Diagnosis is by anti-HDV IgM and IgG enzyme-linked immunosorbent assays (ELISAs).

Treatment and control. HDV infection is controlled through the same measures used to prevent transmission of other hepatitis viruses: following safe sexual practices, screening blood products, avoiding intravenous drug use, and vaccinating susceptible persons with the HBV vaccine. There is no specific vaccine or treatment for HDV.

Hepatitis E

Pathogenesis and clinical manifestations. Hepatitis E virus (HEV) is a common cause of acute hepatitis in developing countries. Cases in developed countries are usually related to recent travel. HEV is an RNA virus spread via the fecal-oral route, especially through contaminated water. Parenteral transmission may occur.

The incubation period is 2 to 9 weeks. The prodrome and icteric illnesses are similar to those of HAV infection but usually last only 2 weeks. It is assumed that many subclinical cases occur, but in the absence of widely available serologic testing, subclinical infection is difficult to determine. Fulminant hepatic failure may occur, especially in pregnant women, and HEV can lead to chronic infection among immunosuppressed transplant patients.

Treatment. Treatment is supportive. Because no vaccine is available, the only prophylaxis is avoiding undercooked foods, performing careful handwashing, and drinking safe water and beverages (i.e., canned, bottled, or purified through the usual means). There is no standard treatment, but anecdotal evidence indicates some efficacy with ribavirin monotherapy.

Chronic Hepatitis

Chronic hepatitis encompasses a group of diseases characterized by inflammation of the liver that lasts 6 months or longer. The most prominent of these conditions is chronic active viral hepatitis, but chronic hepatitis may be due to toxic, autoimmune, or metabolic causes as well.

Chronic Persistent Hepatitis

Chronic persistent hepatitis, often called *triaditis* or *transaminitis,* is an archaic term for a chronic, low-grade liver inflammation of any cause. The inflammation is confined to the portal triads without destruction of normal liver structures, but serum transaminase levels are elevated. The condition may be asymptomatic or may be associated with mild, nonspecific symptoms. Progressive liver disease does not usually develop, and no drug treatment is indicated. The illness has an excellent prognosis. However, other more serious liver diseases may pass through a phase that is histologically indistinguishable from chronic persistent hepatitis and may progress (e.g., chronic viral hepatitis).

Current classification schemes emphasize (1) etiologic factor, (2) histologic grade, and (3) stage in terms of fibrosis. Therefore chronic persistent hepatitis would generally correspond to a liver condition with mild disease activity and minimal or no fibrosis by biopsy.

Chronic Active Hepatitis

Pathogenesis and clinical manifestations. On the other hand, chronic active hepatitis is a progressive, destructive inflammatory disease that extends beyond the portal triad to the hepatic lobule (so-called *"piecemeal" necrosis*). The severity spans the spectrum from mild to severe. The natural history is variable, because the disease could spontaneously arrest with any degree of fibrosis or could progress to cirrhosis and end-stage liver disease. Symptoms typical of acute hepatitis are often seen, including fatigue, malaise, nausea, anorexia, ascites, hepatomegaly, abdominal pain, and jaundice.

Patients with chronic active hepatitis may be grouped into several categories based on etiology. First, as discussed earlier, a minority of newly infected HBV patients but a majority of those with HCV will progress to chronic active hepatitis. The second subgroup (mainly young women) manifests autoimmune hepatitis and exhibits a variety of immunologic markers, including antinuclear antibodies (ANAs) and anti–smooth muscle antibodies (ASMAs). In addition, these patients frequently suffer from a second autoimmune disease such as Hashimoto thyroiditis. In the third subgroup are patients with chronic hepatitis induced by alcohol or other toxins, including therapeutic agents such as minocycline or nitrofurantoin. In the fourth subgroup are patients with a metabolic liver disorder such as Wilson disease or hemochromatosis. A small number of patients have neither a suggestive history nor any detectable markers to suggest an etiology; advanced liver disease in this group is usually termed *cryptogenic cirrhosis*.

Diagnosis. The diagnosis of chronic hepatitis is made on the basis of the clinical setting and abnormal liver enzymes, particularly transaminases. Serologic studies are indicated to assess for viral and autoimmune hepatitis. Serum iron and ferritin studies screen for hemochromatosis, and a serum ceruloplasmin level screens for Wilson disease. A liver biopsy may be performed to confirm the diagnosis and exclude other specific causes if the etiology is not forthcoming. Biopsy also allows for grading and staging, as discussed earlier in the chapter.

Management of chronic active hepatitis. Treatment depends on the cause; management of HBV and HCV infection has already been discussed. Managing toxic hepatitis involves discontinuing the offending drug, including alcohol. Specific antidotes or treatments are available for some of these conditions. Specific treatments for Wilson disease and hemochromatosis are discussed later in the chapter in the Metal Storage Diseases section.

Autoimmune Hepatitis

Diagnosis. Autoimmune hepatitis is a progressive liver inflammation characterized by the presence of several distinctive autoantibodies, as well as a polyclonal hypergammaglobulinemia. ANA is usually positive at a high level. ASMAs are a less sensitive indicator than ANA, but highly specific for autoimmune hepatitis. The finding of highly positive autoimmune markers in the presence of significant hepatitis, in the absence of other tests indicative of viral or metabolic liver disease, is pathognomonic for this condition. There is often some overlap between conditions, however, and in questionable cases a liver biopsy is indicated.

Management of autoimmune hepatitis. Corticosteroids and immunosuppressive drugs have been used since the early 1960s for the management of autoimmune chronic active hepatitis. Their use is based on the assumption that immunologic mechanisms either cause or maintain ongoing hepatic inflammation. Corticosteroids alone or in combination clearly lower mortality, most noticeably in symptomatic patients and those with very severe pathologic lesions demonstrated on liver biopsy. Current guidelines recommend prednisone with azathioprine until a remission is induced (usually within 18 to 24 months, as indicated by significant improvement in liver function test results). Subsequently, steroids can be tapered over 6 weeks or more, with azathioprine tapered more slowly; 65% to 80% of patients will respond to this initial regimen. Cyclosporine is another treatment option for nonresponders. Treatment with corticosteroids is accompanied by well-known complications: hypertension, edema, glucose intolerance, weight gain, and an impaired immune response, among others. Suddenly stopping long-term treatment with corticosteroids may result in acute adrenal insufficiency with hypotension and shock.

KEY POINTS

- Acute viral hepatitis is generally classified as hepatitis A, B, C, D (delta), and E infection. Modes of transmission and severity of symptoms differ among types.
- HAV infection is also known as *enteric hepatitis* because it is generally transmitted by ingestion of contaminated substances. Symptoms are flulike and tend to be less severe than those of HBV infection. Early treatment with γ-globulin and vaccination after exposure may be effective in preventing disease.
- HBV infection is transmitted through infected blood and body fluids, like HIV. The incubation period is longer and the severity of symptoms (particularly jaundice) is greater than in HAV infection. Hepatitis B immunoglobulin (HBIG) is effective after inoculation if given within 7 days of exposure. HBV vaccine is recommended as part of the childhood vaccination regimen and for high-risk individuals, as well as after exposure. Treatment is with intramuscular interferon-α for 6 to 12 months or with one of several nucleoside analogs for one to several years.
- HCV, also known as *non-A, non-B hepatitis virus,* resembles HBV in its routes of transmission. Chronic HCV infection develops in 85% of cases and is usually asymptomatic until advanced liver disease intervenes. Immunoglobulin does not protect against HCV infection. Treatment is with one of the newer protease inhibitors for 12 to 24 months.
- HDV coinfects with HBV and requires the presence of HBV to be active. Infection accelerates and worsens the manifestations of HBV infection. Prevention of HBV infection also prevents HDV infection.
- HEV is a common virus in the developing world that causes an illness similar to HAV infection but has a relatively high mortality in pregnant women.

- Chronic hepatitis is characterized by persistent inflammation of the liver lasting 6 months or more. Autoimmune disease, viral hepatitis (B and C), toxins, and metabolic diseases may result in chronic hepatitis. Chronic active hepatitis may progress to cirrhosis. Corticosteroids (prednisone) and immunosuppressants (azathioprine) are common therapeutics for autoimmune hepatitis.

CIRRHOSIS

Cirrhosis represents the irreversible end stage of many different hepatic injuries, including CAH, steatohepatitis (i.e., "fatty liver"), the metal storage diseases, alcoholic liver disease, and toxic hepatitis. It is characterized histopathologically by diffuse hepatic fibrosis surrounding nodules of liver tissue and results in permanent alteration in hepatic blood flow and liver function. The pathophysiology of cirrhosis and its complications have been reviewed earlier in this chapter. We will discuss briefly the sequelae of chronic biliary disease and alcoholic liver disease.

Biliary Cirrhosis

Etiology and pathogenesis. Biliary cirrhosis represents the end result of continuous, ongoing inflammation of the bile ducts, which may be due to macroscopic or microscopic biliary obstruction. Persistent biliary obstruction results in inflammation and scarring of the liver, with obliteration of the bile ductules. The consequence is diffuse and widespread fibrosis with regenerative nodule formation (islands of healthy liver tissue within a background of fibrosis), with the complications of portal hypertension described earlier. The prototypical form of microscopic biliary disease is *primary biliary cirrhosis (PBC),* an autoimmune condition often associated with systemic lupus erythematosus and other autoimmune illnesses. It is usually associated with positive tests for ANA and antimitochondrial antibodies.

Examples of large-duct obstruction include gallstone disease, chronic biliary fluke infestation, and primary sclerosing cholangitis (PSC). Immigrants from infected areas may be at risk for chronic parasitic infestations associated with liver disease. The biliary flukes endemic in Asia, *Opisthorchis* and *Clonorchis* species, are acquired by eating raw fish that carry larval cyst forms. Fasciola *hepatica* is found in all sheep-producing and cattle-producing areas of the world and infects humans as accidental hosts; it is acquired by eating fecally contaminated watercress and other aquatic plants that harbor the immature cysts.

Diagnosis and treatment. The diagnosis of PBC depends on appropriate serologic and liver biopsy results. Ursodeoxycholic acid (ursodiol, UDCA) is the only approved treatment for PBC and may delay development of end-stage liver disease. Methotrexate and colchicine are other therapies whose efficacy has yet to be elucidated in studies, but which may be added to UDCA in certain cases. Supportive care is similar to that for other conditions with impaired bile metabolism, including supplementation with fat-soluble vitamins. In spite of these measures, most patients will eventually require liver transplantation for prolonged survival.

The diagnosis of biliary flukes in immigrants and (less commonly) foreign travelers can usually be made by serial stool examinations for parasitic ova, and serologic testing is often helpful. Treatment with antiparasitic agents (praziquantel, mebendazole, or albendazole) will eliminate the live worms but will not necessarily prevent recurrent episodes of cholangitis and other consequences.

Primary Sclerosing Cholangitis

Etiology and pathogenesis. PSC is an autoimmune condition usually seen in patients with ulcerative colitis; 80% of PSC patients have coexistent ulcerative colitis, whereas 3% to 5% of ulcerative colitis patients ultimately develop PSC. It is characterized by recurrent episodes of cholangitis, with progressive biliary scarring and obstruction. Secondary forms of sclerosing cholangitis, such as that occurring after bile duct injury during surgery, behave in a similar fashion.

Diagnosis and treatment. A majority of patients with PSC will have a positive perinuclear antinuclear cytoplasmic antibody test, though this is not necessarily pathognomonic. Diagnosis is primarily by means of ERCP or MRCP showing the typical beaded and atrophic appearance of the biliary tree; liver biopsy is often performed for staging reasons or for suspected small-duct PSC. Although a variety of treatments have been investigated, medical and endoscopic treatments (for relief of strictures) are merely palliative, and the only effective recourse is liver transplantation. The end result of recurrent cholangitis of any cause is cirrhosis, and such patients are predisposed to cholangiocarcinoma.

Alcoholic Liver Disease

Alcoholic liver disease is manifested by fatty liver, hepatitis, and cirrhosis. One or more of these manifestations may be found in alcoholic patients.

Alcoholic Fatty Liver

Etiology. Alcoholic fatty liver (also termed *alcoholic steatohepatitis*) is an accumulation of fat in the liver cells caused by more fat being delivered to the hepatocyte than it can normally metabolize or by a defect in fat metabolism within the cell.

Diagnosis and treatment. Steatohepatitis is commonly alcohol related, but can also result from diabetes mellitus, obesity, protein malnutrition, total parenteral nutrition, drugs, and many other factors that result in a similar pathologic process. It is usually mild and asymptomatic, and is often diagnosed incidentally on ultrasound or CT examination or by liver biopsy for another reason. Levels of liver enzymes are often mildly abnormal, generally serum transaminases. Hypertriglyceridemia is commonly found and at times may be dramatically elevated (e.g., greater than 1000 mg/dL). Occasionally, there is significant liver enlargement, abdominal discomfort, and even portal hypertension. Treatment of alcoholic steatohepatitis involves stopping alcohol intake and providing appropriate nutrition.

Nonalcoholic steatohepatitis (NASH) may be managed with reduction of weight, control of diabetes and hyperlipidemia, or other treatment directed at the underlying cause. Although the exact risk is unclear, it has been estimated that between 3% and 15% of persons with untreated NASH will eventually develop progressive liver fibrosis and cirrhosis.

Alcoholic Hepatitis

Pathogenesis and clinical manifestations. Alcoholic hepatitis is an active inflammation of the centrilobular region of the liver. The liver cells show pathologic changes of hepatocyte necrosis with neutrophilic infiltration and intracellular inclusions known as *Mallory bodies.* This form of liver disease often occurs in chronic alcoholics who binge on quantities much greater than their usual intake. Clinically the illness ranges from mild to very severe, with the worst cases characterized by hepatomegaly, fever, signs of acute liver failure, and encephalopathy. Hepatitis may be complicated by acute alcohol withdrawal and delirium tremens. Mortality rates as high as 33% are seen with this condition.

Diagnosis and treatment. The diagnosis is suggested by the history. The finding of a serum AST (SGOT) level markedly higher than the serum ALT (SGPT) level strongly suggests a toxic etiology rather than a viral hepatitis. Viral serologies, serum acetaminophen levels, and tests for certain metabolic disorders (e.g., determination of serum ceruloplasmin and ferritin) may help sort diagnostic dilemmas.

Because malnutrition and vitamin deficiencies are important pathogenic factors in alcoholic hepatitis, special attention should be

given to nutrition. Thiamine 100 mg daily and a standard multivitamin are administered routinely, and vitamin B_{12} and folate levels should be measured and replenished as necessary. An elevated prothrombin time (or international normalized ratio) is an indication of major liver failure, and such patients should receive subcutaneous vitamin K. Corticosteroid therapy with prednisolone is recommended for patients with severe alcoholic hepatitis, especially those with declining liver function and coma, but patients with mild to moderate impairment do not benefit from steroids and should receive supportive care alone.

TOXIC LIVER DISORDERS

Metal Storage Diseases

Hereditary Hemochromatosis

Pathogenesis and diagnosis. Hereditary hemochromatosis (HH) is one of the most common autosomal-recessive disorders in the world. In European populations, approximately 1 in 10 persons is a heterozygous carrier, and 0.5% are homozygous persons with disease. Studies in the United States show a prevalence of HH homozygosity of 0.44% for Caucasians, 0.027% for Hispanics, and 0.014% for African Americans. The prevalence among Asians is extremely low. However, despite the high prevalence of the gene, the number of persons with clinical hemochromatosis is considerably smaller—less than 1% of homozygotes. This is due to incomplete penetrance of the gene and environmental factors.

The disease is caused by the activity of a mutant gene called *HFE,* which allows excessive and uncontrolled iron absorption by the GI tract. The usual *HFE* mutation has been identified as C282Y, but other mutations can occur and act in a similar way. A small number of patients with HH are heterozygotes for *HFE;* disease in this population may result from as-yet-unidentified mutations or other factors. The result of these mutations is iron deposition in numerous organs; in advanced disease, the body may contain 20 g or more of iron, mainly in the liver, pancreas, and heart. Because of menstruation and perhaps endocrinologic factors, hemochromatosis is much less common in women than in men (ratio of 1:5 to 1:10 female-to-male ratio).

The liver is usually the first organ to show evidence of involvement, with hepatomegaly and elevated levels of liver enzymes. Specific manifestations in organ systems other than the liver include diabetes mellitus, hyperpigmentation, polyarthritis, hypogonadism, heart failure, and bronze discoloration of the skin. In advanced disease, fibrosis and macronodular cirrhosis of the liver develop insidiously and represent the major cause of death. Splenomegaly is common, although portal hypertension and its complications (see earlier discussion) occur less frequently than with other forms of liver disease. HCC develops in about 30% of persons with hemochromatosis, exclusively in the setting of cirrhosis. Therefore early diagnosis and treatment are critical.

Clinical manifestations and diagnosis. The diagnosis is suggested by clinical features and family history. Plasma iron and transferrin saturation is increased, and serum ferritin level is dramatically elevated, often to several thousand micrograms per liter (normal 10 to 200 mcg/L). The diagnosis of hereditary hemochromatosis is confirmed by genetic analysis for the *HFE* gene. In selected cases liver biopsy may be performed, demonstrating iron deposition in periportal hepatocytes. The diagnosis of HH should prompt investigation of other family members for carriage and expression of the *HFE* gene.

It should be noted that a second significant category of iron overload syndromes exists: secondary hemochromatosis. This may occur in patients who ingested large amounts of iron or who have received repeated blood transfusions for chronic dyserythropoietic states (e.g., sickle cell disease, thalassemia, or congenital or acquired sideroblastic anemias). Of historical interest is a condition called *Bantu siderosis,* which occurred among South African tribesmen who regularly imbibed beer fermented in iron pots. These secondary forms of hemosiderosis are not associated with the *HFE* gene.

Treatment. Recent observational studies have shown that asymptomatic patients with a relatively low iron load have an essentially normal life span. Therefore treatment is reserved for those patients with a ferritin level greater than 1000 mcg/L. The mainstay of treatment for hemochromatosis is repeated phlebotomy. The typical protocol is weekly phlebotomy of 500 mL (1 unit) of whole blood until the hematocrit drops below 37%, at which time maintenance phlebotomy of 1 unit is carried out every 2 to 3 months. Patients who do not tolerate phlebotomy may be treated with subcutaneous or intramuscular deferoxamine, a drug that chelates iron and facilitates its renal excretion. Deferoxamine is much less efficient than phlebotomy and requires adequate renal function. If identified early, hereditary and acquired hemochromatosis carries an excellent prognosis in terms of preventing heart failure and liver disease. Diabetes may still develop, however, and iron removal does not change hypogonadism or arthritis. Liver transplantation is available for patients with irreversible cirrhosis whose heart involvement does not preclude surgery.

Wilson Disease (Hepatolenticular Degeneration)

Etiology. Wilson disease, or hepatolenticular degeneration, is a rare autosomal-recessive disorder in which excessive amounts of copper accumulate in the liver and other organs. As with HH, it has now been linked to a specific abnormal gene mutation in the Wilson disease protein *ATP7B* gene, which results in retention of copper in the liver as well as impaired incorporation of copper into ceruloplasmin. Most patients are compound heterozygotes with more than one of several mutations involving the *ATP7B* gene. The condition may present at any time before age 30, generally with either significant hepatic dysfunction or neuropsychiatric illness. Patients with a neuropsychiatric presentation virtually always have occult compensated cirrhosis at the time of diagnosis.

Clinical manifestations. Hepatic disease is more common in children than adults and begins as hepatomegaly, fatty infiltration of the liver, and elevated levels of liver enzymes. Great variability in liver disease is seen, and Wilson disease may manifest as acute hepatitis progressing to fulminant hepatic failure, a condition similar to autoimmune hepatitis with numerous extrahepatic symptoms, or insidious development of macronodular cirrhosis with portal hypertension. Neurologic involvement presents as a movement disorder or rigid dystonia, or occasionally as primarily psychiatric symptoms. Other manifestations include renal tubular acidosis with a Fanconi-like syndrome, cardiomyopathy, hypogonadism, metabolic bone disease (i.e., vitamin D–resistant rickets), and arthritis.

Diagnosis. Clinical signs and symptoms suggest the diagnosis, in particular, the finding on slit-lamp examination of the brownish Kayser–Fleischer rings at the margin of the cornea. (However, lack of Kayser–Fleischer rings does not exclude the diagnosis.) The combination of low serum ceruloplasmin level, Kaiser–Fleischer rings, and elevated 24-hour urinary copper excretion are diagnostic of Wilson disease. Results of 24-hour urinary copper excretion after penicillamine administration provide additional diagnostic proof in ambiguous cases. The liver biopsy technique for copper determination is technically demanding and is not routinely performed. Although the heterogeneity of genetic abnormalities precludes routine testing (as with hemochromatosis), genetic screening of close relatives should be carried out when the ATP7B mutation is known to be present.

Treatment. Treatment involves dietary modification and copper removal therapy. Patients should eliminate copper-rich foods from their diet, including organ meats, shellfish, nuts, chocolate, and mushrooms.

Dietary measures include filtering water with a high copper content. Oral chelation therapy is the mainstay of treatment for Wilson disease, and the currently preferred agent is trientine. Alternative agents include penicillamine and zinc. Patients receiving this treatment early in the course of the disease will show marked improvement and protection against liver and neurologic disease. Even patients with advanced Wilson disease may expect some functional recovery. Treatment is lifelong, and noncompliance leads to definite progression.

Side effects are less common with trientine than with the older therapy, penicillamine. A mild anemia and gastritis may occur with this medication, also a transient worsening of neurologic symptoms. Finally, ammonium tetrathiomolybdate is occasionally used for severe neurologic Wilson disease because, unlike penicillamine, it is not associated with early transient neurologic deterioration. As with hemochromatosis, liver transplantation has a limited but useful role.

Toxic Metabolic Agents

Acetaminophen Poisoning

Etiology. Many drugs and toxins cause liver damage. Unfortunately, treatment is often limited to withdrawal of the offending agent and administration of supportive care. Standard measures, including gastric lavage, induced emesis, and activated charcoal, are used in cases of acute ingestive poisoning. Specific antidotes are few, although heavy metal intoxication may be managed with chelating drugs. Acetaminophen overdose is an important exception that bears further discussion.

Pathogenesis and clinical manifestations. Acetaminophen (trade name Tylenol, also known as *paracetamol* in Britain and Europe) is a widely used, nonprescription analgesic and antipyretic frequently implicated in suicide attempts and accidental poisonings. In fact, a multicenter study recently showed that acetaminophen overdose was responsible for 39% of cases of acute hepatic failure in the United States. Oral acetaminophen is rapidly absorbed and metabolized (Fig. 38.15). A toxic metabolite, *N*-acetyl-*p*-benzoquinone imine, is formed and rapidly detoxified by reaction with glutathione. However, acute ingestion of at least 140 mg of acetaminophen per kilogram of body weight may expose the liver to high levels of the toxic metabolite with resultant hepatic necrosis. Importantly, repeated ingestion of smaller amounts may cause harm in children, the elderly, patients with preexisting liver disease, persons abusing alcohol, and persons taking other hepatotoxic drugs. With acute ingestion significant liver damage is rare if the serum acetaminophen levels are less than 150 and 37 g/mL, respectively, 4 and 12 hours after ingestion. Given the many variables affecting acetaminophen toxicity, it is imperative for clinicians evaluating a suspected case to refer to a published nomogram and monitor serial drug levels of acetaminophen. (Screening for ingestion of other hepatotoxic drugs, such as tricyclic antidepressants, should also be done.)

Within several hours of ingestion, the patient generally experiences nausea, vomiting, and diarrhea. After these symptoms, there is often a temporary "window" period, in which the patient feels well and may wish to withdraw from medical care. Within 24 to 48 hours, signs of hepatic injury occur, including abnormal liver enzyme levels. If untreated, progressive liver failure with jaundice, encephalopathy, hypoglycemia, coagulopathy, and even death may occur, generally within the first week. Patients surviving 1 week generally experience complete recovery of normal liver function without sequelae.

Treatment. Initial treatment, as for any poisoning, involves decontamination with induced emesis or lavage and activated charcoal, which will not interfere with use of specific treatment. The proper use of acetylcysteine for patients with clearly toxic levels can effectively prevent hepatic necrosis and its fatal consequences, if started in a timely fashion. Acetylcysteine, a mucolytic solution often used in patients with bronchial diseases, stimulates liver production of reduced glutathione, which can then keep up with blood levels of toxic metabolites. Acetylcysteine is nontoxic but may cause rash and frequently induces vomiting because of its foul odor and taste. It is often given by nasogastric tube in conjunction with antinausea medications. It can also be given intravenously, although this preparation is associated with a 10% incidence of anaphylactic-type reactions requiring immediate attention.

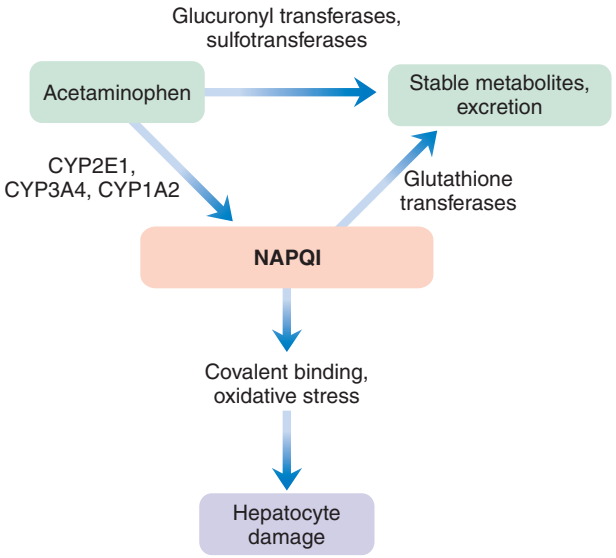


FIG 38.15 Mechanism of acetaminophen toxicity. The majority of an administered dose of acetaminophen is conjugated with sulfate or glucuronic acid to form stable metabolites that are promptly excreted in urine. Under most circumstances, only a minority of the total acetaminophen dose undergoes bioactivation to the reactive intermediate *N*-acetyl-*p*-benzoquinone imine (NAPQI). This species is capable of binding to intracellular proteins and mediating cell injury or death. The liver enzymes capable of bioactivating acetaminophen to NAPQI are CYP2E1, CYP3A4, and CYP1A2. Accumulation of NAPQI in the liver does not generally occur because the liver has sufficient ability to detoxify this metabolite through glutathione conjugation. (Redrawn from Fontana RJ, Watkins PB: Genetic predisposition to drug-induced liver disease. *Gastroenterol Clin North Am* 1995;24:811–823.)

OTHER STRUCTURAL LIVER CONDITIONS

Liver Abscess

Pathogenesis, clinical manifestations, and diagnosis. Pyogenic liver abscess is a common condition worldwide. In the United States it commonly results from one of three situations: ascending cholangitis, with or without gallstones; intraabdominal infection (e.g., appendicitis or diverticulitis), which drains through the portal vein; or hematogenous seeding from a systemic bacteremia or endovascular infection.

Liver abscess should be considered in any patient with fever and right upper abdominal pain. Nausea and vomiting are common, and jaundice from biliary obstruction may rarely be present. Frequently, tender hepatosplenomegaly and sometimes a palpable mass are noted. Typical signs of pyogenic infection are usually present, including fever and leukocytosis, and elevated levels of liver enzymes are present, often in a "mixed" pattern. Because the diagnostic considerations include cholecystitis, biliary obstruction, and liver tumor, imaging of the liver with ultrasonography or CT is usually carried out. Blood cultures should be obtained, and ultrasound-guided aspiration of abscesses is recommended for Gram stain and culture.

Pyogenic liver abscesses from abdominal sources generally contain enteric aerobic and anaerobic gram-negative bacteria. *Escherichia coli,*

Klebsiella species, and *Bacteroides fragilis* are particularly important. Abscesses of hematogenous origin are more heterogeneous. *S. viridans* and related streptococcal species are commonly seen. *Staphylococcus aureus* is seen in the setting of endocarditis or widespread bacteremia.

Other uncommon causes of cystic liver abnormalities should be considered, in particular hydatid liver cysts secondary to *Echinococcus* species and amebic liver abscess. These conditions are virtually always seen among persons who have lived in developing countries for many years. Their appearance on ultrasound and CT is usually characteristic, and serologic tests are available to assist with diagnosis. (Aspiration of probable echinococcal cysts is actually contraindicated due to the possibility of anaphylactic shock from leakage of immunogenic cyst material.) Fungal and mycobacterial infections of the liver are usually granulomatous and diffuse, but in certain settings, such as in neutropenic patients, localized infection with an organism such as Candida or *Aspergillus* may be seen.

Treatment. Large (2.0 cm) solitary or multiple liver abscesses require drainage, formerly a surgical procedure. At present CT- or ultrasound-guided percutaneous drainage is the standard of care. Percutaneous tubes may be placed with minimal morbidity and discomfort and should be kept in place until drainage has essentially resolved.

Antibiotic coverage should be directed at likely organisms. A β-lactam with β-lactamase inhibitor (e.g., ampicillin/sulbactam) or a third-generation cephalosporin such as ceftriaxone plus metronidazole are appropriate choices. Ertapenem is a once-daily carbapenem with broad aerobic and anaerobic coverage and is a good choice for outpatient administration. Because up to 4 weeks of antibiotic administration are generally required, the latter part of the treatment course can often be given orally, for instance, with an oral quinolone such as levofloxacin with or without oral metronidazole. Therapy should be adjusted depending on culture results.

Trauma

Etiology and clinical manifestations. The liver is the most common solid organ to be injured by penetrating abdominal trauma (such as gunshot wounds, stab wounds, or rib fractures) and the second most commonly injured organ in blunt trauma. Damage or injury to the liver should be suspected when any upper abdominal or lower chest trauma is sustained. The liver is frequently injured by steering wheels in vehicular accidents. Common injuries to the liver include simple lacerations, multiple lacerations, avulsions, and crush injuries.

The gravity of liver wounds arises from the fact that the liver is a highly vascular organ that receives approximately 29% of the body's cardiac output. When hepatic trauma occurs, blood loss can be massive. The patient generally exhibits the typical signs of hemorrhagic shock: hypotension, tachycardia, tachypnea, pallor, diaphoresis, and confusion. Hemoglobin/hematocrit levels may be normal early after the trauma but eventually will reflect significant blood loss. Clinical manifestations include right upper quadrant pain with abdominal tenderness, distention, guarding, and rigidity. Abdominal pain exaggerated by deep breathing and referred to the shoulder may indicate diaphragmatic irritation.

Standard trauma imaging studies such as CT or ultrasound are generally diagnostic of this condition.

Treatment. Treatment entails the administration of fresh whole blood or packed red cells and fresh frozen plasma, as well as massive fluid infusion to maintain adequate intravascular volume. Many patients will require surgical management, although some may be treated angiographically or expectantly with medical support. Postoperatively, a patient with hepatic trauma is usually admitted to a critical care unit and monitored for persistent bleeding. The complete blood cell count and coagulation parameters must be closely monitored for trends.

Malignancy

Etiology. Malignancy in the liver usually develops as a metastatic process. Because of the vascularity and lymphatic drainage of the liver, the organ is a common site for metastasis from primary cancers of the esophagus, stomach, colon, rectum, breast, and lung—among many other possibilities.

Primary hepatic malignancy (cancer originating within the liver) is relatively rare in the United States, though it has been increasing in frequency—approximately 36,000 cases in the United States annually. However, in other parts of the world such as Africa, it is one of the most common sites of malignancy because of a high prevalence of chronic hepatitis B virus infection. Primary liver tumors include HCC (discussed in the next section), cholangiocarcinoma, and angiosarcoma. Hepatoblastoma is the most common malignant tumor in children. Lymphoma, especially T-cell lymphoma, may occasionally arise primarily in the liver. Benign liver tumors are much less common, with the exception of cavernous hemangioma.

Clinical manifestations and diagnosis. The most common form of primary hepatic malignancy is hepatocellular carcinoma (HCC), often referred to as *hepatoma*. HCC is a malignancy of middle-aged persons, more frequent in men than in women. The incidence is increasing in the United States, in part as a result of increasing HBV and HCV prevalence. Signs and symptoms of HCC include hepatomegaly, abdominal pain, weight loss, nausea, and, in advanced cases, jaundice and ascites. HCC has an extraordinary number of paraneoplastic syndromes, including hypercalcemia, erythrocytosis, hypoglycemia, thyrotoxicosis, and hypertrophic osteoarthropathy (finger clubbing).

Space-occupying masses of the liver are often first suggested by abnormal measurements of liver-related enzymes, especially alkaline phosphatase. α-Fetoprotein is often dramatically elevated in cases of HCC, although false-positive results can occur. Imaging studies of the liver are usually quite characteristic, although on occasion the tumor may be diffuse and difficult to image. Radiographic studies also allow guided needle biopsy of the liver lesion.

Treatment. The only treatment for HCC is hepatic resection. Unfortunately, because of advanced diffuse liver disease or multifocal tumors, such treatment is not usually possible. Partial resection is preferred, but complete hepatectomy followed by liver transplantation is a radical option for tumor localized to the liver. One of the preferred modes of treatment currently is transarterial chemoembolization (TACE). Other treatments include systemic chemotherapy and chemotherapy directed selectively to the liver via portal vein or hepatic artery cannulation. A newer drug, sorafenib, has been approved for treatment of HCC not amenable to curative surgery and has shown a modest survival benefit. Nevertheless, the 5-year survival for this cancer is only 14%.

TRANSPLANTATION

Patients with end-stage liver disease that has not improved with conventional medical therapy are potential candidates for liver transplantation (Boxes 38.2, 38.3, and 38.4). In adults, diseases currently managed by orthotopic (in-place) liver transplantation include end-stage cirrhosis from chronic active hepatitis, alcoholic liver disease, primary biliary cirrhosis, and primary sclerosing cholangitis, as well as hepatic metabolic diseases such as hemochromatosis and Wilson disease. Liver transplantation is rarely performed for patients with malignant neoplasms, although there is clearly a subset of patients with HCC who are candidates for this procedure. The major indication for pediatric transplantation is biliary atresia after a failed Kasai procedure (portoenterostomy) or delayed recognition of the diagnosis. Other major pediatric indications include α1-antitrypsin deficiency and other metabolic disorders.

BOX 38.2 Conditions Treated With Liver Transplantation

Cirrhosis Secondary to
Viral hepatitis
Alcoholic hepatitis
Autoimmune hepatitis
Cryptogenic source (no cause determinable)

Metabolic Liver Diseases
Wilson disease
α_1-Antitrypsin deficiency
Hemochromatosis, neonatal
Glycogen storage disease
Tyrosinemia
Byler disease
Crigler–Najjar syndrome
Other miscellaneous

Cholestatic Liver Diseases
Primary biliary cirrhosis
Sclerosing cholangitis
Biliary atresia
Other miscellaneous

Acute Liver Failure
Drug reactions
Toxins (e.g., mushroom poisoning)
Viral hepatitis
Acute Budd–Chiari syndrome, other ischemic insult

BOX 38.3 Indications for Liver Transplantation

Acute Liver Disease
Fulminant liver failure with progressive encephalopathy
Prothrombin time >10 sec above control
Bilirubin <15 mg/dL and rising

Chronic Liver Disease
Bilirubin >15 mg/dL
Intractable hepatic encephalopathy
Intractable ascites
Serum albumin <2.5 mg/dL
Prothrombin time ≥20 sec above control
Hepatorenal syndrome

Currently about 5000 liver transplantations are performed yearly, although the waiting list contains about three times that number.

Evaluation of the Transplantation Patient

Potential transplantation patients undergo extensive physiologic and psychological evaluation by physicians, nurses, psychologists, and social workers to identify potential contraindications to the procedure. Conditions that would normally preclude transplantation include uncontrolled bacterial sepsis, failure of other major organ systems, extrahepatic malignancy, and more. Additional identified risk factors include portal vein thrombosis, previous portosystemic shunt operations, current alcohol or drug addiction, a poor psychosocial support system, and psychological instability. Many of these conditions are relative contraindications that diminish the chance of a successful outcome, but they do not always preclude transplantation. Patients with HIV and end-stage liver disease represent a growing group that is increasingly being considered.

BOX 38.4 Contraindications to Liver Transplantation

Absolute Contraindications
End-stage cardiopulmonary disease
Metastatic cancer
Active sepsis
Acquired immunodeficiency syndrome
Psychiatric illness preventing compliance with treatment

Relative Contraindications (Vary Greatly Among Transplant Centers)
Renal failure
Hepatitis B infection
Liver cancer
Portal vein thrombosis
Active alcoholism

Patients with viral hepatitis are particularly susceptible to recurrence in the transplanted organ and must be managed with care. Recent data show that patients transplanted for HBV who are treated indefinitely with high-dose HBIG have low recurrence rates, with acceptable survival and quality of life. In contrast to HBV, it has not been possible to develop an effective regimen to prevent recurrent HCV infection. A promising approach is use of direct-acting agents once clinical recurrence after transplant has been identified; this approach is currently the focus of ongoing clinical trials.

After the patient has been identified as a candidate and a donor organ has been procured, the actual surgical procedure can take 8 to 22 hours to complete. The procedure involves five anastomoses between recipient and donor organs, including the following vascular anastomosis sites: suprahepatic inferior vena cava, infrahepatic vena cava, portal vein, hepatic artery, and biliary tract. The biliary anastomosis site varies, depending on the patient's extrahepatic biliary tract.

Posttransplantation Management

A cornerstone of posttransplantation management is immunosuppression to prevent rejection of the transplant graft. The rejection response after liver transplantation most often occurs between postoperative days 4 and 10. Clinical manifestations of acute rejection include tachycardia, fever, right upper quadrant or flank pain, diminished bile flow through the T-tube drain or a change in bile color and increasing jaundice. Laboratory findings include elevated serum bilirubin, transaminase, and alkaline phosphatase levels and increased PT. After the successful use of cyclosporine, a number of immunosuppressive drugs have appeared and provide several choices for improving outcome. Immunosuppressives may be broadly categorized into three groups: initial immunosuppression, maintenance immunosuppression, and management of acute cellular rejection. Prednisone is the primary posttransplantation immunosuppressive and is steadily tapered in favor of maintenance drugs. The calcineurin inhibitors tacrolimus or cyclosporine are begun during anesthesia induction and represent the mainstays of maintenance therapy, with tacrolimus being the preferred therapy. Adjunctive maintenance agents include either mycophenolate mofetil or the older drug azathioprine. Acute rejection is generally managed with high-dose corticosteroids. Patients who fail to respond to steroids can be treated with infusion of one of several agents, including muromonab (OKT3), mycophenolate mofetil (MMF), or an interleukin-2 receptor blocker (e.g., basiliximab).

Side effects limit the usefulness of all these drugs and are a main source of morbidity and mortality among transplant recipients. The main nonimmunologic side effects of cyclosporine and tacrolimus are hypertension and renal insufficiency. Blood levels must be carefully monitored. Side effects of MMF and azathioprine include bone marrow suppression with cytopenias. Problems with the acute antirejection infusions include hypersensitivity and cytokine reactions and a heightened risk of opportunistic infection, especially cytomegalovirus (CMV), immediately after use. Among the many side effects of prednisone are hypertension and hyperglycemia. The main consequence of all these treatments is immune suppression and increased risk for infection.

Early infections are generally due to issues involving surgical technique and preexisting infection (e.g., cholangitis) and have declined significantly in recent years. They usually represent nosocomially acquired pathogens. Infections in the middle period from 1 to 6 months often represent viral infection or reactivation. TMP-SMX is usually prescribed for prophylaxis against bacterial infection and *Pneumocystis jiroveci* for up to 1 year. Of particular concern and the focus of several prophylactic strategies is CMV. CMV occurs at a high rate among seronegative recipients who receive a liver from a seropositive donor, and reactivation rates are significant as well. Use of antiviral treatments for CMV is usually undertaken for 3 to 6 months in patients at high risk for CMV disease (e.g., CMV-negative recipients who receive a CMV-positive donor liver).

In the middle and late periods fungal infection is important, especially *Aspergillus* species, and lymphoproliferative disorder attributable to Epstein–Barr virus is seen. Transplant recipients should be instructed to avoid exposure to environmental or foodborne mold, which could increase their risk for fungal infection. Another important preventive strategy is appropriate vaccination, particularly annual influenza shots. (Live virus vaccinations, such as varicella and yellow fever, should be avoided.)

Critical issues in the posttransplantation period include the following: hypertension, renal dysfunction, hyperlipidemia and cardiovascular disease, obesity, osteoporosis, and increased risk for cancer. Psychological issues are especially prominent, and caregivers and family should be alert for signs of depression and anxiety. Many of the antirejection medications exacerbate these symptoms. With careful follow-up, recipients can live productive lives for many years after transplantation. Chronic rejection, often in the setting of progressive ductopenia, and recurrence of primary pretransplantation liver disease tend to cause graft failure with time. Actuarial survival at 5 years is approximately 88% for persons with cholestatic liver disease, 78% for patients with noncholestatic liver disease who are HCV negative, and 70% for persons with HCV.

AGE-RELATED LIVER DISORDERS

Liver Diseases and Pediatric Considerations

Liver disease in infants and children not only encompasses pediatric variations of adult liver diseases, but also includes conditions unique to that age group. Many of these conditions present at birth or shortly thereafter, although several may appear in later life.

Abnormal Bilirubin Metabolism in the Neonatal Period

Physiologic jaundice of the newborn is a harmless condition lasting no longer than 2 weeks after delivery. (Bilirubin metabolism in the neonatal period has already been discussed; see earlier Jaundice section.) Immature bilirubin conjugation and transport mechanisms are the primary causes, along with increased gut absorption of bilirubin. Hyperbilirubinemia is not considered physiologic or normal if the bilirubin level is greater than 5 mg/dL on the first postpartum day, 10 mg/dL on the second day, or 13 mg/dL at any time. It should be noted that breast-fed babies have a higher incidence of hyperbilirubinemia than do bottle-fed babies because the β-glucuronidase in breast milk results in increased unconjugated bilirubin in the gut, which can be absorbed. If significant hyperbilirubinemia occurs, breast feeding can be stopped temporarily.

Pathologic bilirubin levels should lead to an immediate evaluation to exclude congenital hemolytic disorders, Crigler–Najjar syndrome, hypothyroidism, congenital pyloric stenosis, sepsis, resorbing hematomas, and other conditions associated with an elevated serum bilirubin level.

Kernicterus refers to brain injury as a result of hyperbilirubinemia. It is a serious complication of the neonatal period, generally occurring in the setting of premature birth, neonatal jaundice, and especially hemolytic disease of the newborn. In brief, the immature blood–brain barrier allows free unconjugated bilirubin to enter the brain, leading to encephalopathy. The term *kernicterus* refers to yellowish staining of permanently damaged brain tissue, primarily in the basal ganglia and thalamus. Despite decades of research, the exact pathophysiologic mechanisms by which elevated levels of bilirubin cause brain damage have not been elucidated, although evidence points toward premature programmed cell death (i.e., apoptosis). Unfortunately, most infants die of this condition, and survivors often suffer from cerebral palsy, movement disorders, and mental retardation. Drugs that displace bilirubin from albumin seriously worsen the condition.

If recognized early, treatment with exchange transfusions, phenobarbital (to increase the levels of UDPGT; see previous discussion), and phototherapy ("bili-lights") may prevent these catastrophic consequences. Phototherapy with light in the 450-nm wavelength band is used to treat unconjugated hyperbilirubinemia in infants. Light at this wavelength converts unconjugated lipid-soluble bilirubin into water-soluble photoisomers that can be excreted by the kidneys, thus lowering the bilirubin level.

Infectious and Acquired Hepatitides in Children

Acute HAV infection is usually mild or asymptomatic in children. The prevalence of childhood infection correlates inversely with the quality of sanitation and hygiene. Treatment is supportive, and prevention guidelines parallel those for adults. Universal childhood vaccination is currently recommended in the United States and in many countries around the world.

Globally, HBV infection is a common childhood disease, with vertical transmission from an HBsAg-positive mother to the infant being the most common mechanism of dissemination. Infected blood products and drugs are modes of infection in less developed areas. Features suggesting immune complex disease such as arthritis, fever, papular acrodermatitis (a rash not seen in adults), renal disease, and hematologic complications are more common in children. The incidence of chronic infection is also much higher after neonatal or childhood infection, which has grave long-term consequences because of the late sequelae of cirrhosis and HCC.

Passive immunization with HBIG should be given within 12 hours of birth to children of HBsAg-positive mothers. Active immunization with HBV vaccine should be administered as a series of intramuscular injections at birth and at 1 and 6 months of age. Identical prophylaxis should be given to children of high-risk mothers even if not screened for HBsAg and to children otherwise exposed to HBV. In 1991 the U.S. Public Health Service recommended universal childhood HBV vaccination. Strategies include infant vaccination, possibly with a booster dose in young adulthood, or adolescent vaccination.

Unlike HBV, HCV is less commonly spread vertically, and effective screening of blood products has greatly reduced the risk of childhood infection. HDV may be transmitted by the intrafamily route as a coinfector with hepatitis B.

Acute HEV is especially virulent in adolescents and young adults, with a high mortality in pregnant women. In endemic regions, it is the most common cause of childhood hepatitis and is indirectly a cause of infant mortality.

Many systemic viruses may cause biochemical or clinical hepatitis in the pediatric population, including Epstein–Barr virus, CMV, herpesvirus, and adenovirus. Neonatal hepatitis is part of the so-called *TORCH syndrome* and may be caused by a variety of congenital infections, including CMV, herpesvirus, varicella, *Toxoplasma,* and syphilis. Encephalitis and retinitis may accompany these conditions, with lifelong sequelae. An enlarged liver and elevated levels of transaminases in the newborn should necessitate a search for congenital infection.

Reye syndrome is primarily a disease of children, although adult cases are reported. The syndrome usually occurs shortly after a viral illness such as influenza or chickenpox, and begins with nausea and vomiting rapidly progressing to coma. The exact pathophysiologic mechanism is unknown, but significant mitochondrial dysfunction of hepatocytes occurs. Reye syndrome is characterized by fatty infiltration of the liver with severe hepatic dysfunction, including encephalopathy, coagulopathy, and elevated levels of hepatocellular enzymes. Mortality may be as high as 40%. A strong association with aspirin use during the preceding viral illness has been noted, but other drugs may be causative as well. (Because of the risk of Reye syndrome, aspirin is contraindicated for childhood viral illnesses.) Treatment of Reye syndrome is supportive, and if the child survives, recovery is generally complete. As with other toxic hepatitides, liver transplantation is reserved for irreversible disease.

Congenital Liver Disease

Many heritable diseases of the liver occur in childhood. These may be broadly characterized as enzyme deficiencies affecting multiple organ systems including the liver (e.g., α1-antitrypsin deficiency), disorders of bilirubin metabolism (e.g., Crigler–Najjar syndrome), inborn errors affecting other metabolic pathways, intrahepatic ductopenic conditions, and extrahepatic ductopenia (biliary atresia).

Multisystem Enzyme Deficiencies

α1-Antitrypsin deficiency is an autosomal-recessive condition that commonly affects children and young adults, although it may not become obvious until later in life. α1-Antitrypsin is an enzyme inhibitor found in many tissues that prevents normal enzymes, such as elastase and collagenase, from causing damage to those tissues. Production of this enzyme is genetically controlled by a gene that has many allelic variations (gene types). Although numerous abnormal alleles have been identified, the most common pathologic form, which causes both liver and lung disease, is the protease inhibitor Z variant PiZZ, which produces α1-ATZ protein. This defective α1-antitrypsin protein accumulates in the liver and produces the diagnostic granules seen microscopically, although the exact mechanism of liver damage is unclear. A characteristic centrilobular emphysema, pancreatic insufficiency, and cirrhosis may occur. Treatment by liver transplantation is often precluded by these other problems. Gene therapy for α1-antitrypsin deficiency is an area of active investigation.

Cystic fibrosis (CF) is an autosomal-recessive condition primarily known as a cause of lung disease in children. CF results from one of many mutations in a single large gene on chromosome 7 that encodes the cystic fibrosis transmembrane conductance regulator (CFTR) protein. Clinical disease requires disease-causing mutations in both copies of the *CFTR* gene.

In addition to pulmonary disease, this relatively common genetic condition (1 in 3000 Caucasian births; 1 in 15,000 African American births) may cause pancreatic insufficiency, intestinal obstruction, gallstone disease, neonatal giant cell hepatitis, bile duct obstruction, and biliary cirrhosis. Treatment is directed at complications. Administration of UDCA improves the biochemical indices of liver injury; however, conclusive evidence that the drug halts the progression to cirrhosis is lacking. Gene therapy for CF is in the experimental phase, and treatment strategies for CF lung disease are continually evolving.

Wilson disease and hemochromatosis are single-gene mutation illnesses with significant liver involvement and have been discussed earlier in the chapter.

Disorders of Bilirubin Metabolism

Inherited defects in bile acid metabolism can manifest as impaired synthesis or transport. The first group includes cerebrotendinous xanthomatosis, a steroid hydroxylase deficiency that leads to premature atherosclerosis and encephalopathy. However, manifestations do not generally include liver disease. Children treated with chenodeoxycholic acid have shown marked improvement. Peroxisomes are responsible for β-oxidation in the final steps of bile acid synthesis, and numerous hereditary peroxisomopathies have been described. The most well known of these conditions is X-linked adrenal leukodystrophy (ALD), which is manifested by progressive neurologic dysfunction and adrenal insufficiency. Treatments for this and related conditions remain generally ineffective, but hematopoietic cell transplantation is emerging as the treatment of choice for patients with early stages of ALD, and gene therapies continue to be developed. One interesting, if controversial, therapy for X-linked ALD is dietary therapy with the so-called *Lorenzo's oil*—a mixture of glycerol trioleate and glycerol trierucate.

Disorders of bile acid transport include Gilbert syndrome, a very common (about 10% of the Caucasian population in the United States), entirely benign, autosomal-dominant condition that results in mild unconjugated (indirect) hyperbilirubinemia. It is caused by decreased bilirubin glucuronidation. Awareness of this disorder is important to avoid inappropriate evaluation of these patients.

Crigler–Najjar syndrome is a rare autosomal-recessive disorder marked by significant unconjugated hyperbilirubinemia. In type I Crigler–Najjar syndrome, the near-total absence of bilirubin conjugation results in high levels of unconjugated bilirubin crossing the immature blood–brain barrier. This condition presents shortly after birth, and neonates usually die of kernicterus or suffer irreversible neurologic damage. Liver transplantation after phototherapy and plasma exchange transfusion has been lifesaving in rare instances. In type II Crigler–Najjar syndrome, some conjugating capability exists, and it is enhanced by the administration of phenobarbital. These patients rarely experience bilirubin encephalopathy and can lead normal lives. Treatment consists of phototherapy, phenobarbital administration, and potentially liver transplantation. Gene therapy is an active area of research for Crigler–Najjar syndrome.

Progressive familial intrahepatic cholestasis (PFIC) is a rare autosomal-recessive disorder involving severe jaundice, pruritus, and malabsorption attributable to a defect in bile salt excretion. PFIC type I, or *Byler syndrome,* is caused by a single-gene mutation and traces back to an Amish kindred descended from Jacob Byler. Other genetic defects cause different types of PFIC with similar manifestations. Medical therapy with UDCA is helpful in some children, and biliary diversion procedures have given symptomatic relief to some patients by decreasing the bile acid pool. In the past the disease was uniformly fatal, but liver transplantation has been shown to normalize bile acid synthesis and growth in selected patients.

Other rare disorders of bile salt transport exist and are generally fatal in infancy. These chronic cholestatic diseases, such as North American Indian childhood cirrhosis and cholestasis-lymphedema syndrome (Aagenaes syndrome), are undergoing investigation at

the molecular genetic level and are generally linked to single-gene mutations.

Inborn Errors of Metabolism

A very broad range of enzyme abnormalities resulting from single-gene mutations, generally autosomally recessive, may appear in children. These result in abnormal processing of lipids, glycogen, amino acids and proteins, lipopolysaccharides, and other substances. The pathologic process generally results from excessive accumulation of precursor substances in target organs, such as the brain and spinal cord. The liver is often the primary site of processing and may be the target of toxic accumulation as well. The latter may result in signs of liver disease, including hepatomegaly, elevation in levels of liver enzymes, and jaundice. Inborn errors that manifest in the neonatal period are usually fatal unless immediate treatment is undertaken. Inborn errors presenting in infancy and later years may be amenable to specific therapies. Liver or bone marrow transplantation may be effective for some of these conditions, and many of these conditions are foci of intense research regarding gene therapy. The diagnosis of any inborn error of metabolism should prompt a thorough investigation of family history and the provision of appropriate genetic counseling and testing of family members.

Given the heterogeneity and rarity of these conditions, a detailed discussion of the manifestations and treatment is beyond the scope of this text, and the interested reader is referred to a standard text on the subject.

Intrahepatic Cholestatic Conditions

Congenital intrahepatic cholestasis may be defined as cholestatic liver disease in which the pathologic process is confined to the liver (i.e., the extrahepatic biliary system is normal). One cause of this condition is neonatal hepatitis, which can be idiopathic, viral, or secondary to an inborn error of metabolism. (Viral hepatitis and inborn errors were discussed earlier in this chapter.) The second category of these illnesses includes conditions in which the number of bile ducts is decreased and inadequate to accommodate normal bile metabolism and transport.

Alagille syndrome, or arteriohepatic dysplasia, is the most common form of inherited intrahepatic cholestasis. This autosomal-dominant condition has incomplete penetrance and expressivity and is associated with typical bony and cardiovascular malformations, as well as a paucity of intrahepatic bile ducts. The disease generally progresses slowly; pruritus, hypercholesterolemia, xanthomas, and neurologic complications due to vitamin E deficiency may occur if untreated. Patients may be maintained on UDCA until liver transplantation, which is the current treatment of choice.

Several other conditions manifest a paucity of intrahepatic bile ducts. Some of these arise from defective bile transport mechanisms and are discussed in the Inborn Errors of Metabolism section.

Extrahepatic Cholestatic Conditions (Biliary Atresia)

Extrahepatic ductopenia is often referred to as *biliary atresia*. Biliary atresia or, as some authors prefer, progressive obliterative cholangiopathy, can be either congenital or acquired. The latter occurs in the setting of certain autoimmune illnesses and is one of the principal forms of chronic rejection of a transplanted liver allograft. Biliary atresia is a rather common birth defect, occurring in 1:10,000 to 1:15,000 live births (and is therefore much more common than intrahepatic cholestatic conditions). Distinguishing between this disorder and idiopathic neonatal hepatitis can be challenging, but the liver biopsy findings are usually characteristic. Infants and children with biliary atresia have progressive cholestasis with all the usual concomitant features: pruritus, malabsorption with growth retardation, fat-soluble vitamin deficiencies, hyperlipidemia, and eventually cirrhosis with portal hypertension. Some children have recurrent episodes of bacterial cholangitis as well. A cholangiogram should be obtained to assess the possibility of a correctable obstruction. The Kasai procedure is often performed to create a hepatoportoenteric connection, which may allow adequate bile drainage. This procedure is not usually curative but ideally does "buy time" for growth of the child until an appropriate time for liver transplantation.

KEY POINTS

- Cirrhosis is the irreversible end stage of many different hepatic injuries. The liver is fibrotic, scarred, and nodular. Symptoms of cirrhosis are due to hepatocellular failure and portal hypertension.
- Biliary cirrhosis is associated with chronic bile duct obstruction with resultant accumulation of bile in the liver. Gallstones and extrahepatic and intrahepatic bile duct inflammation are common causes.
- Alcoholic cirrhosis is associated with chronic alcohol ingestion, which may precipitate fatty liver, hepatitis, and, finally, cirrhosis.
- The liver is subject to damage because of its role in storing and detoxifying potentially injurious substances. Metal storage diseases are genetic disorders in which excessive minerals are absorbed and subsequently deposited in the liver. Hemochromatosis is characterized by excessive iron absorption and is manifested by elevated serum ferritin and iron levels. Phlebotomy is the usual treatment. Wilson disease is due to excessive accumulation of copper in the liver and other organs. Copper chelators are effective in preventing liver damage.
- Acetaminophen is converted to a toxic metabolite in the liver that is normally rapidly detoxified by liver enzymes. In acetaminophen overdose, the detoxification reaction may be overwhelmed and liver necrosis results. Treatment is with acetylcysteine.
- Liver abscesses are suspected in patients with fever, nausea, vomiting, and right upper quadrant pain. Ascending biliary infection, abdominal infections transported by the portal vein, and direct extension of infection from neighboring structures are usual sources of infection. Antibiotics and drainage are commonly prescribed.
- The liver commonly sustains injury during penetrating and blunt trauma to the abdomen. Because the liver is highly vascular, trauma may produce extreme blood loss and hemorrhagic shock. Liver trauma is manifested by abdominal tenderness, distention, guarding, and rigidity.
- In the United States cancer of the liver is usually metastatic and rarely primary. Tumors of the esophagus, stomach, colon, rectum, breast, and lung commonly seed in the liver. Hepatocellular carcinoma (HCC) is more common in other parts of the world, but its incidence is increasing in the United States, with HBV and HCV important contributing factors.

LIVER DISEASES AND GERIATRIC CONSIDERATIONS

Liver size and blood flow decrease with aging, but this observation has little functional significance. Drugs whose metabolism is primarily related to hepatic blood flow and drugs processed by the mixed-function oxidase system (cytochromes) may have a prolonged serum half-life requiring careful monitoring and dose adjustment. Routine blood test results of liver-related enzymes are not changed by aging.

In the United States HCC is usually the result of years of injury from alcohol or chronic viral hepatitis and is therefore more often seen in older people. The prognosis is unfortunately dismal (see discussion earlier in the chapter).

Ischemic hepatitis is usually associated with underlying cardiovascular disease and episodes of hypotension, as during surgery or sepsis, and

is more common in older patients. Typically, the levels of serum transaminases rise rapidly and the PT is prolonged. Recovery may be rapid, but the prognosis depends on the severity of the underlying disorder. Ischemic hepatitis alone is rarely a cause of death in such patients. Right-sided heart failure may result in passive hepatic congestion with ascites and (rarely) liver failure. In developing countries, constrictive pericarditis and uncorrected valvular heart disease attributable to rheumatic fever are still significant causes of intractable ascites and should be considered in the differential diagnosis.

Metabolic liver diseases rarely present in the geriatric population. Hemochromatosis in women often occurs after menopause and may present as new-onset diabetes mellitus, heart failure, arthritis, cirrhosis, or HCC. Typically, autoimmune liver diseases are seen in young to middle-aged people. However, autoimmune chronic active hepatitis may be a cause of "cryptogenic" cirrhosis in older women. Primary biliary cirrhosis is uncommon in later life. Primary sclerosing cholangitis may afflict older persons with long-standing ulcerative colitis, even those who had colectomies many years, even decades, earlier.

Because alcohol abuse generally starts early in life, older patients bear the cumulative injury of years of exposure and are likely to show signs of advanced liver disease. Treatment is the same as in younger patients, but attention must be given to the older patients' social circumstances and intercurrent medical problems. Symptoms and signs of alcohol intoxication and hepatic encephalopathy may be confused with senile dementia and made worse by concomitant drug use (e.g., minor tranquilizers, opiates).

The diagnosis of acute viral hepatitis may be more difficult in older people because of nonspecific symptoms in mild cases and decreased clinical suspicion. Acute HAV infection is less common as a result of the higher incidence of immunity, but it may be more severe with higher mortality rates in the elderly. The end results of HBV and HCV infection often become evident in elderly persons, and in the United States the increased prevalence of persons living with these viruses will mean more elderly patients with liver disease in coming decades.

Older patients with chronic HBV and HCV infection may be treated with the standard treatments, but comorbid medical problems, intolerance to the side effects of treatment, and advanced liver disease may preclude treatment in many persons. The indications for liver transplantation do not change with advancing years, and no arbitrary age limits have been set on transplantation. However, there may be subtle barriers in effect, and the allocation of organs remains a highly controversial issue. Each patient must be evaluated individually regarding the propriety of transplantation and the likelihood of success.

SUMMARY

Disorders of the liver are diverse and complex. Because the liver is vital to most life processes, even mild disorders can cause life-threatening alterations. Health care professionals need a good understanding of hepatobiliary anatomy and physiology to appreciate the effects of these disorders on patients.

Many liver disorders are the consequence of lifestyle choices such as alcoholism and drug abuse. Health care professionals are in a position to explain the risks of detrimental lifestyles and their relationship to liver diseases so as to prevent occurrence of these diseases.

Medical treatment entails the use of drugs from many different classes. Because the liver is central to the metabolism of many of these drugs, their use requires special attention. Before any drug is given to a patient with liver disease, it is essential to become completely familiar with it by consulting a good pharmacology text or drug information source.

RESOURCES

Liver Physiology and Pathophysiology

Feldman M, Friedman LS, Brandt LJ: *Sleisenger and Fordtran's gastrointestinal and liver disease*, ed 10, Philadelphia, 2015, Saunders Elsevier.

Hall JE: *Guyton and Hall textbook of medical physiology*, ed 13, Philadelphia, 2015, Saunders Elsevier.

Johnson LR: *Gastrointestinal physiology*, ed 8, Philadelphia, 2013, Mosby.

Kamisako T, et al: Recent advances in bilirubin metabolism research: the molecular mechanism of hepatocyte bilirubin transport and its clinical relevance. *J Gastroenterol* 35(9):659–664, 2000.

Moore KL, Persaud TVN: *The developing human: clinically oriented embryology*, ed 10, Philadelphia, 2015, Saunders.

Moore KL, Persaud TVN, Shiota K: *Color atlas of clinical embryology*, Philadelphia, 2000, Saunders.

Liver Failure and Portal Hypertension

Arroyo V, Guevara M, Gines P: Hepatorenal syndrome in cirrhosis: pathogenesis and treatment. *Gastroenterology* 122(6):1658–1676, 2002.

Bosch J, et al: Recombinant factor VIIa for upper gastrointestinal bleeding in patients with cirrhosis: a randomized, double-blind trial. *Gastroenterology* 127(4):1123–1130, 2004.

Bosch J, Garcia-Pagan JC: Prevention of variceal rebleeding. *Lancet* 361(9361):952–954, 2003.

Butterworth RF: The astrocytic ("peripheral-type") benzodiazepine receptor: role in the pathogenesis of portal-systemic encephalopathy. *Neurochem Int* 36(4-5):411–416, 2000.

Escorsell A, et al: TIPS versus drug therapy in preventing variceal rebleeding in advanced cirrhosis: a randomized controlled trial. *Hepatology* 35(2):385–392, 2002.

Garcia-Tsao G, et al: Prevention and management of gastroesophageal varices and variceal hemorrhage in cirrhosis. *Hepatology* 46:922, 2007.

Gournay J, et al: Isosorbide mononitrate and propranolol compared with propranolol alone for the prevention of variceal rebleeding. *Hepatology* 31(6):1239–1245, 2000.

Jalan R, et al: Moderate hypothermia prevents cerebral hyperemia and increase in intracranial pressure in patients undergoing liver transplantation for acute liver failure. *Transplantation* 75:2034, 2003.

Schrier RW, Gurevich AK, Cadnapaphornchai MA: Pathogenesis and management of sodium and water retention in cardiac failure and cirrhosis. *Semin Nephrol* 21(2):157–172, 2001.

Shils ME: *Modern nutrition in health and disease*, ed 10, Philadelphia, 2006, Lippincott Williams & Wilkins.

Watt K, Uhanova J, Minuk GY: Hepatorenal syndrome: diagnostic accuracy, clinical features, and outcome in a tertiary care center. *Am J Gastroenterol* 97(8):2046–2050, 2002.

Viral Hepatitis

A comprehensive immunization strategy to eliminate transmission of hepatitis B virus infection in the United States: recommendations of the Immunization Practices Advisory Committee (ACIP). *MMWR Morb Mortal Wkly Rep* 54(RR16):1–23, 2005.

Braconier JH, Wennerholm S, Norrby SR: Comparative immunogenicity and tolerance of Vaqta and Havrix. *Vaccine* 17(17):2181–2184, 1999.

Chopra S, Muir AJ *Treatment for chronic hepatitis C virus genotype 1. In Chopra S, Lamont JT, GI editors, UpToDate*. Available at: www.uptodate.com. (Accessed 19 November 2015).

Czaja AJ, Freese DK: Diagnosis and treatment of autoimmune hepatitis: American Association for the Study of Liver Disease recommendations. *Hepatology* 36(2):479–497, 2002.

General recommendations on immunization: recommendations of the Advisory Committee on Immunization Practices (ACIP) and the American Academy of Family Physicians (AAFP). *MMWR Morb Mortal Wkly Rep* 51(RR02):1–36, 2002.
Gorbach SL, Bartlett JG, Blacklow NR: *Infectious diseases*, ed 3, Philadelphia, 2004, Lippincott Williams & Wilkins.
Kamar N, et al: Ribavirin for chronic hepatitis E virus infection in transplant recipients. *NEJM* 370(12):1111–1120, 2014.
Lok ASF: *Standard and pegylated interferon for chronic hepatitis B virus infection. In Chopra S, Lamont JT, GI editors, UpToDate*. Available at: www.uptodate.com. (Accessed 19 November 2015).
Marcellin P: Advances in therapy for chronic hepatitis B. *Semin Liver Dis* 22(Suppl 1):33–36, 2002.
Updated US: Public Health Service guidelines for the management of occupational exposures to HBV, HCV, and HIV and recommendations for postexposure prophylaxis. *MMWR Morb Mortal Wkly Rep* 50(RR11):1–42, 2001.
Wai CT, Lok AS: Treatment of hepatitis B. *J Gastroenterol* 37(10):771–778, 2002.

Noninfectious, Toxic Hepatitis, and Cancer

Adams PC, et al: Hemochromatosis and iron-overload screening in a racially diverse population. *N Engl J Med* 352(17):1769–1778, 2005.
Brewer GJ, et al: Treatment of Wilson disease with ammonium tetrathiomolybdate, III: initial therapy in a total of 55 neurologically affected patients and follow-up with zinc therapy. *Arch Neurol* 60(3):379–385, 2003.
Crosse KI, Anania FA: Alcoholic hepatitis. *Curr Treat Options Gastroenterol* 5(6):417–423, 2002.
Godara H, et al: *The Washington manual of medical therapeutics*, ed 34, Philadelphia, 2013, Lippincott Williams & Wilkins.
Lee YM, Kaplan MM: Management of primary sclerosing cholangitis. *Am J Gastroenterol* 97(3):528–534, 2002.
Olynyk JK, et al: Evolution of untreated hereditary hemochromatosis in the Busselton population: a 17-year study. *Mayo Clin Proc* 79(3):309–313, 2004.
O'Shea RS, McCullough AJ: Treatment of alcoholic hepatitis. *Clin Liver Dis* 9(1):103–134, 2005.
Ostapowicz G, et al: Results of a prospective study of acute liver failure at 17 tertiary care centers in the United States. *Ann Intern Med* 137(12): 947–954, 2002.
SEER stat fact sheets: liver and intrahepatic bile duct cancer. Available at: http://seer.cancer.gov/statfacts/html/livibd.html. (Accessed 20 November 2015).
Yu AS, Keeffe EB: Management of hepatocellular carcinoma. *Rev Gastroenterol Disord* 3(1):8–24, 2003.

Liver Transplantation

Bezerra JA, Balistreri WF: Cholestatic syndromes of infancy and childhood. *Semin Gastrointest Dis* 12(2):54–65, 2001.
Clark NM, Cotler SJ: *Infectious complications in liver transplantation. In Chopra S, Lamont JT, GI editors, UpToDate*. Available at: www.uptodate.com. (Accessed 23 November 2015).
Keeffe EB: Liver transplantation: current status and novel approaches to liver replacement. *Gastroenterology* 120(3):749–762, 2001.
Shouval D, Samuel D: Hepatitis B immune globulin to prevent hepatitis B virus graft reinfection following liver transplantation: a concise review. *Hepatology* 32:1189–1195, 2000.
Unique patient populations: Patients who develop recurrent HCV infection post-liver transplantation. In Guidance: Recommendations for testing, managing, and treating hepatitis C. Available at: http://www.hcvguidelines.org/full-report/unique-patient-populations-patients-who-develop-recurrent-hcv-infection-post--liver. (Accessed 23 November 2015).
Wiesner RH, et al: Recent advances in liver transplantation. *Mayo Clin Proc* 78(2):197–210, 2003.

Liver Disease in Children

Balistreri WF: Intrahepatic cholestasis. *J Pediatr Gastroenterol Nutr* 35(Suppl 1):S17–S23, 2002.
Bratlid D: Criteria for treatment of neonatal jaundice. *J Perinatol* 21(Suppl 1):S88–S92, discussion S104–S107, 2001.
Casteels-Van Daele M, et al: Reye syndrome revisited: a descriptive term covering a group of heterogeneous disorders. *Eur J Pediatr* 159(9):641–648, 2000.
Hansen TW: Mechanisms of bilirubin toxicity: clinical implications. *Clin Perinatol* 29(4):765–778, 2002.
Ishak KG: Inherited metabolic diseases of the liver. *Clin Liver Dis* 6(2):455–479, 2002.
Ismail H, et al: Treatment of progressive familial intrahepatic cholestasis: liver transplantation or partial external biliary diversion. *Pediatr Transplant* 3(3):219–224, 1999.
Leung DH, Horowitz D *Cystic fibrosis: Overview of the treatment of lung disease.* In Chopra S, Lamont JT, GI editors, *UpToDate*. Available at: www.uptodate.com. (Accessed 23 November 15).
McGuinness MC, Wei H, Smith KD: Therapeutic developments in peroxisome biogenesis disorders. *Expert Opin Investig Drugs* 9(9):1985–1992, 2000.
Sherlock S, Dooley J: The liver in infancy and childhood. In *Diseases of the liver and biliary system*, ed 10, Oxford, 1997, Blackwell.
Suzuki Y, et al: The clinical course of childhood and adolescent adrenoleukodystrophy before and after Lorenzo's oil. *Brain Dev* 23(1):30–33, 2001.
Yeager AM: Allogeneic hematopoietic cell transplantation for inborn metabolic diseases. *Ann Hematol* 81(Suppl 2):S16–S19, 2002.

39

Endocrine Physiology and Mechanisms of Hypothalamic-Pituitary Regulation

Jacquelyn L. Banasik

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- How does the lipid or water solubility of hormones affect their transport in the circulation?
- What are the general mechanisms of action of lipid-soluble and water-soluble hormones on target cells?
- How do target cells regulate their responsiveness to endocrine hormones?
- How do feedback mechanisms control the secretion of hormones?
- What are the anterior and posterior pituitary hormones, their target tissues, and their negative feedback mechanisms?
- How are thyroid and steroid hormones synthesized?
- What are the normal actions on target cells of antidiuretic hormone, growth hormone, thyroid hormone, and steroid hormones?

CHAPTER OUTLINE

Hormone Structure and Action, 784
Chemical Structure of Hormone Classes, 784
Mechanisms of Hormone Action, 784
Hormones With Cell Membrane Receptors, 784
Amplification of Hormone Activity, 786
Hormones With Intracellular Receptors, 786
Hormone Regulation, 786
Hormone Synthesis, Secretion, and Metabolism, 786
Factors Affecting Hormone Secretion, 787
Feedback Control of Secretion, 787
Hormone Metabolism and Excretion, 787
Pharmacologic Hormone Concentrations, 787
Regulation of Receptor Responses, 787
Receptor Specificity and Affinity, 787
Receptor Down-Regulation and Up-Regulation, 787
Permissiveness, 788
Hormone Agonists and Antagonists, 788
Hypothalamic-Pituitary Endocrine System, 788
Hormones of the Posterior Pituitary Gland, 788
Antidiuretic Hormone, 789
Oxytocin, 790
Hormones of the Hypothalamus and Anterior Pituitary Gland, 790
Growth Hormone, 790
Prolactin, 792
Gonadotropins, 792
Thyroid Stimulating Hormone, 793
Adrenocorticotropic Hormone, 793
Thyroid Hormones, 793
Thyroid Hormone Synthesis and Secretion, 793
Thyroid Action on Target Cells, 793
Steroid Hormones, 794
Steroid Hormone Synthesis and Secretion, 794
Steroid Action on Target Cells, 795
Categories of Endocrine Disease, 797
Hyposecretion, 797
Hypersecretion, 797
Hyporesponsiveness, 797

The endocrine system is composed of cells and organs that are specialized to synthesize and secrete hormones into the bloodstream to act at distant target cells. The nervous and endocrine systems are closely integrated, and many of their actions are coordinated at the level of the hypothalamus. Endocrine systems are particularly suited for regulating complex functions that involve numerous tissues and organs such as growth, metabolism, fluid balance, responses to stress, and reproduction. It is now appreciated that all cells in an organism are involved in cell-to-cell communication, not just the specialized endocrine cells; therefore the term *classical* endocrine systems is sometimes used to designate specialized endocrine glands. This chapter describes general principles of endocrine communication systems and explores the details of hypothalamic-pituitary regulation of antidiuretic hormone, oxytocin, growth hormone, prolactin, gonadotropins, thyroid hormone, and corticosteroid hormones. Disorders of these systems are discussed in Chapter 40. Reproductive function and disorders can be found in Unit IX.

HORMONE STRUCTURE AND ACTION

An endocrine hormone may be defined as a blood-borne chemical messenger that has an effect on target cells anatomically distant from the secreting cell (Fig. 39.1). Chemical messengers may also act in a paracrine or autocrine fashion; the hormone molecule is secreted by one cell and affects either adjacent cells or the original secreting cell. Paracrine and autocrine actions of hormones are not traditionally included in endocrine system physiology; however, the same signaling molecules can have autocrine, paracrine, and distant actions. As an example, estrogen acts locally within the ovary potentiating maturation of ova but is required systemically for outward female sexual differentiation. Secretion of hormones into the bloodstream by neurons may be termed neurocrine signaling, a process common to the hypothalamic-pituitary communication system.

Chemical Structure of Hormone Classes

Hormones may be classified according to chemical structure as peptides, tyrosine-derived catecholamines and thyroid hormones, and steroids (Box 39.1). The great majority of endocrine hormones are peptides (small proteins) and are water soluble and easily transported through the circulation to target cells. The catecholamines (dopamine, epinephrine, and norepinephrine) are also water soluble. Water-soluble hormones usually travel free in solution in the plasma, although some are partially protein bound. Water-soluble hormones are unable to cross the plasma membrane to enter cells and therefore must exert their actions by binding to receptors located on the surface of target cell membranes.

Thyroid hormones (T_3 and T_4) are derived from tyrosine amino acids to which iodine has been attached. The steroid hormones are derived from cholesterol. Thyroid and steroid hormones require transport proteins (globulin) to convey them through the circulation because they are poorly soluble in the blood. In some cases these transport proteins are specialized to carry a certain hormone (thyroxine-binding globulin, cortisol-binding globulin), but hormones can also be carried by nonspecific proteins, such as albumin. Transport proteins are manufactured by the liver. At the target cell, the hormone detaches from the transport protein and moves through the cell membrane to activate intracellular receptors in the cytoplasm or nucleus.

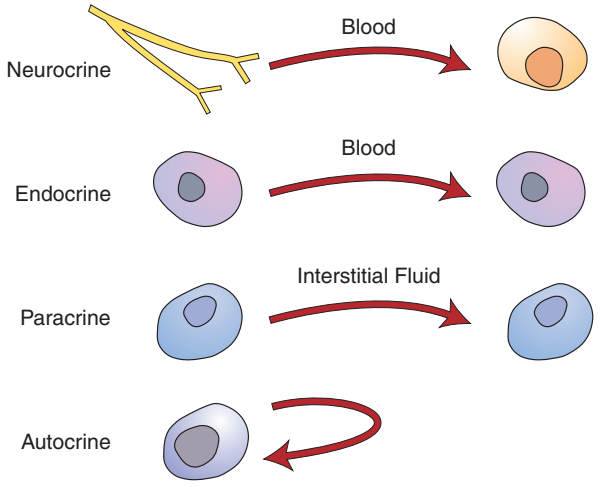


FIG 39.1 Terminology and comparison of cell-to-cell signaling pathways. *IF,* Interstitial fluid.

Mechanisms of Hormone Action

Hormones With Cell Membrane Receptors

Hormones exert their actions by binding to target cell receptor proteins. The target cell receptors for water-soluble hormones have a hormone-binding site located on the external portion of a specific cell-surface receptor. Once hormone-receptor binding takes place, a change in the conformation of the receptor protein conveys a signal to the interior of the cell. Some receptors generate second messengers, such as cyclic adenosine monophosphate (cAMP), within the cell, whereas others become activated enzymes. The intracellular activities occurring after hormone-receptor binding may include changes in cell membrane permeability; activation or inactivation of enzymes; cellular maintenance, growth, and differentiation; protein synthesis; and genetic expression via RNA and DNA synthesis in the cell nucleus. A more detailed explanation of receptor mechanisms can be found in Chapter 3.

G-Protein–Coupled Receptors. Hormones can be described as "first messengers"; the hormone carries a message from the secreting cell to the target cell. A second messenger is generated within the cell in response to the first message. The term *second messenger* is usually applied to the products generated by a class of receptors called *G-protein–coupled*

BOX 39.1 Chemical Classification of Hormones

Water Soluble

Peptides
- Adrenocorticotropic hormone (ACTH)
- Angiotensin II
- Atrial natriuretic hormone
- Antidiuretic hormone (vasopressin) (ADH)
- Calcitonin
- Cholecystokinin
- Follicle-stimulating hormone (FSH)
- Growth hormone (GH)
- Glucagon
- Hypothalamic-releasing hormones
- Insulin
- Insulin-like growth factor-1 (IGF-1)
- Luteinizing hormone (LH)
- Oxytocin
- Parathyroid hormone (PTH)
- Prolactin
- Secretin
- Thyroid-stimulating hormone (TSH)

Tyrosine-Derived Amines
- Dopamine
- Epinephrine
- Norepinephrine

Lipid Soluble

Tyrosine-Derived Thyroid Hormones
- Triiodothyronine (T_3)
- Thyroxine (T_4)

Steroids
- Aldosterone
- Cortisol
- Corticosterone
- Estrogen
- Progesterone
- Testosterone
- Vitamin D

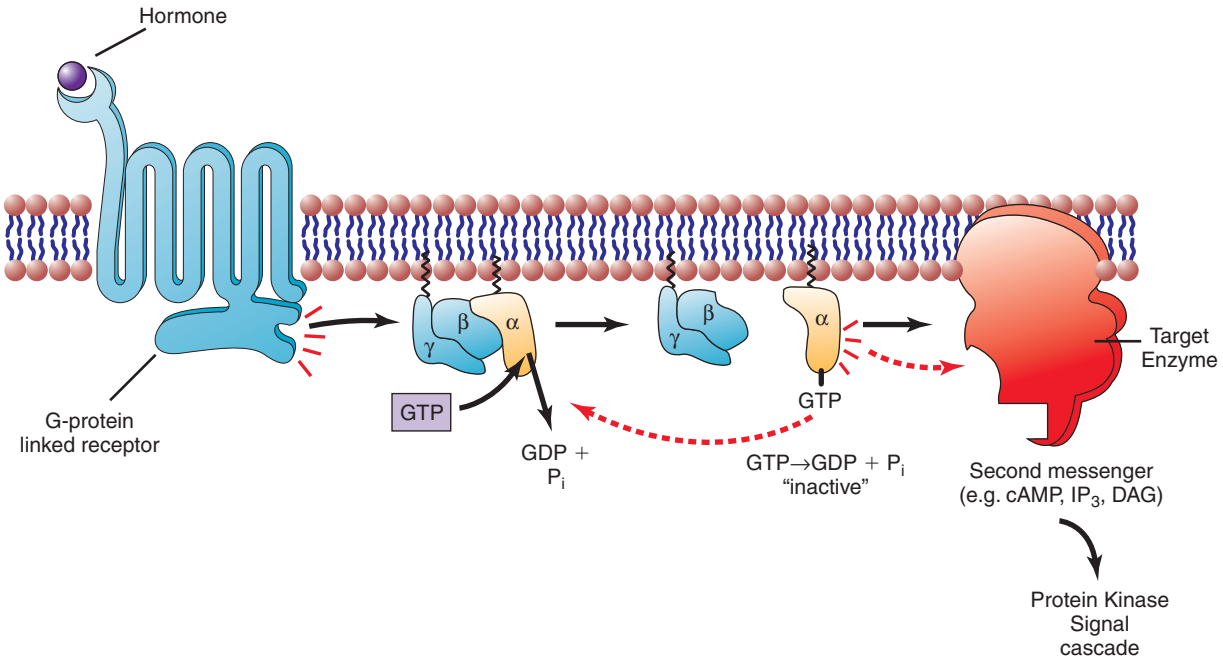


FIG 39.2 Mechanism of G-protein–linked receptor signaling in endocrine cells. Activation of the receptor by a hormone on the outside of the cell changes the conformation of the receptor and activates its target G-proteins. Common G-proteins in endocrine systems are G_s (which activates adenylyl cyclase to produce cAMP) and G_q (which activates phospholipase C to produce DAG and IP_3).

(or *G-protein–linked*) *receptors*. These receptors undergo a conformational change upon binding to a hormone that allows them to interact with a class of proteins, called *G-proteins*, located on the inner side of the cell membrane (Fig. 39.2). G-Proteins have three subunits and when activated by the receptor, one of these subunits (α) binds to a guanine triphosphate (GTP) molecule and then dissociates from the other subunits (β, γ). The activated α subunit is able to move along the inner membrane and interact with target enzymes, inducing them to produce the second messenger. Several different forms of G-proteins have different target enzymes and produce different sets of second messengers (see Chapter 3). The most important of these for endocrine systems are the G_s and the G_q subtypes. Hormones that bind to receptors that are linked to G_s increase the production of cAMP within the cell. The usual downstream target of cAMP is an enzyme called *protein kinase A*. Kinases are enzymes that attach phosphates (phosphorylation) to target proteins to change their activity. Enzymes within the cell called *phosphorylases* work to remove the attached phosphates. Phosphorylation–dephosphorylation is a common strategy for controlling enzyme activity in cells. Activation of protein kinase A causes varied cellular effects depending on the cell type. Important examples of hormones that are linked to increased cAMP production through activation of the G_s pathway include adrenocorticotropic hormone (ACTH), thyroid-stimulating hormone (TSH), follicle-stimulating hormone (FSH), luteinizing hormone (LH), epinephrine (β receptors), parathyroid hormone, antidiuretic hormone (ADH), and glucagon.

The G_q pathway is linked to production of two second messengers: diacylglycerol (DAG) and inositol trisphosphate (IP_3). Norepinephrine and epinephrine can activate the G_q pathway by binding to α_1 receptors on target cells. DAG and IP_3 work together to increase the activity of protein kinase C, which then phosphorylates downstream target proteins to change cell behavior.

Protein Kinase Receptors. A second class of surface receptors important in endocrine signaling is the protein kinase receptor family. These receptors either become activated kinases themselves or are associated with cytoplasmic kinases on their inner domain (Fig. 39.3). Both types of receptors phosphorylate their target proteins upon binding of the hormone to an external binding site. An important example of a receptor kinase is the insulin receptor that phosphorylates tyrosine amino acids in its targets and activates several different intracellular signaling cascades (Chapter 41). Growth hormone (GH) and prolactin (PRL) receptors are associated with Janus kinase (JAK) on the cytoplasmic side and become activated when the hormone binds to the outside of the receptor, causing a conformational change that pulls two JAK enzymes closer together. The JAK enzymes become activated and phosphorylate their target proteins, initiating a signaling cascade (see Fig. 39.3).

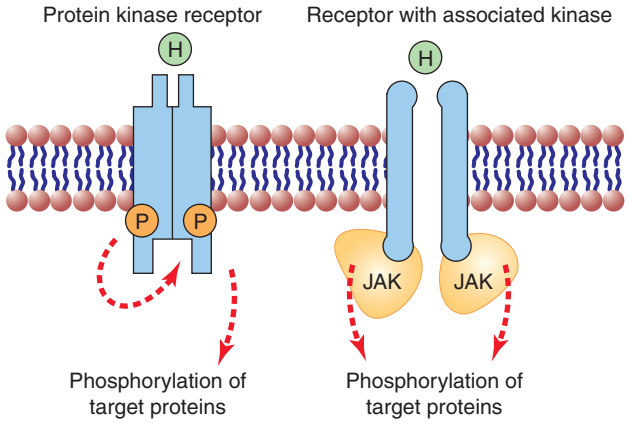


FIG 39.3 Some endocrine receptors have intrinsic kinase activity, whereas others have kinases associated with their inner domain. In either case, hormone binding to the outer part of the receptor triggers a conformational change in the receptor that initiates kinase activities. Kinases phosphorylate target proteins and alter their activity.

Amplification of Hormone Activity

The process of intracellular activation by secondary signals occurs via a cascade effect. Progressively larger numbers of chemical reactions occur at each step, so that activation of one receptor at the cell surface can activate numerous G-proteins. Each activated G-protein may interact with 100 or more target enzyme molecules, such as adenylyl cyclase. Each activated enzyme produces many molecules of a second messenger (cAMP), and each second-messenger molecule activates many molecules of protein kinase, and so forth. This mechanism of signal *amplification* explains why minute amounts of circulating hormone cause significant and rapid cellular and systemic effects.

Hormones With Intracellular Receptors

Receptors for thyroid and steroid hormones are located in the cytoplasm or the nucleus of the target cell (Fig. 39.4). Transport of thyroid hormone across the cell membrane is mediated by several protein carriers in the membrane. Steroid hormones are able to pass directly across the plasma membrane to interact with intracellular receptors. However, they may also bind cell-surface receptors as evidenced by the discovery that estrogen can bind to a G-protein–coupled receptor on some cells. These G-protein receptors achieve a more rapid change in cell behavior than those associated with intracellular receptors. Binding of the hormones to intracellular receptors causes a change in affinity of the receptor for binding sites on DNA in the cell nucleus. Gene expression is changed by binding of the hormone–receptor complex to specific DNA-binding sites. The events that result from the interaction between nuclear DNA and the receptor include messenger RNA transcription, processing, and translation into specific proteins (see Chapter 5 for a review of protein synthesis). In general, the onset of action of thyroid and steroid hormones is slow compared with that of the water-soluble hormones described previously, and there is no amplification cascade. Changes in DNA transcription and protein synthesis occur over hours to days.

KEY POINTS

- Endocrine hormones are chemical messengers that travel via the bloodstream to exert effects on target cells distant from the secreting glands. Hormones regulate complex functions, including reproduction, growth and development, fluid homeostasis, and metabolism.
- Hormones may be classified according to chemical structure as water soluble or lipid soluble. Receptors for water-soluble hormones are located on the cell surface (catecholamines, peptides), and most thyroid and steroid hormone receptors are located intracellularly.
- Activation of G-protein–coupled receptors on the cell surface leads to the production of second messengers within the cell and alteration in intracellular functions. Most endocrine hormones affect their targets through G-protein–coupled receptors.
- Some endocrine hormones, such as insulin, growth hormone, and prolactin, interact with receptors on the cell surface that are protein kinases or directly associated with kinases. Activation of kinase receptors results in phosphorylation of target proteins, which changes their activity and initiates intracellular signaling cascades.
- Steroid and thyroid hormones are carried in the circulation by transport proteins. Thyroid and steroid hormones are able to diffuse directly through the membrane. Thyroid hormone also enters via carrier-mediated facilitated diffusion. Their target receptors are in the cytoplasm or nucleus. Some steroid receptors are associated with the cell membrane. Intracellular hormone-receptor complexes bind to specific targets on nuclear DNA and alter gene expression.

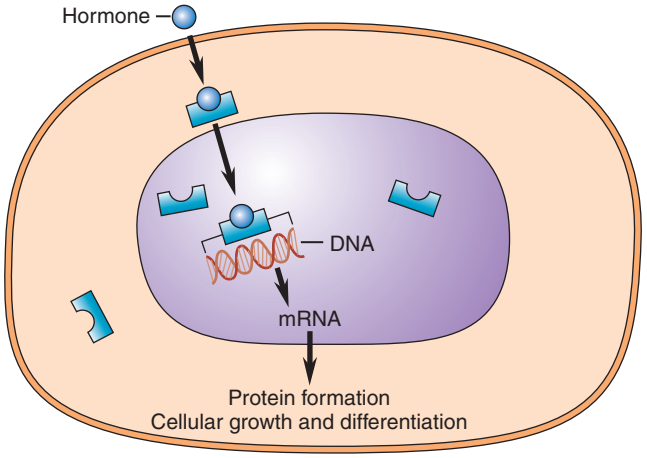


FIG 39.4 Thyroid and steroid hormones are transported through the plasma membrane and bind to intracellular receptors (located in the cytoplasm or nucleus). The hormone–receptor complexes are DNA-regulatory proteins that regulate messenger RNA transcription, processing, and translation into proteins.

HORMONE REGULATION

Hormone Synthesis, Secretion, and Metabolism

Hormone synthesis and secretion may take place in several different cell types, but one site is generally considered to be the primary endocrine tissue and is responsive to feedback about the need for the hormone or its effects. For example, endothelial cells throughout the vascular system produce small amounts of angiotensin II (AII); however, the primary endocrine control is through the renin–angiotensin–aldosterone cascade in which specialized cells in the kidney trigger AII production in response to changes in extracellular volume and pressure.

Most endocrine hormones are polypeptides manufactured on the rough endoplasmic reticulum and stored in vesicles within the cells. A review of protein synthesis can be found in Chapter 5. The initial forms of the polypeptide hormones are usually larger molecules called *prohormones* or *preprohormones*. These must be cleaved by specific enzymes to release the active form of the hormone. Sometimes this cleavage occurs after the hormone is packaged into vesicles and the fragments are released into the bloodstream along with the hormone. In some cases, such as with insulin secretion, this is clinically useful as a marker of hormone synthesis. Insulin and its C-peptide fragment are produced in a 1 : 1 ratio, and the concentration of C-peptide, which is more stable in the circulation, can be used as a measure of insulin production.

Because they are water soluble, peptide hormones can be contained within the lipid bilayer of the vesicles and stored until a trigger results in exocytosis of the hormone into the extracellular space. Catecholamines are also water soluble and stored in vesicles until they are released by exocytosis. However, they are formed by enzymes within the cytoplasm that begin with tyrosine and through a series of steps convert it first to dopamine; depending on the cell type, dopamine may be the final hormone, or it can be converted to norepinephrine and then to epinephrine.

Steroid and thyroid hormones are not stored within vesicles because they are lipid soluble and might leak from the vesicle. The strategy for synthesis and storage is much different for each of these hormones. When a steroid hormone is needed, increased production of the hormone closely precedes hormone release into the circulation. Steroid hormones are formed on demand from cholesterol that is either stored in the cell or retrieved from the circulating lipoproteins. The stimuli that trigger steroid secretion induce the enzymes in the pathway to synthesize the

necessary hormones from cholesterol. Once formed, steroid hormones can simply diffuse through the plasma membrane and into the circulation. The strategy for trapping thyroid hormones within the thyroid gland until they are needed is quite different. Hormone synthesis precedes secretion by weeks or months. Triiodothyronine (T_3) and thyroxine (T_4) are synthesized in the thyroid follicle, bound to a protein called *thyroglobulin.* Thyroglobulin is a large water-soluble protein that is trapped within the thyroid follicle. Secretion occurs via cleavage of the thyroid hormone from thyroglobulin in response to systemic needs determined by the hypothalamus and the pituitary gland. Free T_3 and T_4 are lipid soluble and released from the thyroid cells through membrane diffusion and carrier transport proteins. Thyroid and steroid synthesis and release are detailed later in this chapter.

Factors Affecting Hormone Secretion

Secretion of endocrine hormones is dependent on the interplay of many factors. Most of the hormones controlled by the anterior pituitary gland are secreted in cyclical patterns. These cycles are regulated by the hypothalamus and may occur in response to feeding–fasting cycles, light–dark cycles, sleep–wake cycles, or in a 24-hour (circadian) period. Cycles also occur over longer periods (e.g., the 28-day menstrual cycle) or over years (e.g., the hormones that control reproductive differentiation and maturity). Acute systemic needs or stressors can partially override cyclical patterns and modify hormone secretion. An example of this is activation of the stress response that alters the normal circadian pattern of cortisol secretion. Knowledge of the cyclical nature of hormone release is important when comparing measured serum concentrations to normal ranges and when attempting to provide hormone replacement therapy that mimics the normal physiologic pattern.

Feedback Control of Secretion

The most common mechanism regulating hormone production and secretion is *negative feedback* control in which some aspect of the secreted hormone is sensed and regulates further secretion. An example of negative feedback is the interaction between the hypothalamic-pituitary system and the respective hormone-releasing glands controlled by this system. For example, thyrotropin-releasing hormone (TRH) is secreted by the hypothalamus when the central nervous system senses an inadequate supply of thyroid hormones in the circulation. TRH stimulates the synthesis and release of TSH from specific cells in the anterior pituitary gland; TSH then stimulates the synthesis and release of the thyroid hormones T_3 and T_4 from the thyroid follicle. The resulting increase in plasma thyroid hormone concentration exerts an inhibitory effect, or negative feedback, on the release of TSH by the anterior pituitary gland and TRH from the hypothalamus. The negative feedback loop is a mechanism for maintaining hormone activity within a normal range, or *set point,* of normal activity. The set point is genetically determined and influenced by age, gender, circadian cycles, and current internal and environmental conditions. The regulatory influences are multiple and complex, but often a particular variable is influenced by the hormone's activity and provides most of the feedback regulation. These feedback systems are detailed for the hormones of the pituitary system in later sections of this chapter. Knowledge of feedback controls is helpful when evaluating patients for endocrine diseases (see Chapter 40).

Hormone Metabolism and Excretion

The plasma concentration of a hormone depends not only on the rate of synthesis and release of the hormone, but also on how rapidly the hormone is metabolized and excreted. Like other compounds, hormones are frequently degraded and excreted by the liver and kidneys. Water-soluble hormones may be excreted in the urinary filtrate. Lipid-soluble hormones, which are bound to plasma proteins and stored in adipose tissue, are less readily metabolized and remain in the circulation for a more prolonged period. In addition to metabolism by the kidneys and liver, hormones are often degraded by the target cell after binding to receptors. *Half-life* is a term used to describe the duration of hormone activity in the circulation and may be expressed in minutes, hours, or days. The half-life is the time for a hormone to reach half of its original concentration in the blood and is influenced by the rate of uptake by cells, degradation, and excretion.

Pharmacologic Hormone Concentrations

It is important to differentiate between physiologic and pharmacologic hormone activity. Physiologic hormone concentrations are extremely low in most cases, and pharmacologic levels of hormones are usually several-fold greater than would normally be secreted by endocrine tissues. Pharmacologic levels occur as a consequence of either pathologic processes or the administration of large doses of hormones. The tissue response to pharmacologic hormone concentrations may be significantly different from that caused by physiologic levels of hormones.

Regulation of Receptor Responses

Tissue response to circulating hormone is only partially determined by the amount of hormone present in the circulation. Although virtually all body tissues are exposed to circulating hormones, specific hormones elicit a response only from certain cells and tissues. The ability of a cell to respond to a particular hormone depends on the presence of specific receptors for that hormone on or in the cell. Target cells are able to regulate their responsiveness to hormones by altering the receptor number, affinity, and efficiency of coupling to intracellular responses.

Receptor Specificity and Affinity

The concept of *receptor specificity* is important for understanding the targeted responses of the endocrine system. Specificity refers to the molecular "fit" of a hormone within a receptor-binding pocket. Specificity cannot easily be separated from the concept of *affinity.* Affinity describes the degree of "tightness" of the hormone–receptor bond, or the inclination of the hormone to remain bound to the receptor. Specificity and affinity determine whether a cell with a receptor will respond to a hormonal stimulus and the hormone concentration at which the receptor effectively can bind the hormone. The potency of the hormone, or the amount of hormone required to elicit a cellular response, is dependent in part on the affinity of the receptor for a particular hormone. The higher the affinity of the receptor for a hormone, the lesser the amount of hormone needed to produce a response. Therefore hormones that circulate in very minute amounts may have a large effect on cellular activity because of the tightness of the hormone–receptor bond.

"Cross-specificity" between hormones of similar structure may occur. For example, PRL and GH both can bind to the PRL receptor, but with differing affinities. Under normal circumstances, the plasma concentration of GH is not adequate to bind a significant number of PRL receptors. However, in the condition called *acromegaly,* in which excessive concentrations of GH lead to massive bone and tissue overgrowth, milk may be secreted from the mammary glands secondary to GH stimulation of PRL receptors.

Receptor Down-Regulation and Up-Regulation

A factor determining the degree of response to a circulating hormone is the number of cell receptors available for hormone binding. When cells are exposed to high concentrations of hormone for a prolonged period, a common result is that the cell decreases the number of receptors. This phenomenon is known as *down-regulation.* An example of down-regulation occurs with insulin receptors in obese individuals. An increase in the plasma insulin level commonly occurs in obese individuals. This

increased concentration of insulin does not result in an increase in cellular activities in most cells because the number of insulin receptors is down-regulated in response to the high plasma insulin concentration. Down-regulation probably serves a protective function: the cells are protected against excessive activity despite pathologic processes that cause excessive hormone levels. *Up-regulation,* or an increase in the number of receptors in response to chronically low hormone concentrations, may also occur. Up-regulation would make the cell more sensitive to the hormone, and hormone-dependent cellular activity could occur at normal or nearly normal levels despite a lower-than-normal hormone concentration.

Another mechanism for altering hormone responsiveness is through regulation of the coupling mechanism between receptor activation and its intracellular responses. For example, the cytoplasmic domain of a receptor can be phosphorylated or otherwise altered so that it cannot effectively interact with its target G-protein. Additionally, enzymes such as phosphatases, which remove phosphates from proteins, and phosphodiesterases, which degrade cyclic nucleotides, can be up-regulated or down-regulated to turn off the cellular responses more or less quickly. Many of these regulatory mechanisms are not fully understood. The term *postreceptor signaling mechanisms* is used to describe these downstream processes.

Permissiveness

One effect that hormones may have on target cells is to increase the number of receptors for other hormones, thus enhancing the effect of the second hormone. This phenomenon is known as *permissiveness.* For example, one effect of thyroid hormone on adipose cells is to increase the number of receptors for epinephrine. When the adipose tissue is subsequently exposed to epinephrine, a greater release of fatty acids (used in providing energy for cellular processes) takes place than would occur in the absence of thyroid hormone. Permissiveness also allows cellular events to occur in sequence. One effect of estrogen, secreted early in the menstrual cycle, is to increase the number of uterine receptors for progesterone. The uterus is therefore sensitive to progesterone when it is present during the last part of the menstrual cycle, and normal proliferative changes take place.

Hormone Agonists and Antagonists

To produce a cellular effect, a hormone must bind to the receptor and initiate a series of events that lead to a change in cellular activity. A chemical may bind to a receptor without initiating the typical intracellular changes; this chemical is described as a hormone antagonist or blocking agent**.** Agonists**,** on the other hand, bind hormone receptors and cause the same intracellular events that would occur with normal hormone-receptor binding. A large number of pharmacologic agents have agonist or antagonist effects on receptors for hormones. For example, hormone antagonists that compete with epinephrine and norepinephrine for receptor sites are frequently used to block the cardiac stimulatory properties of these hormones and thus decrease cardiac workload. Conversely, agonist medications that mimic hormone activity are frequently given as hormone replacement therapy in deficiency syndromes.

KEY POINTS

- Hormone concentration in the bloodstream depends on the rate of secretion and on the rate of degradation by the liver, kidneys, and target tissues. Lipid-soluble hormones are less readily metabolized than peptides and usually have a longer half-life.
- The secretion of most hormones is regulated by negative-feedback mechanisms. Target gland hormones provide feedback to inhibit secretion of the initial hormone and keep the response within a normal range or set point.
- Only target cells producing the specific receptor for a hormone will respond to a particular hormone. High-affinity receptors enable hormone binding at very low physiologic concentrations.
- The number of hormone receptors on a cell's surface may adapt to changes in cell environment. Down-regulation refers to a decrease in the number of receptors in response to excessive amounts of hormone. Up-regulation refers to an increase in receptor number in response to low hormone concentration. Receptors may also be uncoupled from their intracellular signaling cascades.
- In the phenomenon known as permissiveness, receptor number and responsiveness may be enhanced by the presence of other hormones.
- Chemicals other than the usual hormone may be able to bind to the hormone's receptors. Chemicals that bind receptors and block activity are antagonists; those that bind receptors and activate them are agonists.

HYPOTHALAMIC-PITUITARY ENDOCRINE SYSTEM

The pituitary gland is located beneath the hypothalamus in the sella turcica, a bony, saddle-shaped enclosure at the base of the skull (Fig. 39.5). In adults, the pituitary gland is composed of distinct anterior and posterior lobes. The pituitary gland is connected to the hypothalamus by the pituitary stalk, which contains a portal system that transports capillary blood from the hypothalamus to the capillaries of the anterior pituitary gland. Neurons whose cell bodies lie within the hypothalamus send their axons down the pituitary stalk and terminate in the posterior pituitary gland. Synthesis and secretion of the various pituitary hormones are controlled, directly or indirectly, by the hypothalamus. Release of posterior pituitary hormones occurs when action potentials generated in the hypothalamic neurons travel down the axons of the pituitary stalk and trigger exocytosis of hormone from the nerve terminals in the posterior pituitary gland. The hypothalamus regulates endocrine function of the anterior pituitary gland by secreting releasing hormones and inhibiting hormones from hypothalamic neurons that are subsequently carried to anterior pituitary endocrine cells by the pituitary portal system.

The pituitary gland is also called the *hypophysis,* and the anterior and posterior lobes are sometimes referred to as the *adenohypophysis* and the *neurohypophysis,* respectively. The posterior pituitary gland secretes two important peptide hormones: ADH (vasopressin) and oxytocin. The anterior lobe of the pituitary gland has five endocrine cell types that secrete six different peptide hormones: (1) somatotropes, which secrete GH; (2) gonadotropes, which secrete gonadotropins (LH, FSH); (3) thyrotropes, which secrete TSH; (4) corticotropes, which secrete ACTH; and finally (5) lactotropes, which secrete PRL. The hormones produced by the anterior pituitary gland have direct actions on other endocrine tissues and the liver (Table 39.1).

Hormones of the Posterior Pituitary Gland

The posterior pituitary gland consists of nerve axons whose neuronal cell bodies originate in the supraoptic and paraventricular nuclei of the hypothalamus (Fig. 39.6). Posterior pituitary hormones, oxytocin and ADH, are produced in the neurons of the hypothalamus and, packaged in vesicles, travel along the length of the nerve axons to the posterior pituitary gland. Release of ADH or oxytocin occurs as a result of depolarization of the appropriate hypothalamic neurons and conduction of action potentials to the nerve terminals in the posterior pituitary gland. ADH is released in response to altered serum osmolality and hypotension and causes water retention by increasing water reabsorption

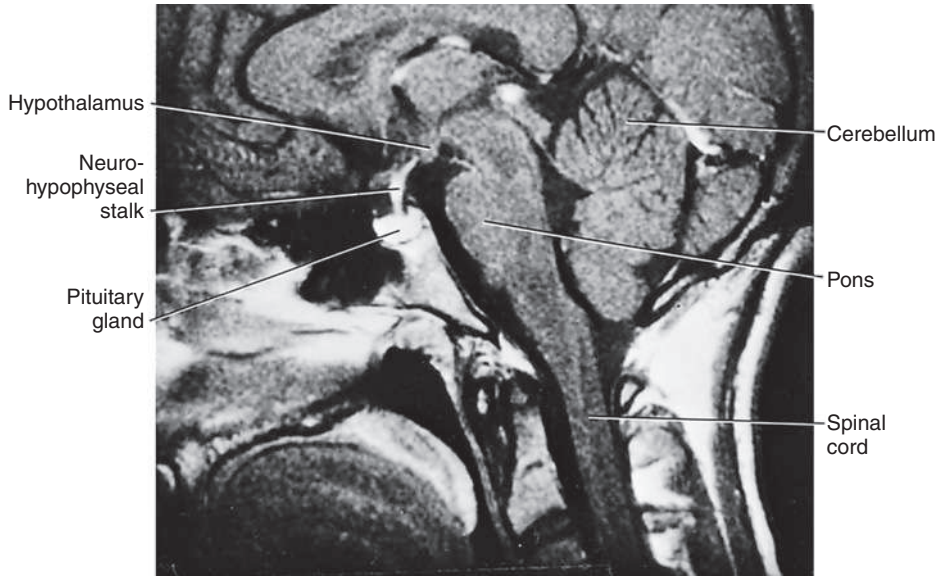


FIG 39.5 Cross-section showing anatomic relationship of the hypothalamus and pituitary gland. (From Koeppen BM, Stanton BA, editors: *Berne & Levy physiology,* ed 6 [updated edition], Philadelphia, 2010, Mosby, p 707.)

TABLE 39.1 Endocrine Cell Types of Adenohypophysis

	Corticotrope	Thyrotrope	Gonadotrope	Somatotrope	Lactotrope
Primary hypothalamic regulation	Corticotropin-releasing hormone (CRH) (41-aa peptide) (stimulatory)	Thyrotropin-releasing hormone (TRH) (3-aa peptide) (stimulatory)	Gonadotropin-releasing hormone (GnRH) (10-aa peptide) (stimulatory)	Growth hormone–releasing hormone (GHRH) (44-aa peptide) (stimulatory); somatostatin (14-aa peptide) (inhibitory)	Dopamine (catecholamine) (inhibitory); prolactin-releasing factor (stimulatory)
Tropic hormone secreted	Adrenocorticotropic hormone (ACTH) (39-aa peptide)	Thyroid-stimulating hormone (TSH) (28-kDa glycoprotein hormone)	Follicle-stimulating hormone and luteinizing hormone (FSH and LH) (28- and 33-kDa glycoprotein hormone)	Growth hormone (GH) (ca. 22-kDa protein)	Prolactin (ca. 23-kDa protein)
Receptor	MC2R (G_s-linked GPCR)	TSH receptor (G_s-linked GPCR)	FSH and LH receptors (G_s-linked GPCRs)	GH receptor (JAK/STAT-linked cytokine receptor)	PRL receptor (JAK/STAT-linked cytokine receptor)
Target endocrine gland	Zona fasciculata and zona reticularis of adrenal cortex	Thyroid epithelium	Ovary (theca and granulose*) and testis (Leydig, Sertoli)	Liver (but also direct actions, especially in terms of metabolic effects)	No endocrine target organ; not part of an endocrine axis
Peripheral hormone involved in negative feedback	Cortisol	Triiodothyronine (T_3)	Estrogen,† progesterone, testosterone, inhibin‡	IGF-1	None

From Koeppen BM, Stanton BA, editors: *Berne & Levy physiology*, ed 6 [updated edition], Philadelphia, 2010, Mosby, p 712.
aa, Amino acid(s); *IGF-1,* insulin-like growth factor-1; *GPCR,* G-protein–coupled receptor; *kDa,* kilodalton(s).
*Both follicular and luteinized thecal and granulosa cells.
†Estrogen can also have positive feedback in women.
‡Inhibin selectivity inhibits FSH release from the gonadotrope.

by the renal collecting duct. Oxytocin is released during sexual activity, childbirth, and breast feeding and causes uterine and milk duct contractions.

Antidiuretic Hormone

ADH is a 9–amino acid peptide that differs by 2 amino acids from oxytocin, another 9–amino acid peptide. The circulating half-life of ADH is about 15 to 20 minutes, and it is destroyed by proteolysis in the kidney and liver. The most important regulator of ADH release is the osmolality of plasma, which is detected by specialized neurons called *osmoreceptors* that are found in the hypothalamus. The osmoreceptors have a set point that influences the stimulation and suppression of ADH. When body fluids become too concentrated, ADH is released, leading to increased reabsorption of water in the kidneys. ADH secretion is also stimulated by the baroreceptors in response to hypovolemia and low arterial blood pressure.

The primary targets for ADH are vasopressin 2 (V2) G-protein–coupled receptors on the basolateral membrane of distal renal tubule

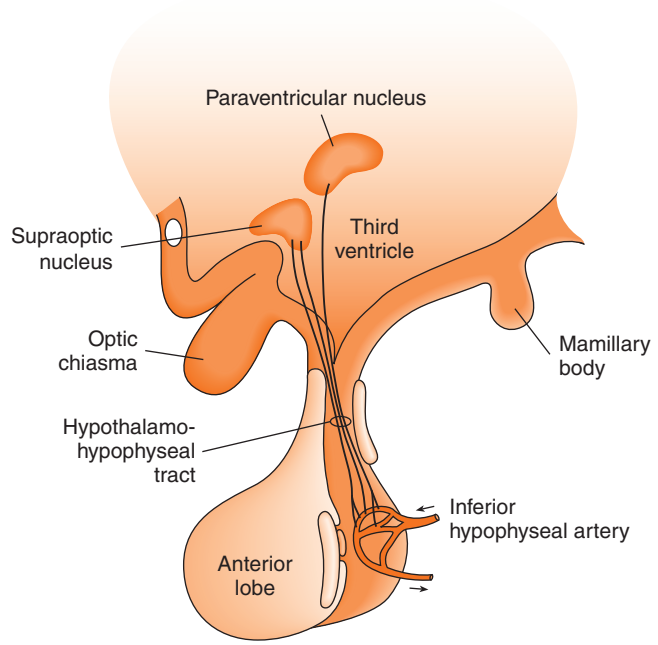


FIG 39.6 Axons from the paraventricular and supraoptic nuclei in the hypothalamus project down the pituitary stalk, and the axon terminals lie within the posterior pituitary gland. Oxytocin and antidiuretic hormone (ADH) are produced in the hypothalamic neuron cell bodies and then transported to the axon terminals, where they are released into the bloodstream in response to action potentials. (From White BA, Porterfield SP: *Endocrine and reproductive physiology,* ed 4, Philadelphia, 2013, Mosby, p 102.)

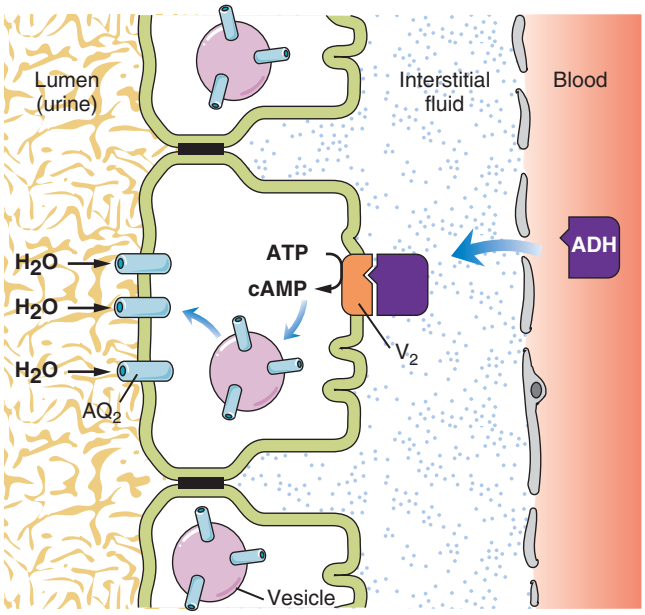


FIG 39.7 Antidiuretic hormone (ADH) acts on renal tubule G_s-protein–coupled receptors (V2) to trigger movement of vesicles containing water channels (aquaporins, AQ2) to the apical cell membrane. The aquaporins increase membrane permeability to water, allowing it to be absorbed through the process of osmosis. ADH thus decreases the osmolality of the extracellular fluids and creates a more concentrated urine. *ATP,* Adenosine triphosphate; *cAMP,* cyclic adenosine monophosphate.

cells. ADH causes pores, called *aquaporins,* to move from the cytoplasm to the cell membranes of apical tubular epithelial cells (Fig. 39.7). These pores allow free diffusion of water from the tubular fluid into the cell. Water then flows out the basolateral membrane and into the interstitium. The enhanced reabsorption conserves water in the body, creates concentrated urine, and reduces serum osmolality (see Chapter 26).

Oxytocin

Oxytocin is structurally similar to ADH, and there is some overlap in biological activity. Oxytocin can inhibit diuresis. Oxytocin has a half-life of 3 to 5 minutes. Oxytocin is known for its actions on the breast and uterus. Sexual arousal, intercourse, and orgasm can stimulate oxytocin release in men and women. Its role in men is not well understood, but in women it is proposed to affect sperm motility through uterine contractions. Oxytocin also stimulates uterine contractions during labor. It is not thought to initiate labor, but once labor has begun, stretching of the cervix increases oxytocin release, which increases the intensity of uterine contractions. This is an example of a positive feedback loop in which stretching of the cervix induces oxytocin release, resulting in more forceful uterine contractions and further stretching of the cervix. Oxytocin has been characterized as the "tend and befriend" hormone because of its effects on social interaction and its participation in the stress response in men and women.

Oxytocin release is triggered by stimulation of the nipple and areola, which have sensory receptors that send neuronal signals to the hypothalamic nuclei. Oxytocin then binds to myoepithelial cells surrounding the milk ducts, causing them to contract and eject milk during breast feeding. This reflex can be triggered centrally without breast stimulation as a conditioned response to the sight or sound of an infant.

Hormones of the Hypothalamus and Anterior Pituitary Gland

The hypothalamic-pituitary endocrine system is a three-tiered axis that includes hypothalamic-releasing and -inhibiting hormones, anterior pituitary hormones, and target organ hormones (Fig. 39.8). Releasing and inhibiting factors are secreted into a capillary bed that is contained within the hypothalamus by hypothalamic neurons with short axons. Hypothalamic hormones diffuse into the capillary network, travel down the portal vein, and then diffuse from the anterior pituitary capillary network into the tissue where they bind receptors on pituitary cells. All of the hypothalamic-releasing and -inhibiting factors are peptides, with the exception of dopamine, and all are water soluble. Therefore all releasing and inhibiting factors interact with cell-surface receptors on the target pituitary cells.

The three-tiered hypothalamic-pituitary target system provides an opportunity for fine modulation of hormone action on target tissues. The hypothalamic and pituitary hormones exemplify the complexity of endocrine interactions. In addition to feedback from target organ hormones, the hypothalamus processes numerous neuronal signals into an integrated response by the anterior pituitary gland. The hypothalamic-releasing and -inhibiting hormones are typically secreted in a pulsatile and circadian rhythm. Two structures thought to modulate this activity are the suprachiasmatic nucleus (SCN) and the pineal gland. Changes in light and dark patterns influence the SCN and pineal gland, resulting in a diurnal (daily) pattern of hormone secretion (Fig. 39.9). Each of the major anterior pituitary hormones is briefly described next.

Growth Hormone

Pituitary GH secretion is controlled by hypothalamic release of growth hormone–releasing hormone (GHRH) and growth hormone–inhibiting hormone (somatostatin). GHRH binds to receptors on pituitary

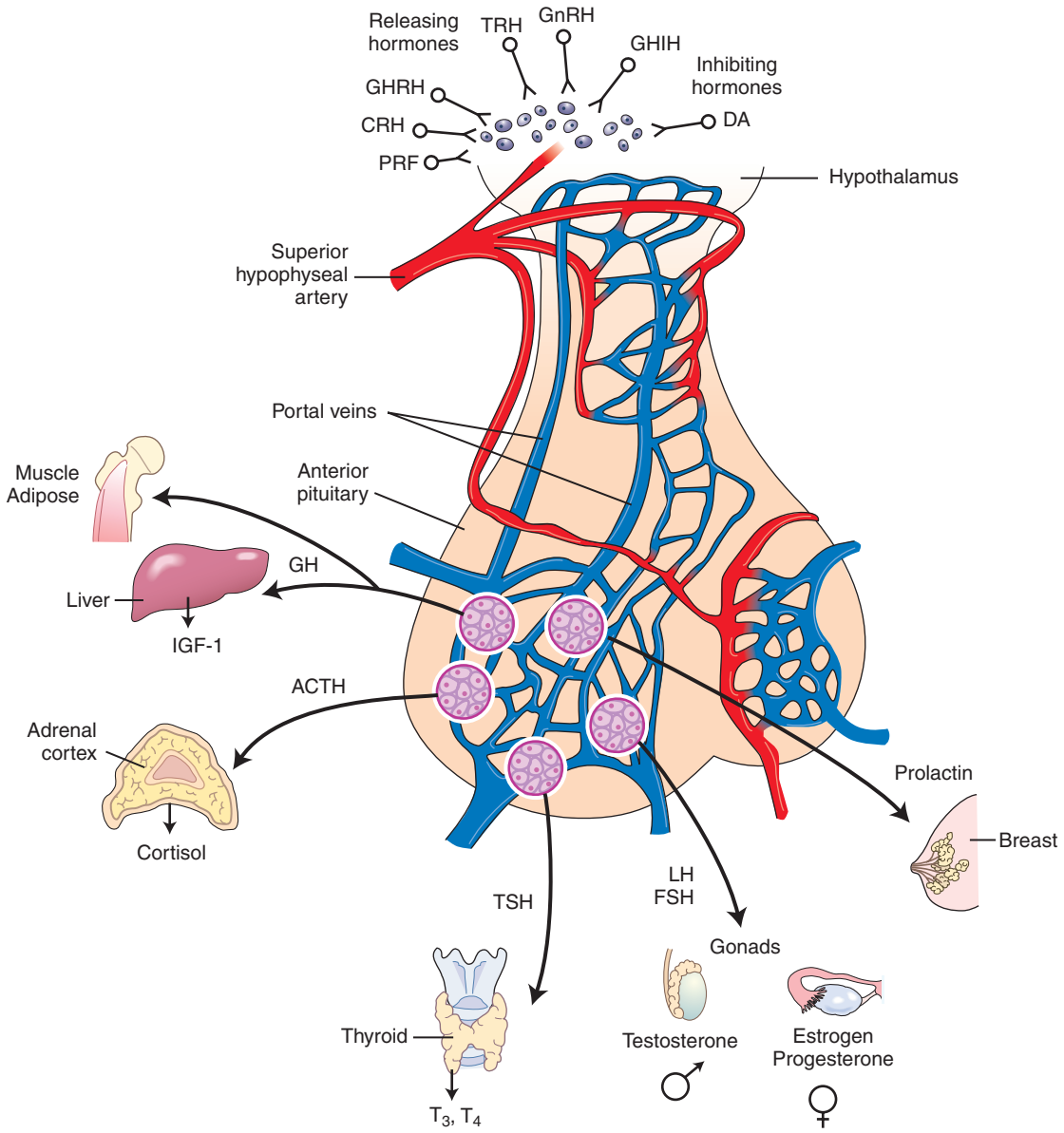


FIG 39.8 Relationship of the hypothalamic-releasing and -inhibiting hormones, anterior pituitary hormones, and target gland hormones. *ACTH,* Adrenocorticotropic hormone; *CRH,* corticotropin-releasing hormone; *DA,* dopamine, *FSH,* follicle-stimulating hormone; *GH,* growth hormone; *GHIH,* growth hormone inhibiting hormone (somatostatin); *GHRH,* growth hormone releasing hormone; *GnRH,* gonadotropin releasing hormone; *IGF-1,* insulin-like growth factor-1; *LH,* luteinizing hormone; *TRH,* thyrotropin-releasing hormone, *TSH,* thyroid-stimulating hormone; T_3, triiodothyronine; T_4, thyroxine.

somatotropes and stimulates release of GH, whereas somatostatin inhibits GH release. Under normal physiologic conditions, the anterior pituitary gland secretes small pulsatile amounts of GH each day. Sleep studies indicate a circadian pattern to GH secretion, with secretion being greatest during deep, slow-wave sleep (stage 3 or 4). GH secretion is greatest during adolescence and decreases in the elderly.

GH is a 191–amino acid protein that is similar in structure to prolactin. A major target for GH is the liver, where it affects liver metabolism and induces the production of another endocrine hormone called *insulin-like growth factor-1* (IGF-1). Previously, IGF-1 was called *somatomedin.* IGF-1 is an anabolic hormone that increases the growth of bone and cartilage tissues of the body. GH is not the only stimulus for the production of IGF-1, nor is the liver the only source. Insulin is an important costimulator of IGF-1 along with GH. In the liver, GH stimulates the breakdown of glycogen reserves, leading to the release of glucose into the bloodstream. GH also has direct effects on some cells and binds to receptors on muscle and adipose tissue, increasing muscle mass and decreasing fat mass by inducing lipolysis. It affects metabolic processes by increasing the rate of protein synthesis, decreasing protein catabolism, slowing carbohydrate utilization, and increasing mobilization of fats and the use of fats for energy (Fig. 39.10). An insufficient level of GH is exhibited as short stature in children, whereas excesses may lead to increased height in children and excessive growth of nonadipose tissues in adults (see Chapter 40).

Feedback regulation of GH release is complex and is affected by several hormones, including estrogen, testosterone, and thyroid hormone, which stimulate GH release. Perhaps the most important negative feedback is through IGF-1. Hypoglycemia and a rise in the concentration

of amino acids in the blood stimulate the release of GH, as does starvation and exercise.

Prolactin

The lactotropes that secrete PRL are the second most numerous cell type in the anterior pituitary gland after the somatotropes. PRL is a 199–amino acid protein that acts directly on numerous cell types. It is best known for its trophic effects on breast tissue development and lactation and its ability to suppress reproductive function in both men and women. PRL exerts its inhibitory effects on fertility through suppression of hypothalamic gonadotropin-releasing hormone. Numerous other functions for PRL have been proposed, and it has been noted that its level increases during sexual orgasm, stress, and exercise.

PRL secretion is normally under tonic (constant) inhibition by PRL-inhibiting factor. The PRL-inhibiting factor is dopamine produced in the hypothalamus. Drugs that block dopamine receptors can release PRL from its normal inhibition, resulting in abnormal breast development in males (gynecomastia) or inappropriate milk secretion (galactorrhea). During pregnancy, the number of pituitary lactotropes increases under the influence of placental estrogen. Estrogen production and nipple stimulation increase PRL synthesis and release from the anterior pituitary gland.

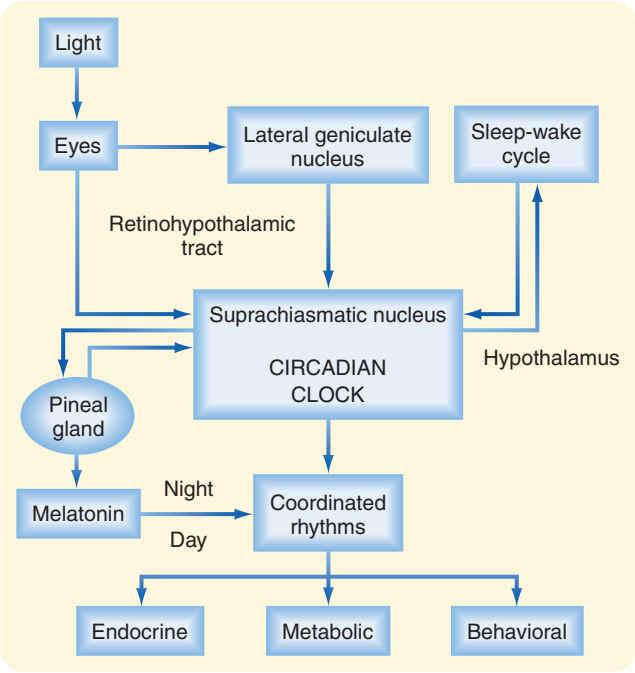


FIG 39.9 Origin of circadian rhythms in endocrine gland secretion. (From Koeppen BM, Stanton BA, editors: *Berne & Levy physiology,* ed 6 [updated edition], Philadelphia, 2010, Mosby, p 656.)

Gonadotropins

The gonadotropins FSH and LH are produced together within the gonadotropes in the anterior pituitary gland; however, they are segregated into different packaging vesicles and are not necessarily released in equal amounts. FSH and LH are proteins and bind to their specific respective receptors on target cells in the ovary or testes. Gonadotropins stimulate testosterone production in men (LH) and estrogen and progesterone production in women. In men, FSH promotes the physical maturation of developing sperm. These relationships are complex in women and change over the menstrual cycle (see Chapter 32). Release of FSH and LH is stimulated by hypothalamic gonadotropin-releasing hormone in a pulsatile fashion and inhibited by negative feedback from sex steroids (except for the midcycle positive feedback LH surge induced by estrogen at the time of ovulation in women).

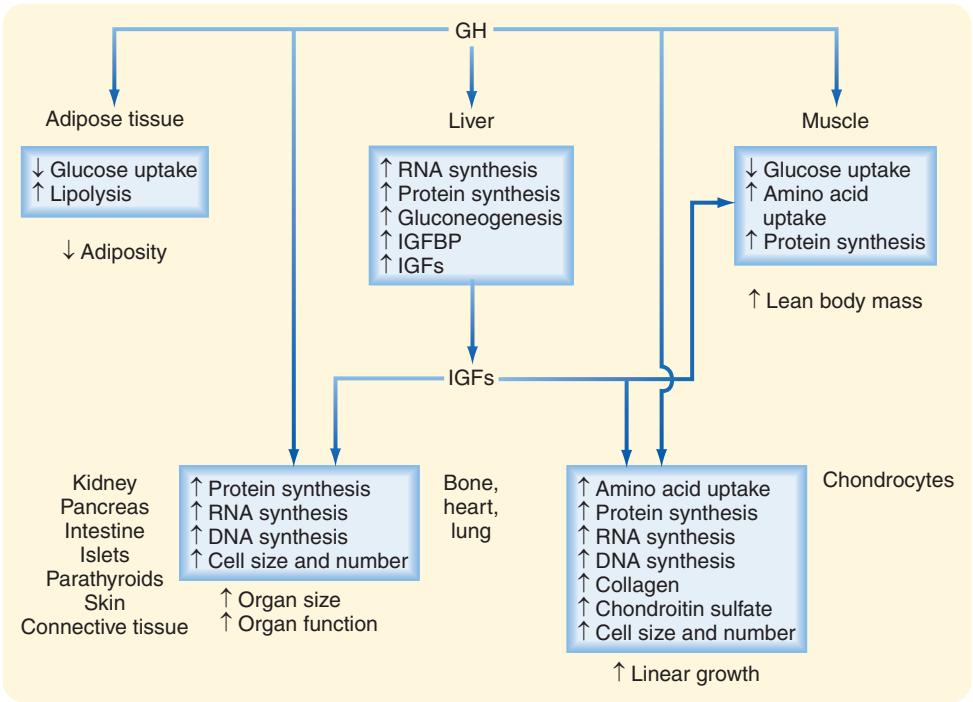


FIG 39.10 Biological actions of growth hormone (GH). *IGF,* Insulin-like growth factor; *IGFBP,* insulin-like growth factor–binding protein. (From Koeppen BM, Stanton BA, editors: *Berne & Levy physiology,* ed 6 [updated edition], Philadelphia, 2010, Mosby, p 721.)

Thyroid-Stimulating Hormone

Thyrotropes release TSH in response to hypothalamic TRH. TSH is a glycoprotein that binds to TSH receptors on follicle cells of the thyroid gland. TRH is released according to a diurnal (daily circadian) rhythm with its lowest period in the evening. Stress, starvation, and infection reduce the secretion of TRH. Negative feedback regulation of TRH and TSH is achieved primarily by the concentration of circulating T_3. TSH regulates all aspects of thyroid function, including growth of the gland and synthesis and secretion of hormone. The details of thyroid hormone synthesis and action on target tissues are described later in this chapter.

Adrenocorticotropic Hormone

ACTH is produced by corticotropes in the anterior pituitary gland in response to hypothalamic corticotropin-releasing hormone (CRH). ACTH is a 39–amino acid peptide that is derived from cleavage of a large preprohormone called *proopiomelanocortin (POMC)*. Although POMC has the protein sequences for melanocyte-stimulating hormone and endorphin in its structure, human corticotropes express only the enzyme for cleaving ACTH as the sole hormone from this preprohormone. ACTH circulates unbound in the circulation and has a half-life of about 10 minutes. It binds to G_s-coupled receptors on cells in the adrenal cortex and stimulates the production of cortisol and adrenal androgens. ACTH also has trophic effects on the adrenal cortex and supports the structure and synthetic enzymes of the gland. Withdrawal of ACTH results in adrenal cortical atrophy. ACTH in high concentration stimulates darkening of the skin by stimulating melanocyte receptors on skin cells. Normally keratinocytes in the skin produce melanocyte-stimulating hormone (MSH) in response to ultraviolet light, which can bind to MSH receptors on neighboring skin melanocytes in a paracrine fashion. ACTH can cross-react with melanocyte MSH receptors, and in high concentrations ACTH can produce abnormal skin pigmentation.

CRH and ACTH secretion demonstrate a significant diurnal pattern, with a peak upon wakening and a valley during the usual time of sleep (Fig. 39.11). The diurnal pattern can be altered when there are chronic changes in sleep–wake patterns. Negative feedback regulation of ACTH is accomplished by the actions of cortisol, which suppress CRH and ACTH release. Regulation of the hypothalamic-pituitary-adrenal axis is complex and is affected by acute and chronic stress (see Chapter 2). The adrenal cortex produces three types of steroid hormones: glucocorticoids, mineralocorticoids, and androgens. The details of adrenocortical steroid hormone synthesis and action on target tissues are described later in this chapter.

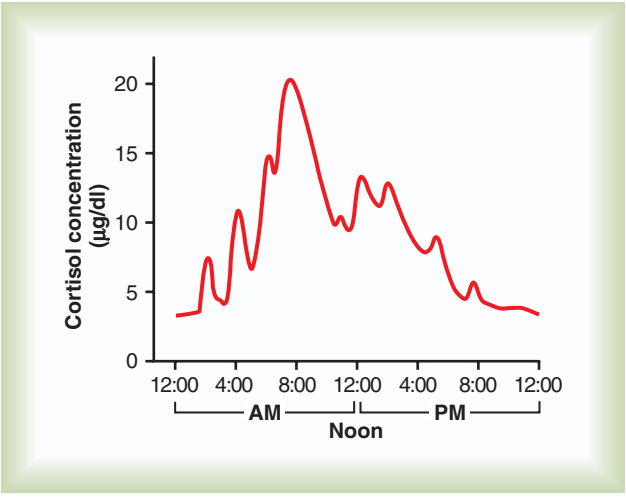


FIG 39.11 Typical pattern of cortisol secretion showing a diurnal (circadian) rhythm, with highest secretion occurring in the morning upon awakening and lowest levels in the late evening. The pattern can change when the usual wake and sleep times are altered, for example, in night-shift workers. (In Hall JE, editor: *Guyton and Hall textbook of medical physiology,* ed 13, Philadelphia, 2015, Saunders, p 977.)

KEY POINTS

- Oxytocin and ADH (vasopressin) are synthesized in hypothalamic neurons that send axons to the posterior pituitary gland. Oxytocin is released during childbirth and suckling. ADH is released in response to increased serum osmolality and decreased blood volume.
- Major hormones of the anterior pituitary gland are GH, TSH, PRL, ACTH, and gonadotropins (FSH, LH). The release of anterior pituitary hormones is regulated by releasing and inhibiting hormones secreted into pituitary portal blood by the hypothalamus. Many factors influence the secretion of releasing and inhibiting hormones, including circadian rhythms, hormone release from target cells, stress, and pain.

THYROID HORMONES

Thyroid Hormone Synthesis and Secretion

The thyroid gland, a two-lobed gland, lies in the neck region on either side of and anterior to the trachea. It secretes the thyroid hormones thyroxine (T_4) and triiodothyronine (T_3). The thyroid gland is composed of microscopic follicles made up of a single layer of epithelial cells forming a ball, with their apical surfaces toward a pocket in the middle filled with colloid and their basal surfaces facing the interstitial fluid and capillary system (Fig. 39.12). Follicular cells perform all the functions required to make and secrete thyroid hormones. They trap dietary iodine and transport it into the colloid, synthesize thyroglobulin protein, and transport it into the colloid along with the enzyme thyroid peroxidase. Thyroid peroxidase acts on thyroglobulin to produce thyroid hormones (Fig. 39.13). First it oxidizes iodide (I^-) and couples it to tyrosine amino acids in thyroglobulin. Some tyrosines receive two iodides (diiodotyrosine, DIT) and others have only one (monoiodotyrosine, MIT). Two iodotyrosines are then attached to each other by thyroid peroxidase to form T_4 (two DIT) or T_3 (one DIT and one MIT). Approximately 90% of the thyroid hormone in the colloid is in the form of T_4, and 10% is in the form of T_3. The thyroid hormones remain attached to thyroglobulin, a storage protein, which accumulates in the thyroid follicles. By remaining attached to thyroglobulin, T_3 and T_4 are trapped within the water-soluble colloid. When stimulated by TSH, the follicular cells endocytose a portion of the colloid containing thyroglobulin with its attached thyroid hormones. The endocytic vesicle combines with a lysosome, and the thyroid hormones are cleaved from thyroglobulin. Released T_4 and T_3 are transported from the follicle and into the circulation through carriers on the basal membrane.

Thyroid Action on Target Cells

Thyroid hormones are carried in the circulation bound to thyroid-binding proteins. Ninety percent of the circulating thyroid hormone is T_4, and 10% is T_3 (Fig. 39.14). A small percentage of thyroid hormone (less than 1%) is dissolved in the plasma as free hormone and is able to cross into the target cells through specific membrane carrier proteins. Once in the cell, T_3 is able to bind to its receptor and exert its actions. T_4 is acted upon by cellular enzymes that cleave one of the iodine molecules to form either the active T_3 or a mirror image of T_3 called *reverse T_3*. This stereoisomer of T_3 has no known biological activity. The ratio of production of T_3 and reverse T_3 is normally about 1:1, but is influenced by altered metabolic states such as starvation. Most of the actions of

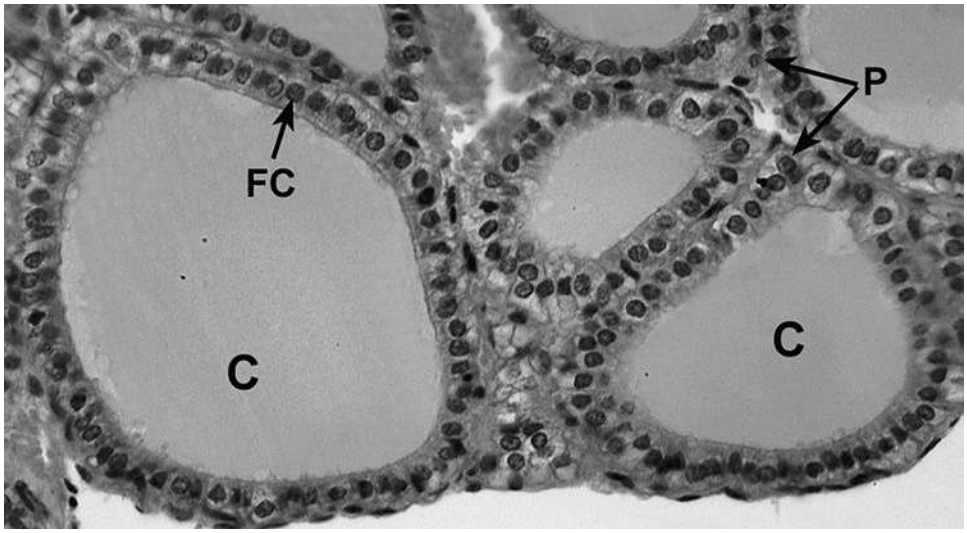


FIG 39.12 Histology of thyroid follicle demonstrating single layer of epithelial follicular cells *(FC)* surrounding a core of colloid *(C)*. The parafollicular cells *(P)* produce another hormone called calcitonin, which is involved in calcium regulation. (From White BA, Porterfield SP: *Endocrine and reproductive physiology,* ed 4, Philadelphia, 2013, Mosby, p 131.)

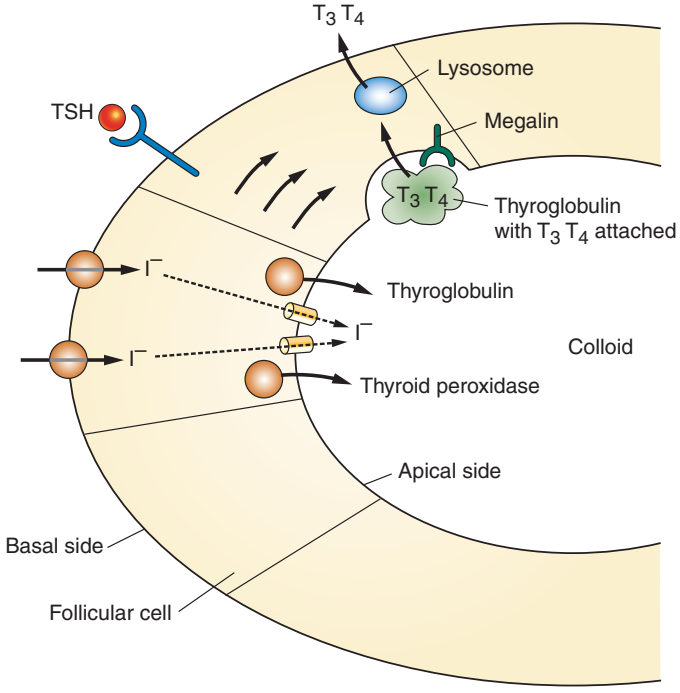


FIG 39.13 The follicular cells in the thyroid gland perform all of the functions required for thyroid hormone synthesis. Iodide (I^-) is transported from the circulation into the inner core along with the protein thyroglobulin, which is synthesized within the cells. Follicles also produce a thyroid peroxidase enzyme and secrete it on the apical surface next to the colloid. Thyroid peroxidase activates and attaches iodides to tyrosine amino acids in the thyroglobulin. Two iodotyrosines are then coupled together, but remain attached to thyroglobulin. Stimulation by TSH initiates a cascade that induces the follicle cell to endocytose some of the colloid, followed by combining it with a lysosome that cleaves the T_3 and T_4 from the thyroglobulin. T_3 and T_4 can then leave the cell into the circulation.

thyroid hormone on target cells are mediated through alterations in gene transcription. The nuclear thyroid receptor has much higher affinity for T_3 than T_4 and primarily binds T_3 to form a complex that binds to specific DNA sequences and alters gene activity. Many of these genes influence the metabolic rate of the cell. The general physiologic effects of thyroid hormones are shown in Table 39.2. Excesses and deficits of thyroid hormone are exhibited by alterations in growth, development, and metabolic rate (see Chapter 40).

KEY POINTS

- Follicular cells in the thyroid gland perform all the functions required to make and secrete thyroid hormones. They trap dietary iodine and transport it into the colloid, synthesize thyroglobulin protein, and transport it into the colloid along with the enzyme thyroid peroxidase.
- Thyroid peroxidase oxidizes iodide and couples it to tyrosine amino acids in thyroglobulin. Two iodotyrosines are then attached to each other to form T_4 (two DIT) or T_3 (one DIT and one MIT). Approximately 90% of the thyroid hormone is in the form of T_4, and 10% is in the form of T_3.
- When stimulated by TSH, the follicular cells endocytose a portion of the colloid. The endocytic vesicle combines with a lysosome, and the lipid-soluble T_4 and T_3 are released and leave the follicle where they are more than 99% bound to proteins in the circulation. Only free hormone is physiologically active.
- Actions of thyroid hormone on target cells are mediated through alterations in gene transcription. The nuclear thyroid receptor binds T_3 to form a complex that binds to specific DNA sequences and alters gene activity.
- Thyroid hormone increases metabolic rate and is essential for normal growth and development.

STEROID HORMONES

Steroid Hormone Synthesis and Secretion

The adrenal glands are located atop the kidneys and are composed of an inner medulla and an outer cortex. The adrenal medulla secretes epinephrine and norepinephrine in response to sympathetic nervous system stimulation and is discussed in Chapter 40. The hormones produced by the adrenal cortex are called steroids and include (1)

TABLE 39.2 Physiologic Actions of Thyroid Hormones

	Hypothyroid	Euthyroid	Hyperthyroid
Metabolic rate	Decreased BMR	—	Increased BMR
Proteins	↓ Synthesis, ↓ degradation, ↓ turnover (% BW as protein will ↓)	Protein anabolic	↑ Synthesis, ↑ degradation, ↑ turnover (catabolic if insufficient dietary protein)
Lipids	↓ Synthesis, ↓ degradation, ↓ turnover (% BW as lipid increases), ↑ serum cholesterol	↑ Beta oxidation, ↑ lipolysis, ↑ lipogenesis	↑ Synthesis, ↑ degradation, ↑ turnover (% of BW as lipid decreases), ↓ serum cholesterol
Glucose	Normal	Normal	Normal serum glucose; abnormal glucose tolerance test
Glycogen	↓ Synthesis, ↓ degradation, ↓ turnover, glycogen accumulates	—	↑ Synthesis, ↑ degradation, ↑ turnover; glycogen is depleted
Actions with SNS	—	—	Excess mimics effects of ↑ β-adrenergic stimulation; can ↑ number and affinity of β-receptors and ↑ adenylyl cyclase sensitivity
Direct cardiovascular actions	↓ Amplitude of ECG waves	↑ HR, ↑ CO, ↑ contractility, ↑ pulse pressure, ↑ actin and myosin	↑ Amplitude of ECG waves

BMR, Basal metabolic rate; *BW,* body weight; *CO,* cardiac output; *ECG,* electrocardiogram; *HR,* heart rate; *SNS,* sympathetic nervous system.

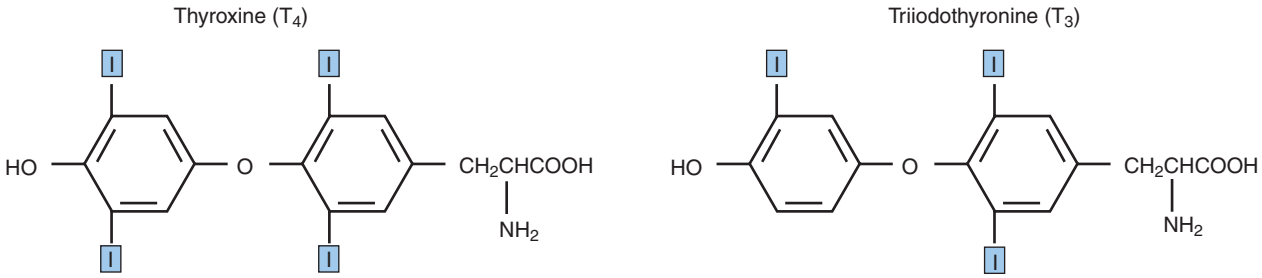


FIG 39.14 Structures of thyroxine (T_4) and triiodothyronine (T_3).

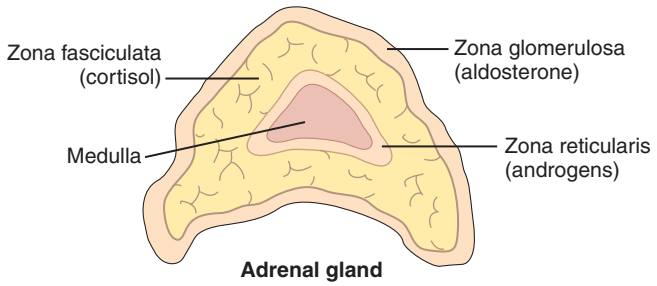


FIG 39.15 The adrenal gland is composed of the adrenal medulla at the innermost core, surrounded by the cortex. The adrenal cortex is composed of three anatomically and physiologically distinct layers: the zona glomerulosa, the zona fasciculata, and the zona reticularis. Each zone produces a different adrenocortical hormone.

glucocorticoids (cortisol), (2) mineralocorticoids (aldosterone), and (3) sex steroids (dehydroepiandrosterone [DHEAS], androgen). The synthesis and secretion of these hormones, especially cortisol, are considered essential for life, regulating the body's response to normal and abnormal levels of physiologic and psychological stress. The activities of these three hormones can be remembered as regulating the "three S's": sugar, salt, and sex.

The adrenal cortex has distinct zones that differ in histologic appearance, hormonal regulation, and enzyme pathways available for steroid synthesis (Fig. 39.15). The outer *zona glomerulosa* produces the mineralocorticoid (aldosterone) in response to stimulation by AII. The middle *zona fasciculata* produces the glucocorticoid (cortisol) in response to stimulation by ACTH from the pituitary gland. The inner *zona reticularis* is adjacent to the adrenal medulla and produces the androgen (DHEAS). All of these hormones are synthesized from cholesterol, and some of the cholesterol-derived precursors feed enzymatic pathways in all three zones (Fig. 39.16). This is an important concept because it explains why an enzyme deficiency in one pathway (e.g., cortisol synthesis) may lead to an overproduction in another pathway (e.g., DHEAS) as precursors are shunted. This is what happens in a disorder called *congenital adrenal hyperplasia* (see Chapter 40).

As previously noted, steroid hormones are lipid soluble and diffuse from the adrenocortical cells as they are synthesized, so they are made on demand and not stored in the cells. The stimulus for both cortisol and androgen synthesis and secretion is ACTH, whereas aldosterone synthesis and secretion are under control of AII (Fig. 39.17). The activity of aldosterone synthase, the enzyme that performs the last step to produce aldosterone, is stimulated by AII binding to G_q receptors on adrenocortical cells in the zona glomerulosa.

Steroid Action on Target Cells

Steroid hormones travel in the circulation bound to proteins, including corticosteroid-binding globulin (transcortin) and albumin. Cortisol and androgens bind tightly with their binding proteins, but aldosterone does not and has a shorter half-life in the circulation. Steroids have long been known to diffuse through their target cell membranes and bind with their respective cytoplasmic receptors, and then the hormone–receptor

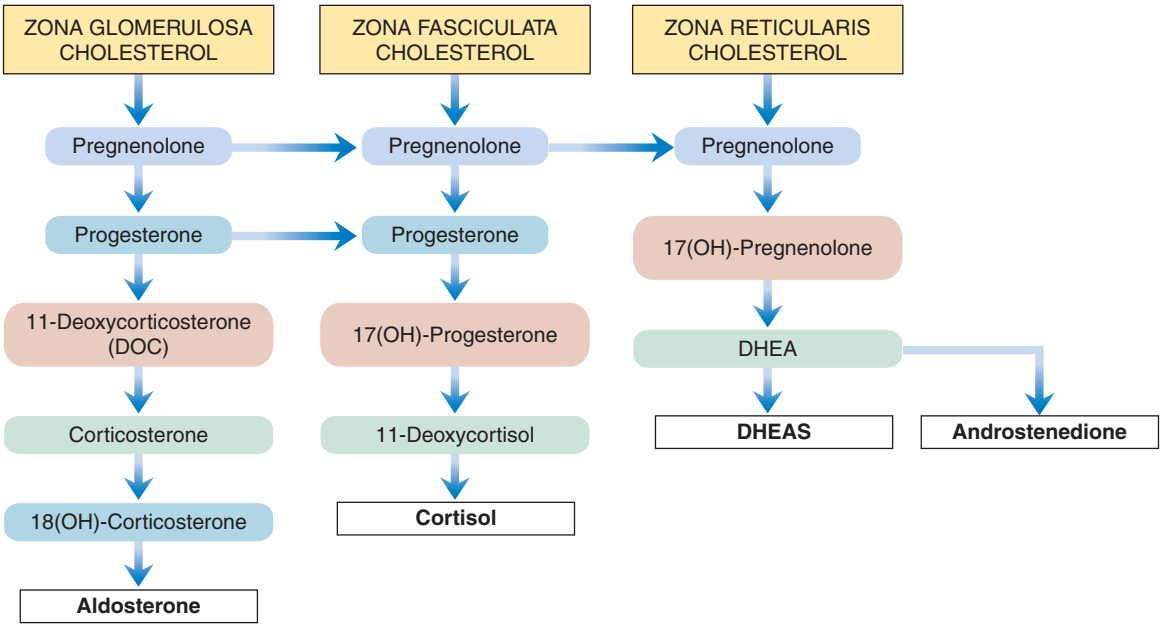


FIG 39.16 Adrenal cortex steroid hormone synthesis. Steroid hormones are synthesized from cholesterol and have some common precursors. *DHEAS,* Dehydroepiandrosterone.

BOX 39.2 Biological Actions of Cortisol

Hyperglycemic
Gluconeogenic
Lipolytic
Protein catabolic
Insulin antagonist in muscle and adipose tissue
Inhibits bone formation, stimulates bone resorption
Necessary for vascular response to catecholamines
Antiinflammatory
Suppresses immune system
Inhibits antidiuretic hormone secretion and action
Stimulates gastric acid secretion
Necessary for integrity and function of gastrointestinal tract
Stimulates red blood cell production
Alters mood and behavior
Permissive for calorigenic, lipolytic effects of catecholamines

Adapted from White BA, Porterfield SP: *Endocrine and reproductive physiology,* ed 4, Philadelphia, 2013, Mosby, p 159.

complex rapidly translocates into the nucleus. The complex binds to specific DNA sequences to alter responsive genes. In addition to this classic pathway, steroid hormones may bind to receptors already located in the nucleus or to cell membrane–associated receptors. Membrane-associated receptors mediate more rapid changes in cell function than the classic DNA-binding pathway.

Glucocorticoids, principally cortisol, are named for their primary effect on glucose metabolism. Cortisol affects nearly every tissue in the body and exhibits a broad range of manifestations (Box 39.2). Glucocorticoids oppose the effects of insulin and raise the blood glucose level. This is accomplished by decreasing glucose uptake by many body cells and increasing glucose synthesis in the liver from glycogen and amino acid and glycerol substrates in protein and fat stores. Glucocorticoids also contribute to protein catabolism by releasing muscle stores of proteins, providing amino acids for glucose production in the liver. Glucocorticoids promote lipolysis and increased blood cholesterol level. Glucocorticoids protect against the damaging physiologic effects of stress (see Chapter 2) and regulate the inflammatory and immune responses.

Mineralocorticoids, principally aldosterone, function to maintain normal salt and water balance by promoting sodium retention and potassium excretion at the distal renal tubules. Aldosterone production is regulated primarily by the renin–angiotensin system associated with the juxtaglomerular cells of the kidney in response to a reduction in renal perfusion and by a high serum potassium level (see Chapter 26).

KEY POINTS

- The hormones produced by the adrenal cortex are called *steroids* and include (1) glucocorticoids (cortisol), (2) mineralocorticoids (aldosterone), and (3) sex steroids (androgens).
- The adrenal cortex has distinct zones. The outer *zona glomerulosa* produces aldosterone in response to stimulation by AII. The middle *zona fasciculata* produces cortisol in response to stimulation of ACTH from the pituitary gland. The inner *zona reticularis* is adjacent to the adrenal medulla and produces androgens.
- Steroids diffuse through their target cell membranes and bind with their respective cytoplasmic (or nuclear) receptors. The hormone–receptor complex rapidly translocates into the nucleus where it binds to specific DNA sequences to alter responsive genes. More rapid actions of steroid hormones have been noted and are mediated by cell membrane–associated receptors.
- Cortisol affects nearly every tissue in the body and exhibits a broad range of effects on metabolism, including increasing plasma glucose level, regulating immune and inflammatory reactions, and inhibiting bone and collagen synthesis. Cortisol is an important hormone in the response to acute and chronic stress.
- Aldosterone secretion increases in response to low blood pressure and reduced perfusion of the kidney, which stimulate the renin–angiotensin–aldosterone cascade. Aldosterone increases sodium and water reabsorption in the distal tubule of the kidney and promotes the excretion of potassium in the urine.

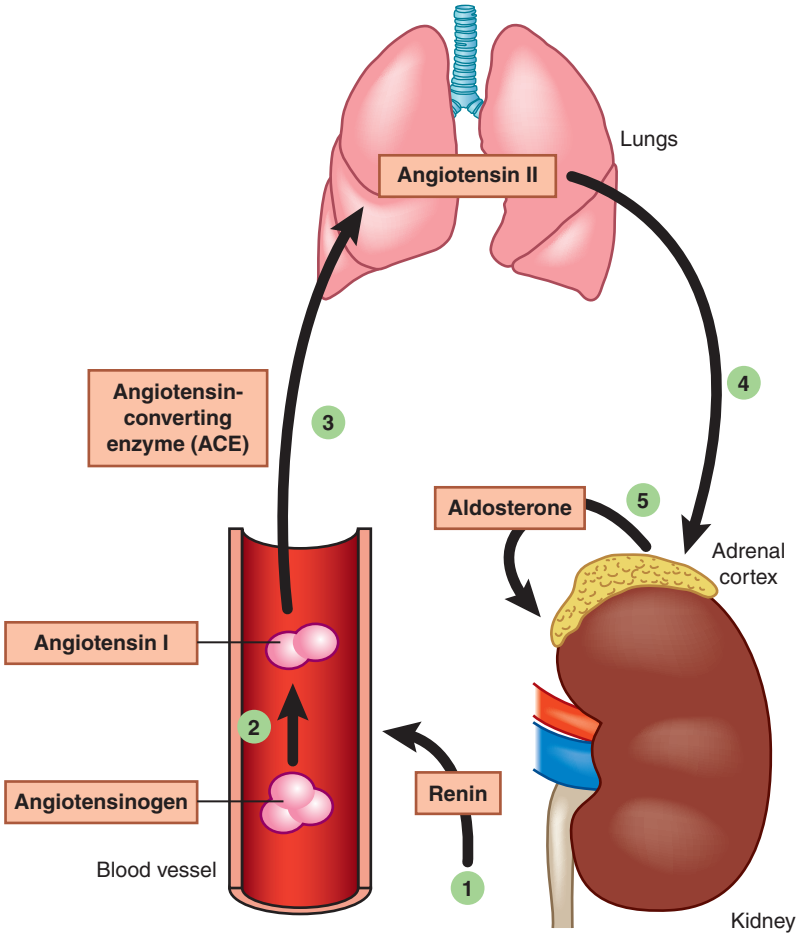


FIG 39.17 Renin–angiotensin mechanism for regulating aldosterone secretion. (1) When the incoming blood pressure in the kidneys drops below a certain level, the juxtaglomerular apparatus secretes renin into the blood. (2) Renin, an enzyme, causes angiotensinogen (a normal constituent of blood) to be converted into angiotensin I. (3) Angiotensin I circulates to the lungs, where converting enzymes in the capillaries split the molecule, forming angiotensin II. (4) Angiotensin II circulates to the adrenal cortex, where it stimulates the secretion of aldosterone. (5) Aldosterone causes increased reabsorption of sodium, which causes increased water retention. As water is retained, the volume of blood increases. The increased volume of blood creates higher blood pressure, which then causes the renin secretion to stop. (Modified from Patton KT, Thibodeau GA: *Anatomy & physiology,* ed 7, St Louis, 2010, Mosby, p 561.)

CATEGORIES OF ENDOCRINE DISEASE

Endocrine pathologic processes can be divided into three general categories: disorders of hyposecretion, hypersecretion, and target cell hyporesponsiveness. These are discussed in relation to specific endocrine diseases in Chapter 40.

Hyposecretion

Primary hyposecretion occurs when an endocrine gland, such as the thyroid or adrenal cortex, releases an inadequate amount of hormone to meet physiologic needs. Secondary hyposecretion occurs when secretion of a tropic hormone, such as TSH or ACTH, is inadequate to cause the target gland to secrete adequate amounts of hormone. The diagnosis of hormone deficiency is complex because knowledge of tropic and releasing hormone levels, as well as the deficient hormone level, is necessary. For example, in a *primary* thyroid hormone deficiency (common), serum thyroid hormone level would be low but TSH levels would be high because the anterior pituitary gland would not be receiving negative feedback from thyroid hormone. However, in *secondary* thyroid hormone deficiency (rare), both thyroid hormone and TSH concentrations would be abnormally low.

Hypersecretion

Hypersecretion disorders can also be either primary or secondary. When a diseased endocrine gland secretes an abnormally high amount of its hormone, the tropic pituitary hormone will be at an unusually low plasma level because of excessive negative feedback. Alternatively, if hypersecretion is secondary to elevated tropic hormone levels (rare), the plasma concentration of both hormones will be elevated. For example, in Cushing disease, the pituitary gland becomes hyperactive and oversecretes ACTH, which induces the adrenal cortex to produce too much cortisol (causing symptoms of cortisol excess). Excessive plasma hormone levels can also occur from hormone secretion by an ectopic source, as sometimes occurs with malignancies.

Hyporesponsiveness

Hyporesponsiveness (hormone resistance) of the target tissues will cause the same set of clinical symptoms as hyposecretion. The usual reason

for hyporesponsiveness is lack of or a deficiency in cellular receptors, although postreceptor mechanisms such as second-messenger dysfunction can also cause decreased cellular response. If the target cell does not have appropriate receptors for a hormone, the clinical symptoms will be the same as if inadequate hormone levels were reaching the target cells. However, plasma concentrations of hormone would be expected to be normal or high because of the lack of negative feedback to hormone-secreting organs. Nephrogenic diabetes insipidus is an example of kidney tubule resistance to the effect of ADH because of defective receptors on the tubular cells. Some forms of diabetes mellitus also are characterized by tissue resistance to the effects of insulin (see Chapter 41).

KEY POINTS

- Endocrine disorders occur because of hyposecretion, hypersecretion, or lack of responsiveness by target cells. Hyporesponsiveness is clinically similar to hyposecretion and usually results from a lack of functional receptors or a defect in postreceptor signaling.
- Endocrine disorders may be due to abnormal tropic signals from the pituitary gland (secondary disorder) or to dysfunction of target glands (primary disorder).

SUMMARY

The endocrine system, together with the nervous system, is responsible for coordinating cellular activity between many body systems and organs. Hormone secretion occurs in response to a variety of stimuli, including psychological or physiologic stress, electrolyte and metabolite levels, and normal circadian cycles. Increases or decreases in the quantity of a particular circulating hormone tend to regulate levels of that hormone through negative feedback mechanisms. The cellular responses to hormones are complex and controlled by many other factors in addition to the circulating hormone levels.

RESOURCES

General Endocrinology

Alberts B, Alberts B, et al: Cell signaling. In *Molecular biology of the cell*, ed 6, New York, 2015, Garland Science, pp 813–888.

Hall JE: Introduction to endocrinology. In Hall JE, editor: *Guyton and Hall textbook of medical physiology*, ed 13, Philadelphia, 2015, Saunders, pp 925–938.

Koeppen BM, Stanton BA: The male and female reproductive systems. In Koeppen BM, Stanton BA, editors: *Berne & Levy physiology*, ed 6, Philadelphia, 2010, Mosby, pp 758–798.

Van der Deure WM, Peeters RP, Visser TJ: Molecular aspects of thyroid hormone transporters, including MCT8, MCT10 and OATPs, and the effects of genetic variation in these transporters. *J Mol Endocrinol* 44:1–11, 2010.

White BA, Porterfield SP: Introduction to the endocrine system. In White BA, Porterfield SP, editors: *Endocrine and reproductive physiology*, ed 4, Philadelphia, 2013, Mosby, pp 1–25.

Hypothalamic-Pituitary System, Thyroid and Adrenal Glands

Hall JE: Adrenocortical hormones. In Hall JE, editor: *Guyton and Hall textbook of medical physiology*, ed 13, Philadelphia, 2015, Saunders, pp 965–982.

Koeppen BM, Stanton BA: The adrenal gland. In Koeppen BM, Stanton BA, editors: *Berne & Levy physiology*, ed 6, Philadelphia, 2010, Mosby, pp 738–757.

Koeppen BM, Stanton BA: The hypothalamus and pituitary gland. In Koeppen BM, Stanton BA, editors: *Berne & Levy physiology*, ed 6, Philadelphia, 2010, Mosby, pp 706–724.

Koeppen BM, Stanton BA: The thyroid gland. In Koeppen BM, Stanton BA, editors: *Berne & Levy physiology*, ed 6, Philadelphia, 2010, Mosby, pp 725–737.

Martini FH, Ober WC, Bartholomew EF, Nath JL: *Visual essentials of anatomy & physiology*, 2013, Pearson.

White BA, Porterfield SP: The adrenal gland. In White BA, Porterfield SP, editors: *Endocrine and reproductive physiology*, ed 4, Philadelphia, 2013, Mosby, pp 147–176.

White BA, Porterfield SP: The hypothalamus-pituitary complex. In White BA, Porterfield SP, editors: *Endocrine and reproductive physiology*, ed 4, Philadelphia, 2013, Mosby, pp 99–128.

White BA, Porterfield SP: The thyroid gland. In White BA, Porterfield SP, editors: *Endocrine and reproductive physiology*, ed 4, Philadelphia, 2013, Mosby, pp 129–146.

40

Disorders of Endocrine Function

Jacquelyn L. Banasik

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- How can primary and secondary endocrine disorders be differentiated?
- What etiologic factors would lead to clinical manifestations of hormone excess or deficiency?
- What are the etiologic factors, clinical findings, and management of excess and deficiency of the following endocrine hormones: growth hormone, thyroid hormone, adrenocortical hormones, adrenal medullary hormones, parathyroid hormone, and antidiuretic hormone?

CHAPTER OUTLINE

Basic Concepts of Endocrine Disorders, 799
Etiology of Endocrine Disorders, 799
Classification of Endocrine Disorders, 800
Growth Hormone Disorders, 801
Growth Hormone Deficiency, 801
Growth Hormone Excess, 802
Thyroid Hormone Disorders, 803
Hypothyroidism, 803
Hyperthyroidism, 804
Adrenocortical Hormone Disorders, 806
Adrenocortical Insufficiency, 806
Congenital Adrenal Hyperplasia, 808
Hypercortisolism, 808
Hyperaldosteronism, 809
Adrenal Medulla Disorder, 810
Pheochromocytoma, 810
Parathyroid Gland Disorders, 811
Regulation and Actions of Parathyroid Hormone, 811
Hyperparathyroidism, 811
Hypoparathyroidism, 812
Antidiuretic Hormone Disorders, 812
Diabetes Insipidus, 812
Syndrome of Inappropriate Antidiuretic Hormone Secretion, 813

Together with the nervous system, the endocrine system (the glands and the hormones they secrete) regulates body processes involving growth, maturation, metabolic functions, fluid balance, responses to stress, and reproduction. This regulation is carried out through the actions of the hormones produced and secreted by the endocrine cells. Endocrine hormones are chemical messengers that travel through the bloodstream to exert physiologic effects on specific target cells and tissues. In the healthy state, hormones are released by endocrine glands when their action is needed and inhibited when their effect is attained.

Endocrine disease is marked by either hyperfunction (excessively high blood concentrations of a hormone or conditions that mimic high hormone levels) or hypofunction (depressed levels or conditions that mimic low hormone levels). Some endocrine disorders have such striking characteristics that recognition is obvious. Other symptoms of endocrine disease may be nonspecific and more difficult to detect. Observing and interviewing skills are important because, with the exception of the thyroid and testicles, the endocrine glands cannot be directly examined. Laboratory diagnostic tests are especially important in assessing the endocrine system.

This chapter describes alterations in the anterior pituitary regulatory system, including growth hormone, thyroid hormone, and adrenal hormones, as well as parathyroid hormone disorders and posterior pituitary disorders of antidiuretic hormone (ADH; vasopressin) secretion. Disorders of prolactin and the gonadotropins (follicle-stimulating hormone and luteinizing hormone) are discussed in Unit IX. Disorders of insulin secretion are discussed in Chapter 41.

BASIC CONCEPTS OF ENDOCRINE DISORDERS

Etiology of Endocrine Disorders

Dysfunction, either hyposecretion or hypersecretion, may originate in the hypothalamus/pituitary, the hormone-producing gland, or the target tissue (Fig. 40.1). The etiology of endocrine disorders may be congenital, infectious, autoimmune, neoplastic, idiopathic, or iatrogenic. The onset

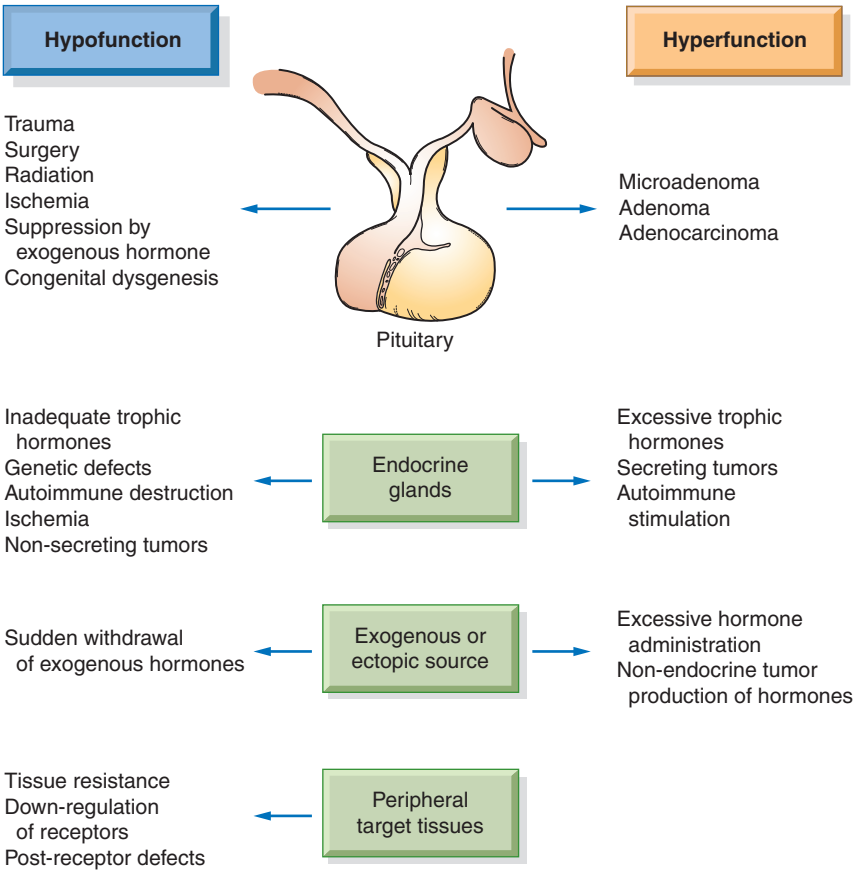


FIG 40.1 Common etiologies of endocrine disorders.

of the disorders can be slow and insidious or abrupt and life threatening. The age at onset may range from birth to old age.

Abnormal hormone production occasionally results from an inborn genetic defect. Such defects may cause excessive production of hormone precursors because of an enzymatic block in the synthetic pathway and enzyme deficiencies that impair hormone synthesis. An important example is congenital adrenal hyperplasia in which infants develop enlarged adrenal glands, but have a deficiency of cortisol production. If genetic defects do not cause a complete block of synthesis, increased pituitary stimulus may compensate by causing glandular hyperplasia, resulting in near-normal hormone levels.

Autoimmune disorders commonly cause endocrine dysfunction, particularly in women. The pathogenesis of autoimmunity is incompletely understood, but involves both a genetic predisposition and an environmental trigger (see Chapter 10). Antibodies are produced against certain antigens on self tissue cells, resulting either in hyperfunction of the endocrine gland (e.g., hyperthyroidism of Graves disease) or in immune destruction of the gland, eventually leading to hypofunction (e.g., adrenal insufficiency of Addison disease).

Hormones may be produced by abnormal tissue sites. Such *ectopic* hormone production is usually associated with a malignant tumor. Although different tumors can produce hormones, some cell types are more commonly associated with specific tumors. For example, some lung tumors produce ADH, leading to water intoxication and hyponatremia. Endocrine disorders can also be classified as functional disorders caused by nonendocrine disease such as chronic renal failure, liver disease, or heart failure.

In some cases, endocrine disease occurs when the target tissue fails to respond to a hormone. The presence of normal or elevated hormone levels without normal hormonal action indicates target tissue resistance. This problem is also demonstrated by diminished or absent response to the administration of exogenous hormones. The mechanisms of hormone resistance may be genetic or acquired and may include defects at receptor sites, antibody reaction to hormone receptors, and defective postreceptor hormone action.

Finally, some endocrine disorders may be induced by medical treatments, such as therapy for a nonendocrine disorder. These iatrogenic disorders can be caused by chemotherapy, radiation therapy, or surgical removal of glands. Commonly, a treatment for endocrine hyperfunction involves removal or destruction of glandular tissue with resultant chronic hypofunction. Long-term hormone replacement therapy may then be needed.

Classification of Endocrine Disorders

Endocrine disorders involving control by the anterior pituitary gland commonly are classified as *primary* (intrinsic malfunction of the hormone-producing target gland) or *secondary* (malfunction of the hypothalamus/pituitary cells that control the hormone-producing target gland). The clinical presentation of an endocrine disorder of primary or secondary etiology is similar; however, diagnosing the source of the problem may be important in determining the best treatment. Measurement of serum concentrations of pituitary and target gland hormones allows differentiation between primary and secondary endocrine etiologies.

Clinically useful laboratory measures are available for diagnosing and monitoring hormone disorders, including those affecting thyroid and adrenal function. Laboratory diagnosis is based on an understanding of the feedback loop communication between the hypothalamic-pituitary

system and the target gland. When the *primary* (target) gland fails, inadequate hormone is produced and low levels of hormone are present in the circulation, but blood levels of the corresponding trophic pituitary hormone become very elevated (Fig. 40.2). For example, in primary hypothyroidism, the thyroid fails to secrete thyroid hormones and serum levels of thyroxine (T_4) become lower. Thyroid-stimulating hormone (TSH) levels rise as the pituitary gland attempts to stimulate the malfunctioning thyroid. In contrast, in secondary hypothyroidism, the pituitary gland fails to release TSH, secondarily reducing thyroid gland production, so both thyroxine and TSH levels are abnormally low in the circulation. It is important to recall that the hormones released by the target gland are the ones that produce clinical signs and symptoms and that they are the starting point for interpretation of laboratory test results.

KEY POINTS

- Endocrine disorders occur because of hypersecretion, hyposecretion, or nonresponsiveness by target cells.
- *Hypersecretion* is usually due to secreting tumors, autoimmune disease, or excessive stimulation of the gland by trophic signals.
- *Hyposecretion* may be due to failure or congenital absence of glandular tissue, autoimmune destruction, surgical removal of the gland, or lack of normal trophic signals.
- *Hyporesponsiveness* is clinically similar to hyposecretion and is from target tissue dysfunction. This phenomenon is called *tissue resistance.*
- Endocrine disorders involving the hypothalamic-pituitary system are often classified as primary or secondary.
- Primary endocrine disorders result from intrinsic defects within the hormone-secreting target gland.
- Secondary disorders result from abnormal hypothalamic-pituitary secretion of trophic signals. Manifestations of an endocrine disorder are due to abnormal target gland function and are therefore similar whether the etiologic classification is primary or secondary.

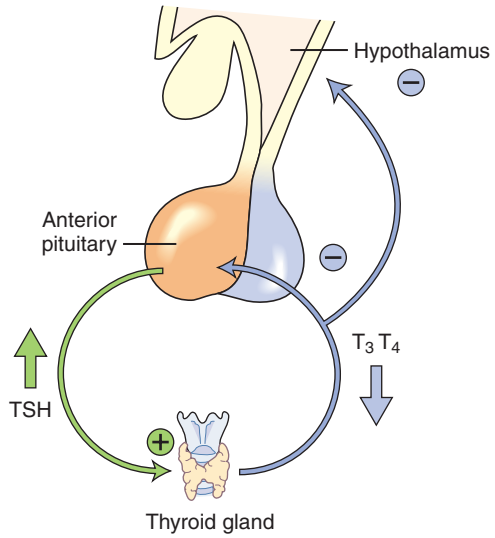


FIG 40.2 Diagnosing endocrine disorders as primary or secondary is based on the hypothalamic-pituitary and target tissue feedback loops. In this example, the low T_3 and T_4 levels are indicative of hypothyroidism. The elevated TSH level indicates that the pituitary gland is responding appropriately and is not the source of the hypothyroidism. A diagnosis of primary hypothyroidism would be made on the basis of these laboratory results.

GROWTH HORMONE DISORDERS

Growth hormone (GH) is produced in the anterior pituitary gland under the influence of hypothalamic releasing (GH–releasing hormone) and inhibiting (somatostatin) factors. Its primary target organ is the liver, but GH also has direct effects on several tissue types. In general, GH increases lean body mass, reduces fat mass, and induces the liver to release glucose under conditions of hypoglycemia. Many of the effects of GH are mediated by a peptide called *IGF-1* (insulin-like growth factor-1) that is released from the liver when stimulated by GH. Please refer to Chapter 39 for details of GH synthesis, regulation, and activity. The major signs and symptoms of GH imbalance are summarized in Box 40.1.

Growth Hormone Deficiency

Etiology and pathogenesis. Deficiencies in GH secretion can be classified into several major categories: (1) decreased GH secretion, (2) defective GH action (structurally abnormal GH or defective GH receptor), and (3) defective IGF-1 (somatomedin) generation.

GH deficiency is most clinically relevant in children. A birth history of prolonged labor or breech delivery is common, but GH deficiency may also be present in children who are born with midline craniocerebral defects, most likely attributable to congenital malformations or as sequelae of a chromosomal anomaly. Deficiencies in GH and other pituitary hormones should be considered in any child with nystagmus, retinal abnormalities, and other midline or midfacial abnormalities, such as cleft lip or palate. The association between GH deficiency and other midline abnormalities appears to occur because the pituitary gland is developing during the same stage of fetal life as the other midline structures.

Children with GH deficiency have a variety of presentations, depending on the cause of the deficiency, the age at onset, and the severity of the disorder. The basis for this defect may be failure of the hypothalamus to stimulate pituitary GH secretion or failure of the pituitary to produce GH.

The most common tumors to influence hypothalamic-pituitary function are midline brain tumors. These include gliomas of the optic nerve and craniopharyngiomas. Craniopharyngiomas arise from cells at the junction of the anterior and posterior pituitary gland, are believed to be present at birth, and are slow growing. Craniopharyngiomas may grow to a large size without producing typical signs of increased intracranial pressure (vomiting, headache, oculomotor abnormalities). In older children, delayed growth may be the first sign of a craniopharyngioma.

BOX 40.1 Signs and Symptoms of Growth Hormone Imbalance

Growth Hormone Excess	Growth Hormone Deficiency
Children	
Increased linear growth and tall stature	Delayed growth Fine features Short stature, proportionate
Adults	
Soft tissue hyperplasia Increased bone density Large hands, feet Coarse facial features Thick, leathery skin Weight gain Glucose intolerance	May be associated with hyposecretion of other pituitary hormones

Radiation therapy for brain tumors or leukemia may cause damage to hypothalamic and pituitary function. Traumatic insult to the skull or sella turcica may damage the pituitary gland, interrupting vascular connections and hypothalamic stimulation.

Clinical manifestations. GH-deficient infants usually have normal birth length and weight. They may manifest hypoglycemia because GH and cortisol are necessary to maintain the euglycemic state. Hypoglycemia may present after fasting, which could be as brief as 3 hours in an infant. Recurrent episodes of hypoglycemia may lead to seizures and permanent cerebral damage. Boys developing GH deficiency in utero may have micropenis and undescended testicles.

GH-deficient children fall below the third percentile of growth in comparison with their peers. Dental eruption is delayed, and the development and setting of the permanent teeth are irregular. The hair is thin, and the nail growth is poor. Older children have greater fat mass and decreased muscle mass, with delayed bone formation. Delayed puberty is common if other anterior pituitary hormones are also affected.

Children's growth should be evaluated annually. If growth velocity is abnormal, an endocrinologic evaluation and physiologic tests to stimulate GH release can be planned. Many pharmacologic agents are available that stimulate GH secretion in children, including insulin, arginine, levodopa, and clonidine. Children's neurosecretory GH patterns can be studied by obtaining timed serum GH samples during a normal nighttime sleep cycle.

Treatment. Hormonal replacement therapy for GH-deficient children has been available for several decades. The most obvious effect of GH replacement is stimulation of linear growth in children whose bones have not yet fused. With treatment, children experience an increase in growth velocity, as well as depletion of the excess fat stores noted in GH-deficient children.

Adults may become GH deficient after resection of pituitary tumors or after traumatic head injuries. There is controversy regarding the manifestations of GH deficiency acquired in adulthood. GH-deficient adults may have diminished lean body mass, hypercholesterolemia, and decreased bone density.

Growth Hormone Excess

Etiology and pathogenesis. GH excess is nearly always because of uncontrolled production of the hormone by a benign somatotropic tumor in the pituitary gland (adenoma). GH stimulates the liver to produce IGF-1, and these two hormones act in concert to cause up-regulated growth of soft and bony tissues. Because GH secretion varies significantly over the course of the day, the serum level of IGF-1, which is more stable, may be measured as an indicator of GH secretion. An elevated IGF-1 level is a useful indicator of GH hypersecretion. If the tumor presents in childhood before the skeletal epiphyses are closed, rapid growth results in *pituitary giantism.* These children experience markedly accelerated growth velocity and quickly exceed the ninety-fifth percentile on pediatric growth charts. When the disorder is allowed to progress untreated, some of these children may grow to 8 feet or more in height and usually suffer an early cardiovascular death related to cardiomegaly and heart failure.

In adults, GH excess is called *acromegaly* and may be clinically subtle. Acromegaly occurs with equal frequency in men and women during the fourth and fifth decades of life. After the skeletal epiphyses close, bony growth increases bone density and thickening of the short bones, such as the hands and feet.

Clinical manifestations of acromegaly. Patients usually notice increased ring and shoe sizes, which progressively advance over several years. Enlargement of the frontal sinus causes a prominent brow, and growth of the mandible results in progressive underbite *(prognathism)* (Fig. 40.3). Soft tissues also slowly hypertrophy, causing coarsening of facial features and skin tags. Internal organs increase in size, resulting in goiter (thyroid enlargement) and cardiomegaly. Other manifestations include deepening of the voice secondary to vocal cord thickening and enlargement of the tongue, resulting in sleep apnea. Colonic polyps become more common, with the potential for malignant degeneration. It is estimated that the average patient with acromegaly has an active pituitary tumor for 7 or more years before seeking evaluation. Most often the changes are attributed by the patient and his or her family as "just growing older." Abnormalities in bone and soft tissue growth are mostly irreversible. Patients may develop symptoms of increased intracranial pressure if the pituitary tumor enlarges significantly, including headache and visual disturbances. Some patients may develop glucose intolerance or hyperglycemia, and GH has been called a *diabetogenic* hormone for this reason.

Treatment. Effective therapy for acromegaly involves surgically removing the tumor while counteracting the effects of excess GH. Octreotide, a synthetic form of somatostatin, suppresses production of GH. GH receptor antagonists are also available (pegvisomant) that block the effect of GH on target tissues. Surgery is usually performed using a transsphenoidal approach, but often the tumor is too large to completely resect. Radiation may be used postoperatively in these cases.

FIG 40.3 Progressive development of facial features of acromegaly. (From Lewis SM et al: *Medical-surgical nursing,* ed 10, St Louis, 2017, Elsevier. Courtesy Linda Haas, Seattle, WA.)

KEY POINTS

- Hyposecretion of growth hormone (GH) results in decreased linear growth in children. In some cases decreased linear growth occurs despite normal GH levels, and abnormalities of IGF-1 generation or responsiveness are suspected.
- GH deficiency may be idiopathic or related to tumors, radiation, or trauma. The diagnosis is confirmed by a finding of decreased GH levels in the blood and deficient GH release in response to hypoglycemia or other stimulants.
- Excessive GH production is usually due to pituitary adenoma. Excess GH during childhood results in increased linear growth and giantism. Excess GH secretion after closure of bone epiphyses results in increased bulk and acromegaly.
- Features of acromegaly include a protruding jaw, increased bone density, increased growth of soft tissues (e.g., nose, ears), and large hands and feet. Excessive GH secretion causes persistent hyperglycemia and increased insulin production in some individuals.
- Serum IGF-1 measurement is used to assess for acromegaly. High IGF-1 concentration and an elevated GH level that is not suppressed by administration of oral glucose aid in the diagnosis. Treatment entails surgical removal or pharmacologic inhibition of GH release or tissue actions.

THYROID HORMONE DISORDERS

Secretion of the thyroid hormones, triiodothyronine (T_3) and thyroxine (T_4), is under control of TSH secretion from the anterior pituitary gland. In turn, TSH secretion from the pituitary is under control of thyroid-releasing hormone (TRH) from the hypothalamus. Thyroid hormones are important for normal growth and development of tissues throughout the body and an important regulator of metabolism. The details of thyroid hormone synthesis, regulation, and activity are described in Chapter 39.

Hypothyroidism

Etiology and pathogenesis. Hypothyroidism may be congenital in origin or acquired later in life. The great majority of cases of hypothyroidism are primary, due to intrinsic dysfunction of the thyroid gland. Congenital hypothyroidism may result from a variety of causes. Thyroid dysgenesis (lack of thyroid gland development) accounts for most of the cases of congenital hypothyroidism. Abnormal TSH receptors and defective synthesis of thyroid hormone are other mechanisms causing congenital hypothyroidism. Congenital hypothyroidism that results in significant defects in mental and physical development may be referred to as *cretinism*.

Lymphocytic thyroiditis (Hashimoto thyroiditis or autoimmune thyroiditis) is the most common cause of acquired hypothyroidism. Lymphocytic thyroiditis is characterized by an enlarged thyroid gland (Fig. 40.4) caused by lymphocytic infiltration. Thyroid hormone production decreases, stimulating the release of TSH from the pituitary gland and resulting in elevated serum TSH levels. Hypothyroidism and its clinical symptoms progress as the gland becomes fibrotic.

Other causes of acquired hypothyroidism include irradiation of the thyroid gland, surgical removal of thyroid tissue, and iodine deficiency. Iodine is essential for the formation of T_4 and T_3. Lack of iodine prevents production of both T_4 and T_3 but does not stop the formation of thyroglobulin. As a result, insufficient hormone is available to inhibit production of TSH by the anterior pituitary gland. The elevated TSH level then causes the thyroid cells to secrete excessive amounts of thyroglobulin (colloid) into the follicles, and the gland grows larger and larger, producing a goiter (see Fig. 40.4). An enlarged thyroid (goiter) is not always associated with hypothyroidism and can be present in euthyroid and hyperthyroid states.

Some foods contain "goitrogenic" substances that interfere with thyroid hormone synthesis. Such goitrogenic substances occur in some varieties of turnips and cabbage, but the clinical significance of these goitrogens is considered to be minimal. The drug lithium inhibits thyroid hormone synthesis and secretion and causes hypothyroidism in up to 20% of patients.

Secondary hypothyroidism is caused by defects in TSH production and is uncommon. Individuals who have been exposed to severe head trauma, cranial neoplasms, brain infections, cranial irradiation, and neurosurgery can be left with secondary hypothyroidism.

Clinical manifestations. Routine screening of newborns has resulted in early treatment of most infants with congenital hypothyroidism. Few clinical manifestations are present at birth. In untreated infants, symptoms appear in the first months of life and include a dull appearance; a thick, protuberant tongue; and thick lips (leading to feeding difficulties). Other signs include prolonged neonatal jaundice, poor muscle tone, bradycardia, mottled extremities, umbilical hernia, and a hoarse cry. Thyroid hormone is essential for normal central nervous system development; significant and irreversible intellectual disability will occur unless thyroid hormone replacement therapy is started early in infancy. Older children who acquire hypothyroidism have essentially the same clinical manifestations as seen in adults. In addition, growth retardation, delayed bone development, and delayed or precocious puberty may occur.

In general, individuals with hypothyroidism have decreased basal metabolic rates as the basis for many of their signs and symptoms. Patients report subjective feelings of weakness, lethargy, cold intolerance, and decreased appetite. Bradycardia, narrowed pulse pressure, and mild to moderate weight gain may occur. Elevated levels of serum cholesterol

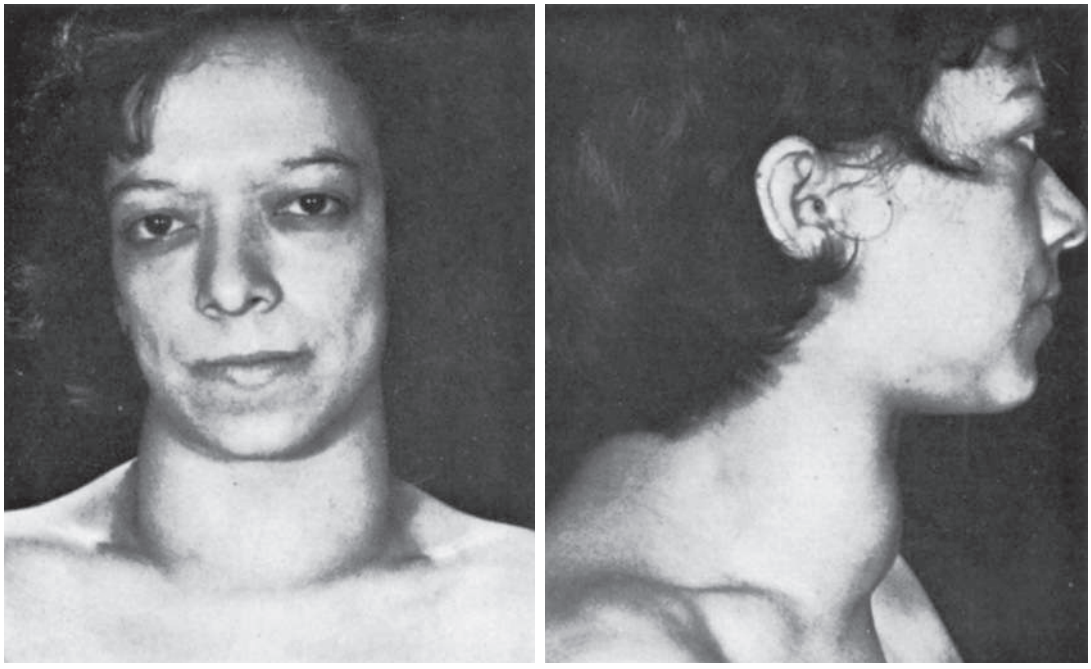


FIG 40.4 An enlarged thyroid gland (goiter) can be present in hypothyroid, hyperthyroid, and euthyroid states. Note enlargement at the base of the neck. (From Wilson JD, Foster DW, editors: *Williams textbook of endocrinology,* ed 8, Philadelphia, 1992, Saunders, p 425.)

and triglycerides are common, as is an increased incidence of atherosclerosis. The thyroid gland may become enlarged (goitrous), the skin may be cool and dry, and constipation may be present. Depression and difficulties with concentration and memory occur. Women with acquired hypothyroidism may experience menstrual irregularities, with increased flow and clotting. Box 40.2 summarizes the general signs and symptoms of thyroid imbalance.

Myxedema occurs in severe or prolonged thyroid deficiency. The term is in reference to the generalized, nonpitting edema that patients with long-term hypothyroidism tend to demonstrate. The edematous-looking skin is from the accumulation of glycosaminoglycans (mucopolysaccharides) in the interstitial spaces, which then retain fluid. Thyroid hormone normally prevents the accumulation of glycosaminoglycans within subcutaneous tissues. Individuals with hypothyroid-induced myxedema usually present in an altered mental state, with alterations in thermoregulation and a history of a precipitating event such as sepsis, trauma, or the use of certain medications. Without medical intervention, patients may lapse into so-called *myxedema coma,* a medical emergency with significant mortality. Fig. 40.5 shows the typical features of patients with myxedema.

The diagnosis of hypothyroidism can be confirmed by measuring serum TSH and thyroid hormone levels. Because the most common cause of hypothyroidism is thyroid failure (primary hypothyroidism), nearly all patients will have elevated TSH levels because of the lack of negative feedback exerted by T_3 and T_4. The serum TSH level is a sensitive indicator of thyroid hypoactivity, and an increase in TSH level often is detectable long before many symptoms develop. A decline in T_4 and T_3 levels may not occur until later in the course of disease. In the rare case of hypothalamic-pituitary dysfunction, both serum TSH and serum T_4 concentrations will be inappropriately low (Table 40.1).

BOX 40.2 Typical Signs and Symptoms of Thyroid Imbalance

Hyperthyroidism	Hypothyroidism
Sleeplessness, nervousness	Lethargy
Muscle weakness, fatigue	Weakness
Susceptibility to infection	Dry, pale, cool, coarse skin
Skin texture warm, silky, damp	Cold intolerance
Heat intolerance	Weight gain
Increased appetite with weight loss	Constipation
Increased gastric emptying, increased intestinal motility	Bradycardia, wide pulse pressure
Tachycardia, narrow pulse pressure, palpitations, angina	Dyspnea, chest pain
Dyspnea	Thyroid may be diffusely enlarged (goiter) or may not be palpable
Enlarged thyroid, may be diffuse or nodular	Hair coarse
Hair silky, nail loose or detached from nail bed	Sluggish return of reflexes; mental impairment: slowed cognitive ability, poor memory, forgetfulness, depressed affect; deafness (in one third of population)
Hyperreflexia, fine tremor	Facial edema (especially periorbital); thinned lateral aspect of eyebrows
Eye symptoms: burning, tearing, diplopia, lid lag, prominent eyes (exophthalmia with Graves disease), stare, eyelid tremors when closed	Heavy, prolonged menses; infertility; decreased libido
Absence of forehead wrinkling on upward gaze	
Decreased or absent menses	

Treatment. The goal of treatment is to return the individual with congenital or acquired hypothyroidism to a euthyroid state. When serum thyroid levels are replaced too quickly, patients may experience insomnia, anxiety, and mood lability. Once treatment has begun, those individuals with a goiter usually experience a regression in glandular enlargement.

Oral levothyroxine is used to replace or supplement hormone production from an underactive thyroid. Patients notice an increase in exercise tolerance, decreased fatigue, and improved mentation with therapy. Resolution of symptoms occurs gradually over weeks. As therapy returns the patient to a "euthyroid" state, the serum TSH level should return to the normal range. Overtreatment with thyroid hormone can be detected by a serum TSH level that is below the normal range. Overtreatment should be avoided in part because it contributes to osteoporosis.

Hyperthyroidism

Etiology and pathogenesis. Mechanisms that produce hyperthyroidism include thyroid follicular cell hyperfunction with increased synthesis and secretion of T_4 and T_3 (e.g., Graves disease), thyroid follicular cell destruction with release of preformed T_4 and T_3 (e.g., Hashimoto thyroiditis), and ingestion of excessive thyroid hormone. The increased serum levels of thyroid hormones increase the metabolic rate (see Table 40.1).

Hyperfunction of thyroid follicular cells can be either autonomous (primary) or mediated through stimulation of TSH receptors by TSH

FIG 40.5 Typical facial puffiness and dull expression of patients with myxedema. (From Seidel HM et al: *Mosby's guide to physical examination,* ed 6, St Louis, 2006, Mosby. Courtesy Paul W. Ladenson, MD, The Johns Hopkins University and Hospital, Baltimore, MD.)

TABLE 40.1 Thyroid Hormone Levels in Various States

State	Serum T_4 (μg/dL), Range	Serum T_3 (ng/dL), Range	Serum TSH (μU/mL), Range
Euthyroid	4.5–11.5	60–180	0.5–4.5
Infants (<2 wk)	8.0–15.0	—	0.5–4.5
Children (prepubertal)	6.5–11.5	80–220	0.5–4.5
Hyperthyroid (primary)	>11.5	—	<0.15
Hypothyroid (primary)	variable	—	>5.0

(secondary). Primary hyperfunction can be caused by adenomas and, rarely, thyroid carcinoma. Inflammation of thyroid follicular cells, with release of preformed thyroid hormone, can be associated with viral or autoimmune processes. Examples are the toxic thyroiditis of Hashimoto disease and subacute thyroiditis. Hashimoto thyroiditis typically presents first with hyperthyroid symptoms because the injured thyroid gland releases stored hormone and then progresses to a hypothyroid state when the autoimmune process destroys the gland. Acute or chronic ingestion of thyroid hormone preparations can produce excess levels of thyroid hormones.

The most common etiology of hyperthyroidism is from autoantibodies that bind and stimulate TSH receptors on the thyroid gland. This stimulation leads to a diffuse toxic goiter and a type of primary hyperthyroidism called Graves disease. The etiology of Graves disease is autoimmune, and it has been associated with certain genetic markers (e.g., *HLA B8, HLA DR3*). There is an increased incidence during the second and third decades of life. In response to the high circulating T_3 and T_4 levels that are stimulated by autoantibodies, the pituitary gland stops producing TSH, and serum TSH falls to very low levels.

Clinical manifestations. Symptoms of hyperthyroidism are from an increase in metabolic rate and enhanced sympathetic nervous system activity. Symptoms may be mild to severe and include insomnia, restlessness, tremor, irritability, palpitations, increased heart rate, heat intolerance, diaphoresis, and an inability to concentrate that interferes with work performance (see Box 40.2). Increased basal metabolic rate may result in weight loss, even though appetite and dietary intake increase. In women, amenorrhea or scant menses is a frequent finding. The term *thyrotoxicosis* is used to describe a hyperthyroid condition that is associated with significant symptoms of the disorder.

Individuals with Graves disease usually present with thyromegaly (diffusely enlarged thyroid), thyrotoxicosis, and, often, exophthalmos (enlargement of retroorbital muscles causing protrusion of the eyes) (Fig. 40.6). Spasm and retraction of the eyelids lead to widening of the palpebral fissure, resulting in exposed sclera. Lid lag develops, and severe, progressive exophthalmos may occur. Eye complaints may include vision changes and photophobia. The exophthalmos of Graves disease is not from elevated levels of thyroid hormones, but is thought to develop as a result of autoimmune injury to retroocular structures. Tissues behind the eye become infiltrated with immune cells, followed by release of inflammatory cytokines that stimulate local cells to secrete excessive glycosaminoglycans. The tissue behind the eye swells and pushes the eyeball forward. Reduction of circulating thyroid hormone levels often does not prevent progression of the exophthalmos.

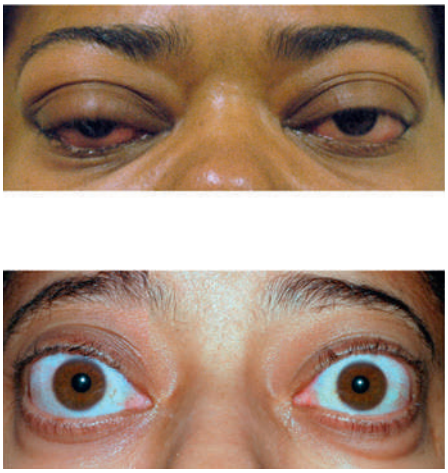


FIG 40.6 Exophthalmos typical of Graves disease. Note lid retraction and eye protrusion. (From Cruz AAV et al: Graves upper eyelid retraction. *Survey of Ophthalmology* 2013;58(1):63–76, Elsevier Inc.)

Undetectable TSH levels are the best indicator of primary hyperthyroidism. Serum T_4 and T_3 levels are elevated. A 24-hour radioactive iodine uptake study can confirm the diagnosis of Graves disease, when the scan shows diffuse homogeneous uptake of tracer, and can exclude the presence of thyroid neoplasms.

Thyroid storm (accelerated hyperthyroidism) is a form of life-threatening thyrotoxicosis that occurs when excessive amounts of thyroid hormones are acutely released into the circulation. This may occur under conditions of psychological or physiologic stress or with physical manipulation of the gland during diagnostic or surgical procedures. Thyroid storm presents with the clinical features of elevated temperature, significant tachycardia, cardiac dysrhythmias, and congestive heart failure. Extreme restlessness, agitation, and psychosis may occur. Prompt treatment of thyroid storm is required.

Treatment. Depending on the degree of symptoms, hyperthyroidism may be treated with medications, such as β-blockers to block acute symptoms or drugs to inhibit thyroid hormone production, or with radioactive iodine to destroy part of the thyroid gland. Surgical removal of the thyroid gland usually is reserved for hyperthyroidism associated with tumors and may result in acute hypocalcemia from inadvertent resection of one or more parathyroid glands. In the rare case of hyperthyroidism secondary to pituitary adenoma, surgical or laser destruction of the anterior pituitary gland may be undertaken.

Antithyroid drugs called *thionamides* (e.g., methimazole, propylthiouracil) block the activities of the enzyme thyroid peroxidase and inhibit thyroid hormone production by follicular cells. These drugs do not prevent release of hormone, and because the typical thyroid gland contains a 2- to 3-month supply of previously synthesized hormone, the onset of symptom reduction is slow. Thionamides produce a gradual reduction in the basal metabolic rate and reduction of symptoms. Relapse frequently occurs once medications are stopped.

Radioactive iodine treatment to ablate the gland, thereby curtailing its ability to produce excess thyroid hormones, is the treatment of choice for Graves disease. Hypothyroidism occurs after radioactive iodine therapy in 50% to 80% of patients, so patients should be prepared for the likelihood of lifelong thyroid hormone replacement therapy.

Because thyroid storm is a life-threatening form of thyrotoxicosis, urgent management is needed. Antithyroid drugs such as methimazole may be used, followed by iodine administration to further inhibit release of T_3 and T_4. High-dose iodine paradoxically inhibits thyroid hormone synthesis. β-Blockers used for their antiadrenergic effects on the heart also inhibit the peripheral conversion of T_4 to T_3. Antipyretic therapy (e.g., use of cooling blankets or ice packs, administration of acetaminophen) may be started to achieve peripheral cooling. Fluid replacement may be needed, and the cardiovascular status should be monitored. Thyroid storm can be fatal if untreated. With treatment, the mortality is between 10% and 30%.

KEY POINTS

- Thyroid hormone (T_3, T_4) is produced in follicular cells of the thyroid gland. The synthesis and secretion of thyroid hormone are stimulated by thyroid-stimulating hormone (TSH) from the pituitary gland. TSH release from the pituitary gland is stimulated by TRH from the hypothalamus. Thyroid hormone is an important stimulator of growth and cellular metabolism.
- Hypothyroidism may be primary (due to congenital agenesis, autoimmune destruction, irradiation, trauma, surgical removal of the gland, or iodine deficiency) or secondary to pituitary hyposecretion of TSH.

- TSH level is helpful in differentiating between primary (high TSH level) and secondary (low TSH level) causes of hypothyroidism. Hypothyroidism is nearly always from a primary etiology.
- Manifestations of hypothyroidism are attributable to a generalized decrease in metabolism and include nonpitting edema (myxedema), slowed mentation, weight gain, dry skin, constipation, decreased heart rate, decreased pulse pressure, lethargy, and loss of the outer third of the eyebrow. Severe hypothyroidism may lead to myxedema coma, characterized by bradycardia, hypothermia, hypotension, and decreased level of consciousness. Untreated congenital hypothyroidism results in profound mental and physical retardation (cretinism). Treatment centers on hormone replacement therapy.
- Hyperthyroidism may be primary (Graves disease, autoimmune, tumor related, inflammatory) or secondary, attributable to pituitary hypersecretion of TSH. The blood level of TSH is helpful in differentiating primary (low TSH level) from secondary (high TSH level) hyperthyroidism. High levels of T_3 and T_4 confirm the diagnosis of hyperthyroidism.
- The manifestations of hyperthyroidism result from a generalized increase in metabolism. Hyperactivity, irritability, insomnia, weight loss, increased appetite, heat intolerance, diarrhea, and palpitations are common. Most individuals have a detectably enlarged thyroid gland. Exophthalmos is immune mediated and occurs with Graves disease.
- Thyroid storm may be precipitated by stress or manipulation of the gland. It is characterized by tachycardia, hypertension, high temperature, and cardiac dysrhythmias. Treatment includes β-blockers to control cardiovascular symptoms, antithyroid drugs to reduce thyroid production, radioactive iodine to ablate the gland, and surgical removal of tumors.

ADRENOCORTICAL HORMONE DISORDERS

The adrenal cortex synthesizes three different classes of steroid hormones, including glucocorticoids, mineralocorticoids, and androgens. Cortisol is the primary glucocorticoid, and its concentration in the circulation provides negative feedback regulation of the hypothalamus and pituitary gland to suppress corticotropin-releasing hormone (CRH) and adrenocorticotropic hormone (ACTH) release. Aldosterone secretion is not regulated by the hypothalamic-pituitary system and is instead regulated by the presence of angiotensin II in the circulation. Androgenic hormone secretion by the adrenal cortex plays a relatively minor role in the development and maintenance of secondary sex characteristics, except in children with adrenogenital syndromes, which produce virilization in the female and precocious sexual development in the male. Physiologically, adrenal androgens are the main source of androgens in the female. As with mineralocorticoids, there is no known feedback mechanism to suppress ACTH production associated with adrenal sex hormone plasma levels. The details of adrenocortical hormone synthesis, regulation, and activity are discussed in Chapter 39.

Adrenocortical Insufficiency

Etiology and pathogenesis. Hyposecretion of adrenocortical hormones can result from disease of the adrenal cortex (primary adrenocortical insufficiency, Addison disease), from the inadequate secretion of ACTH from the anterior pituitary gland (secondary adrenal insufficiency), or from a lack of CRH secretion from the hypothalamus attributable to hypothalamic malfunction or injury (tertiary adrenal insufficiency). Although hyposecretion of all the adrenocortical hormones may occur, the most severe clinical manifestations of adrenocortical insufficiency occur because of inadequate levels of circulating cortisol.

The syndrome of congenital adrenal hyperplasia, a rare cause of adrenal insufficiency in pediatric populations, is due to specific enzymatic defects in the biosynthesis of cortisol by the adrenal glands. The lack of negative feedback results in overproduction of ACTH, leading to hyperplasia of the adrenal glands and excessive androgen secretion. Congenital adrenal hyperplasia is discussed at the end of the section on adrenal insufficiency.

Primary adrenal insufficiency (Addison disease) is caused by destruction of the adrenal cortex through idiopathic or autoimmune mechanisms, tuberculosis, trauma or hemorrhage of the adrenals (often associated with anticoagulant therapy), fungal disease (e.g., histoplasmosis), and neoplasia. Because of the high functional reserve, symptoms of adrenal insufficiency may not be recognized until 90% of the cortical tissue has been rendered nonfunctional.

Secondary adrenal insufficiency (hypothalamic-pituitary dysfunction) is usually iatrogenic in origin because of the large numbers of patients receiving corticosteroid therapy for chronic illnesses. Prolonged exposure to pharmacologic doses of exogenous corticosteroids suppresses CRH and ACTH stimulation of the adrenal gland through negative feedback. A lack of ACTH results in atrophy of the adrenal cortex. If corticosteroid administration is suddenly halted, or if the individual experiences a sudden stress-induced increase in need for cortisol, the adrenal gland will be unable to respond by increasing cortisol secretion. Acute and severe manifestations of adrenal insufficiency ensue. Secondary adrenal insufficiency occurs because of damage to the anterior pituitary gland or hypothalamus by tumors, infection, radiation, postpartum necrosis, trauma, or surgery.

Addisonian crisis, or acute adrenal insufficiency, represents a true medical emergency caused by inadequate levels of glucocorticoids and mineralocorticoids in the circulation. This may result from a slowly developing and unrecognized ACTH or cortisol deficiency in which secretion is adequate for the normal demands of life but inadequate for increased stress or trauma. Diminished vascular tone, reduced cardiac output, and inadequate circulating blood volume all contribute to potentially lethal vascular collapse. Hypotension, tachycardia, and symptoms of shock may occur.

Clinical manifestations. The clinical manifestations of adrenal insufficiency (Box 40.3) occur because of inadequate levels of circulating cortisol and aldosterone (Fig. 40.7). Clinical manifestations may appear gradually, especially if adrenal destruction is slow and incremental, such as in autoimmune adrenal insufficiency. Symptoms are more dramatic if adrenal destruction is sudden (hemorrhage) or if a stressor, such as trauma, causes sudden decompensation in a patient with chronic adrenal insufficiency.

Early signs of primary adrenal insufficiency include anorexia, weight loss, weakness, malaise, apathy, electrolyte imbalances, and hyperpigmentation of the skin caused by unsuppressed ACTH production (Fig. 40.8). ACTH is able to stimulate receptors on melanocytes and promote

BOX 40.3 Signs and Symptoms of Adrenocortical Hormone Imbalance

Cushing Syndrome	Adrenocortical Insufficiency
Truncal obesity	Weakness
Moon face	Hypotension
Dorsocervical fat pad	Hypoglycemia
Hirsutism	Hyperpigmentation (Addison disease)
Muscle wasting	Hyperkalemia
Striae	Weight loss
Petechiae	
Glucose intolerance	
Hypertension	
Hypokalemia	

pigment development in the skin. Salt craving may be present as a result of sodium deficit. If the condition is unrecognized or left untreated, gastrointestinal (GI) symptoms can develop, including nausea, vomiting, diarrhea, and dehydration. The patient may be hypotensive or tachycardic. The sudden onset of symptoms suggests acute adrenal insufficiency, which is a medical emergency.

Diagnosis. The diagnosis of acute adrenal insufficiency is assisted by the patient's medical history (use of steroids and/or anticoagulant therapy, previous trauma), physical examination, and laboratory findings. Decreased plasma cortisol levels assist in the diagnosis; however, because acute decompensation progresses to death so rapidly, cortisol samples often are obtained for laboratory analysis and then therapy is initiated presumptively. Resolution of symptoms may be remarkably rapid with administration of intravenous (IV) glucocorticoids.

In cases of chronic adrenal insufficiency, an ACTH provocation test can be given. Cosyntropin, synthetic ACTH, is given, and serum samples of cortisol are measured 30 and 60 minutes after administration. Serum cortisol levels should increase after this stimulus if the adrenal cortex is functioning normally. A failure to produce cortisol indicates a primary adrenal insufficiency. Abdominal computerized tomography (CT) or magnetic resonance imaging (MRI) may be performed to determine the size of the adrenal glands. Small adrenal glands occur with autoimmune destruction, whereas tuberculous glands are large and calcified, and hemorrhagic glands are large and smooth.

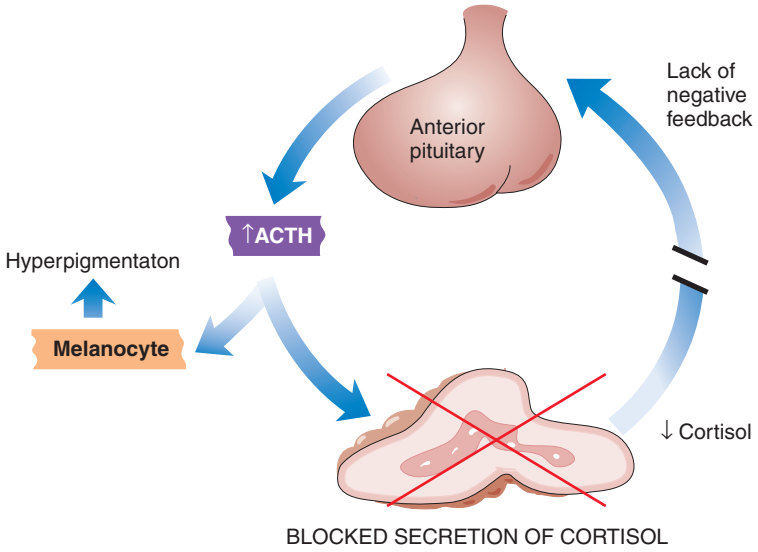


FIG 40.7 Primary adrenocortical insufficiency (decreased cortisol production) leads to hypersecretion of adrenocorticotropic hormone (ACTH) because of lack of negative feedback. ACTH binds to receptors on melanocytes and stimulates pigment development in the skin. Even though ACTH levels are high, the adrenal gland is unable to produce adequate levels of cortisol.

FIG 40.8 Altered pigmentation in adrenocortical insufficiency. **A,** Increased pigmentation across the bridge of the nose. **B,** Generalized hyperpigmentation with vitiligo. (From Bondy PK, Rosenberg LE: *Metabolic control and disease,* ed 8, Philadelphia, 1980, Saunders, p 1462.)

Treatment. The treatment for adrenal insufficiency entails replacing the absent or deficient hormones usually produced by the adrenal cortex in a manner that mimics natural production as closely as possible. About two-thirds of the daily dosage may be given in the morning and one-third in the evening to more closely mimic physiologic adrenal cortical function. In the case of adrenal crisis, IV glucocorticoids can be administered intermittently until the symptoms (hypotension, hypoglycemia) resolve; the dose is then titrated downward. In addition, volume replacement is needed to replace the increased urine output associated with lack of mineralocorticoid activity.

Stress situations increase the dose requirements for individuals receiving chronic replacement therapy. Acute illness (increased temperature causes an increase in metabolic rate), injury (e.g., trauma, surgery, burns), and psychological episodes that affect an individual's ability to function normally (death of a significant relative) may double or triple the needed daily corticosteroid dose. If illness or injury restricts the patient's ability to tolerate oral intake, replacement must be given parenterally.

Congenital Adrenal Hyperplasia

Congenital adrenal hyperplasia is also called *adrenogenital syndrome* and occurs when an enzyme needed for cortisol production is lacking because of a gene defect. The disorder usually follows an autosomal-recessive inheritance pattern. Because circulating cortisol levels are inadequate to provide negative feedback to the anterior pituitary gland, ACTH secretion is elevated. This leads to adrenal hypertrophy and overproduction of steroid precursors in the gland, which are shunted into the production of androgens. In the newborn, classic congenital adrenal hyperplasia is a life-threatening condition because of inadequate circulating cortisol.

Infants are frequently diagnosed at birth because of the effects of excessive androgens on the genitals of the newborn. Virilization of the genitalia of a female fetus occurs. The female infant may be born with an enlarged clitoris and fused labia, resembling a scrotal sac (Fig. 40.9).

FIG 40.9 Female infant with congenital adrenal hyperplasia demonstrating virilization of the genitalia. Note the enlarged clitoris and the fused labia, which resemble a scrotal sac. (From Hurwitz LS: Nursing implications of selected endocrine disorders, *Nurs Clin North Am* 1980;15:528.)

Male infants with congenital adrenal hyperplasia may have an enlarged penis and hyperpigmented scrotum, but the examiner may not recognize these subtle signs.

Depending on the enzymes affected, androgen overproduction may occur at any time from birth to early adult life. If it occurs in an adult female, she may develop such virile characteristics as a beard, a much deeper voice, baldness, masculine distribution of pubic hair, growth of the clitoris to resemble a penis, and deposition of proteins in the skin and muscles to yield typical masculine characteristics.

In adult men, the virilizing characteristics of adrenogenital syndrome are less obvious because masculine characteristics are normal and associated with testosterone secreted by the testes. Therefore the diagnosis is more difficult. In the prepubertal male, adrenogenital syndrome usually causes precocious puberty.

In all cases, treatment with glucocorticoids is necessary to avoid the complications associated with adrenal insufficiency as previously described. In addition, exogenous glucocorticoid therapy suppresses pituitary secretion of ACTH, allowing the adrenal cortex to atrophy and the overproduction of adrenal androgens to cease.

Hypercortisolism

Etiology and pathogenesis. Hyperfunction of the adrenal cortex results in conditions characterized by hypercortisolism. Primary adrenocortical hyperfunction is caused by disease of the adrenal cortex (adrenal adenoma). Secondary disease is caused by hyperfunction of the anterior pituitary ACTH-secreting cells, and tertiary disease is caused by hypothalamic dysfunction or injury. The term Cushing *syndrome* is used to describe the clinical features of hypercortisolism, regardless of cause. Cushing *disease* is the diagnosis reserved for pituitary-dependent conditions.

In pediatric and adult populations, hypercortisolism is frequently caused by the excessive production of pituitary ACTH by microadenomas or adenomas. Ectopic ACTH production by nonpituitary tumors can also stimulate the adrenal glands. In the United States exogenous steroids used in the management of various diseases, such as allergic and autoimmune diseases, are a common cause of Cushing syndrome.

Clinical manifestations. An individual with excess circulating glucocorticoids typically develops a round face with prominent, flushed cheeks, often referred to as "*moon face*" (see Box 40.3). There is a noticeable weight gain with increasing total body fat, especially in the abdomen. A dorsocervical fat pad, capillary friability, and thinning of the skin with the formation of purple striae and ecchymoses over the abdomen, arms, and thighs develop. Muscle mass decreases, and muscle weakness develops. Cortisol increases tissue resistance to the effects of insulin and may contribute to glucose intolerance or hyperglycemia. Fig. 40.10 shows the common clinical manifestations of Cushing syndrome.

Hypertension may develop as a consequence of the salt-retaining activity of cortisol and of the increased blood volume. An associated reduction of serum potassium level is commonly found because of the excessive excretion of potassium by the kidneys. With chronic Cushing syndrome, demineralization of the bones (osteoporosis) and resulting fractures may occur. The cortisol excess may be accompanied by increased androgen production (excessive hair production, acne, menstrual irregularities).

Emotional changes include depression, emotional lability, anxiety, and irritability. Rarely, euphoria or psychosis may develop at higher concentrations of cortisol. Decreases in short-term memory, concentration, and attention span may be present. Appetite is usually increased.

The diagnosis of adrenocortical excess depends on reliable, accurate laboratory measurements. A serum ACTH level may be helpful in differentiating between primary (low ACTH level) and secondary (high ACTH level) hypercortisolism. Urinary free cortisol levels will be elevated

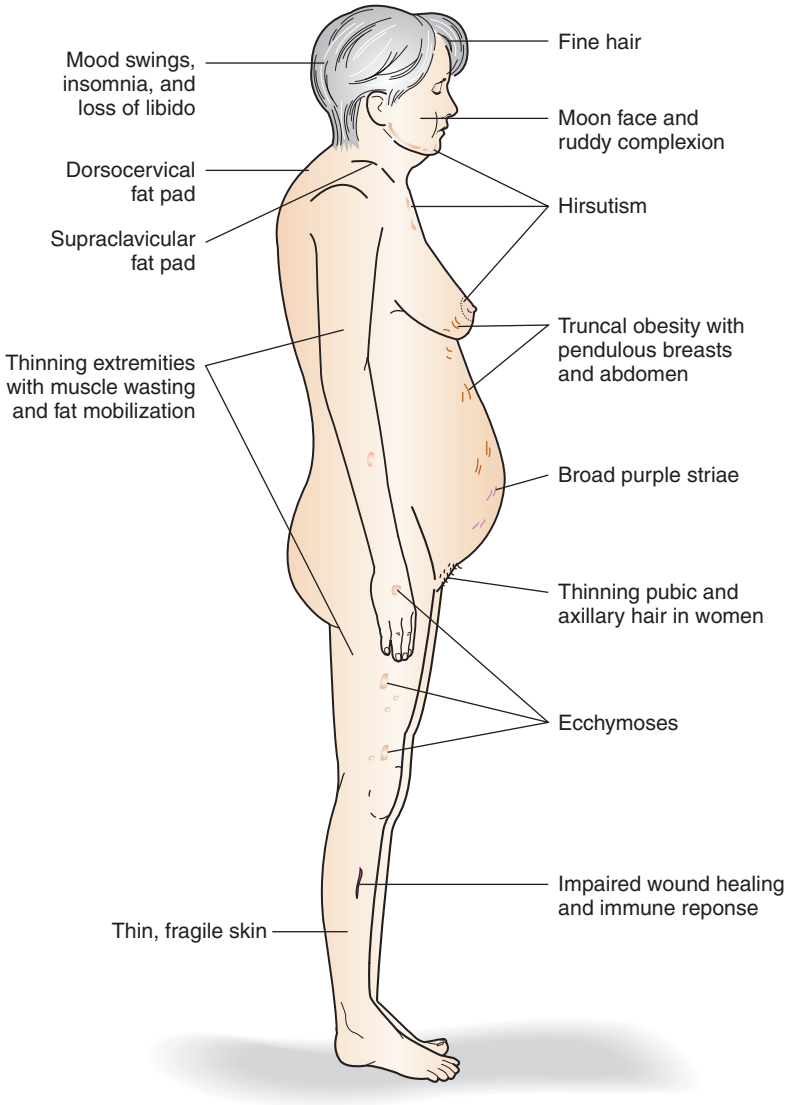


FIG 40.10 Common clinical manifestations of Cushing syndrome.

in all forms. A 24-hour urine collection is typically done to exclude inappropriate diagnoses attributable to diurnal variations in cortisol production. If the ACTH level is elevated or normal and the 24-hour urinary free cortisol level is found to be elevated, a dexamethasone suppression test may be used to differentiate between pituitary causes and ectopic causes of excessive ACTH production. Most pituitary adenomas (Cushing disease) demonstrate a relative resistance to feedback inhibition by cortisol. When a more potent glucocorticoid is given (dexamethasone), the pituitary gland responds by reducing ACTH production. Sometimes Cushing syndrome is a result of ectopic production of ACTH by cancer cells. In this case the suppression test will fail to reduce ACTH production because the cancer cells do not respond to feedback control. A negative suppression test will necessitate a diagnostic evaluation to find the ectopic cancer source.

Treatment. The choice of treatment for Cushing syndrome is based on its etiology. Patients who have Cushing syndrome as a result of exogenous glucocorticoid therapy should have doses reduced if possible. Care must be taken to taper doses slowly to avoid acute adrenal insufficiency. For pituitary disease (Cushing disease), transsphenoidal hypophysectomy or laser ablation of the anterior pituitary gland may be done.

Unilateral adrenalectomy is used if the cause of the hypercortisolism is an adrenal tumor. Bilateral removal of the adrenal glands is rarely necessary because most tumors are unilateral. Radiation therapy may be an option if surgery is contraindicated. Fig. 40.11 shows a woman before and after treatment for Cushing syndrome.

Pharmacologic agents that block cortisol production can also be utilized either alone or in conjunction with surgery and radiation. When any of these therapies is used, the patient should be assessed for treatment-induced adrenal insufficiency.

Hyperaldosteronism

Excessive aldosterone production may be from primary hyperaldosteronism (Conn syndrome) or secondary to conditions associated with poor kidney perfusion. Conn syndrome usually occurs as a result of aldosterone-secreting tumors. Aldosterone is not under pituitary control, and in this case "secondary" refers to disease processes or conditions that stimulate the renin–angiotensin–aldosterone cascade (see Chapter 39). These include heart failure, reduced kidney perfusion, and liver cirrhosis.

Aldosterone facilitates salt and water retention by the kidney. Because aldosterone acts on the distal renal tubule to promote sodium exchange

FIG 40.11 A woman with Cushing syndrome before (**A** and **C**) and after (**B**) removal of an adrenal adenoma. (From Wyngaarden JB et al: *Cecil textbook of medicine,* ed 19, Philadelphia, 1992, Saunders, p 1285.)

for the potassium lost in the urine, individuals with hyperaldosteronism may have decreased potassium levels. The drug spironolactone is an aldosterone antagonist and therefore is useful in the medical management of aldosterone excess. Spironolactone increases sodium excretion and potassium retention. Sodium restriction and potassium replacement may also be necessary.

KEY POINTS

- The adrenal cortex produces three classes of steroid hormones: (1) glucocorticoids, (2) mineralocorticoids, and (3) androgen steroids. Glucocorticoid synthesis is regulated by the pituitary secretion of adrenocorticotropic hormone (ACTH), which is controlled by hypothalamic corticotropin-releasing hormone (CRH). Mineralocorticoid synthesis is regulated by the renin–angiotensin system. The glucocorticoid cortisol provides the primary negative feedback mechanism to inhibit CRH and ACTH release.
- Adrenocortical insufficiency may be primary (Addison disease), in which case it is characterized by high ACTH levels in the blood and hyperpigmentation of skin related to excessive pituitary secretion, or it may be secondary, in which case it is characterized by low ACTH levels.
- Primary adrenal insufficiency may follow autoimmune destruction, surgical removal, or trauma of the gland. Exogenous administration of steroids suppresses ACTH, resulting in adrenocortical atrophy. Sudden withdrawal of exogenous steroids may result in adrenal insufficiency.
- Inherited defects in biosynthetic enzymes necessary for cortisol production may affect one or more of the steroid hormone synthesis pathways. Cortisol deficiency results in pituitary release of ACTH, stimulating the adrenal gland to enlarge (congenital adrenal hyperplasia). Excess androgens may be synthesized, leading to masculinization of females and precocious puberty in males.
- Manifestations of primary adrenocortical insufficiency include weight loss, salt wasting, volume depletion, low blood pressure, hypoglycemia, and hyperkalemia. Stress may lead to severe symptoms (addisonian crisis), including circulatory collapse (hypotension). Treatment includes hormone replacement therapy. Dosages are generally increased during periods of stress (e.g., surgery).
- Excess cortisol production attributable to pituitary hyperstimulation of the adrenal cortex is termed *Cushing disease.* Hypercortisolism of any other cause is termed *Cushing syndrome.* ACTH excess may be due to pituitary adenoma or exogenous production by nonpituitary tumors. Cushing syndrome is commonly due to administration of exogenous steroids.
- Clinical manifestations of Cushing disease and Cushing syndrome include moon face, cervical fat pad, central obesity, thin extremities, weight gain, thin skin, striae, hypertension, and hyperglycemia. Plasma cortisol levels and the urinary excretion of cortisol metabolites are increased. Surgical removal of ACTH-producing tumors or removal of the adrenal gland is the usual treatment.
- Primary hyperaldosteronism (Conn syndrome) is usually due to adrenal tumor. Aldosterone enhances sodium and water reabsorption and potassium excretion from the kidney, leading to hypervolemia, hypertension, and hypokalemia.

ADRENAL MEDULLA DISORDER

Pheochromocytoma

Etiology and pathogenesis. The adrenal medulla secretes two important catecholamine hormones in response to stimulation by the sympathetic nervous system. Epinephrine, or adrenaline, accounts for about 80% of the adrenal medulla's secretion; norepinephrine accounts for the other 20%. Norepinephrine is also the neurotransmitter produced by the postganglionic sympathetic fibers. Sympathetic effectors such as the heart, smooth muscle, and glands have adrenergic receptors for norepinephrine and epinephrine. Both epinephrine and norepinephrine

can bind to adrenergic receptors to prolong and enhance the effects of sympathetic stimulation.

Pheochromocytoma is a tumor of chromaffin tissue that results in the excessive production and release of catecholamines. It is usually benign, but in about 10% of cases the tumor exhibits malignant behavior. It is usually found in the adrenal medulla, but it may also arise in other sites where there is chromaffin tissue, such as the sympathetic ganglia. Like adrenal medullary cells, the tumor cells of a pheochromocytoma produce and secrete the catecholamines epinephrine and norepinephrine in response to sympathetic stimulation. Intermittent excessive release of these catecholamines results in periods of hypertension. The hypertension in individuals with pheochromocytoma is influenced by the level of sympathetic nervous system stimulation, the circulating catecholamine levels, and the cardiovascular response to these changes.

Clinical manifestations. The most common problem experienced by individuals with pheochromocytomas is intermittent or persistent hypertension. Bouts of the classic triad of headache, tachycardia, and diaphoresis strongly suggest the diagnosis of pheochromocytoma. Sporadic hypertensive episodes may occur with stress, excitement, physical activity, ingestion of certain drugs, and the smoking of tobacco products. Other symptoms may include tremor, nervousness, emotional lability, pallor, fatigue, generalized GI complaints, and orthostatic hypotension. Signs of a hypermetabolic state may be present, such as fever and weight loss. CT and MRI are commonly employed diagnostic tools to locate tumors on the adrenal glands.

Treatment. Uncontrolled hypertension can lead to end organ damage and stroke, so prompt diagnosis and aggressive therapy are necessary. The usual treatment for this condition is surgical removal of the tumor. Before surgery, sympathetic blocking medications may be prescribed to manage blood pressure and relieve symptoms.

If surgery is contraindicated, treatment with drugs to block catecholamine production is possible. However, surgery is the only curative therapy. After a period of stabilization, the sympathetic nervous system is able to compensate for the loss of adrenal medullary function and hormone replacement is no longer necessary. Because there is an increased risk of recurrence of the tumor in later years, annual follow-up is recommended.

KEY POINTS

- The adrenal medulla releases catecholamines into the bloodstream when stimulated by the sympathetic nervous system. Catecholamines increase heart rate, blood pressure, and glucose release from the liver.
- A pheochromocytoma is a catecholamine-secreting tumor that is usually located in the adrenal medulla. Excessive catecholamine release from the tumor causes intermittent or persistent hypertension, headache, tachycardia, tremor, and irritability. Most tumors are benign, and surgical removal relieves the disorder. Adrenergic blocking agents may be used to manage the hypertension until surgical treatment is accomplished.

PARATHYROID GLAND DISORDERS

Regulation and Actions of Parathyroid Hormone

The parathyroid glands are small glands located at the upper and lower poles of the thyroid. There are usually four parathyroid glands, although there are reports of fewer or more than four being found during surgery. The parathyroid glands detect serum calcium concentration and help maintain constant levels through the regulation of calcium absorption and resorption from bone (Fig. 40.12). The absorption of calcium from the intestine and renal tubules is dependent on vitamin D and is impaired in conditions such as renal failure, in which active vitamin D is deficient (see Chapter 28).

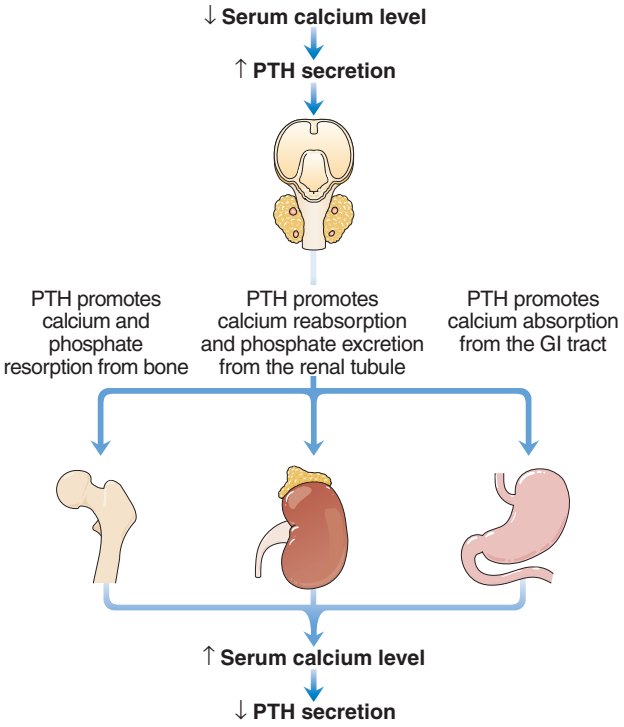


FIG 40.12 Parathyroid hormone (PTH) increases serum calcium level through its effects on bone, renal tubules, and intestine. *GI,* Gastrointestinal.

Serum calcium levels provide the feedback necessary to regulate parathyroid hormone (PTH) secretion. A decrease in serum calcium level causes a release of PTH. An elevated serum calcium level leads to suppression of PTH secretion. Calcium sensing is accomplished by a cell surface G-protein coupled receptor called *CaSR*. Calcium ions bind to CaSR on the cell membrane and signal through Gq pathways to regulate PTH gene expression. PTH is not under control of the hypothalamic-pituitary system. PTH acts on bones, intestine, and renal tubules to increase serum calcium levels. In the bone, it increases osteoclastic activity, resulting in the release of calcium (and phosphate) from bone into extracellular fluid. Renal calcium reabsorption increases under the effect of PTH, thus decreasing urinary calcium excretion.

Calcitonin, produced by thyroid parafollicular cells (C cells), also influences the processing of calcium by bone cells. Calcitonin increases bone formation by osteoblasts and inhibits bone breakdown by osteoclasts. Although the role of calcitonin in calcium homeostasis is not entirely clear, calcitonin tends to decrease blood calcium levels and promote conservation of hard bone matrix.

The clinical manifestations of PTH dysfunction are those of hypercalcemia and hypocalcemia and related changes in neuromuscular excitability. A discussion of serum calcium regulation and disorders can be found in Chapter 24.

Hyperparathyroidism

Etiology and pathogenesis. The causes of primary hyperparathyroidism often remain unclear. Despite an elevated serum calcium level, PTH continues to be secreted. Some forms of hyperparathyroidism have a genetic origin involving mutations of the CaSR sensor. Hyperparathyroidism from a single parathyroid adenoma occurs in the majority of cases. Hyperplasia of the parathyroid is found in the remainder of the cases.

In hyperparathyroidism, bone resorption and formation rates are increased. Serum calcium levels do not rise uncontrollably; indeed, excessive parathyroid gland secretion rarely causes hypercalcemic crisis. Malignant tumors elsewhere in the body can also release PTH-like hormones and are a more frequent cause of extreme hypercalcemia and hypercalcemic crisis.

A hyperparathyroid state during pregnancy leads to perinatal and neonatal complications. The newborn's PTH production will be suppressed by maternal hypercalcemia, and neonatal hypocalcemia and tetany can develop. This presentation in a newborn may be the first indication of the need for investigative studies in the mother if the disorder was asymptomatic during pregnancy. In chronic renal failure, hyperparathyroidism may result from reduced production of active vitamin D (which impairs calcium absorption) and from impaired glomerular filtration (which limits excretion of phosphate in the urine). Some drugs, such as lithium and thiazides, may increase serum calcium levels, leading to a misdiagnosis of hyperparathyroidism.

Clinical manifestations. The presentation of hyperparathyroidism is related to the level of hypercalcemia and the hyperparathyroid state. Hyperparathyroidism may present as asymptomatic hypercalcemia. Individuals are prone to kidney stones and to bone demineralization (osteoporosis). Severe hypercalcemia causes a wide variety of effects, including polyuria and dehydration. Anorexia, nausea, vomiting, and constipation may develop. Various cardiac problems can arise, including bradycardia, heart block, and cardiac arrest.

Often, asymptomatic cases of hyperparathyroidism are found on screening serum chemistry laboratory reports that note mild elevations in serum calcium levels. In primary hyperparathyroidism (not secondary to renal disease), serum calcium levels are elevated and serum phosphorous levels are low to low-normal. Urinary excretion of calcium and phosphate is elevated, as are serum PTH levels.

Treatment. Surgical removal of the abnormal parathyroid gland(s) is the treatment of choice. Individuals with asymptomatic hyperparathyroidism may defer surgery. In such cases, medical management may work for a time. Medical management includes hydration (to prevent kidney stone formation) and ambulation to maintain bone density.

For hypercalcemic crisis, rapid volume expansion with normal saline reverses dehydration. Volume replacement also results in improved glomerular filtration rate and increased calcium excretion. Diuretics, other than thiazide diuretics, may be used to increase calcium excretion by the kidney.

Hypoparathyroidism

Etiology and pathogenesis. Hypoparathyroidism most frequently occurs as a consequence of parathyroid or thyroid surgery or radiation in the area. Transient or permanent hypoparathyroidism may develop after thyroidectomy because of damage to parathyroid gland blood supply, postsurgical swelling, or fibrosis.

Hypoparathyroidism can occur after the removal of one hyperfunctioning parathyroid gland. The hyperfunctioning gland had been suppressing the function of the other parathyroid glands, and when removed a temporary state of deficiency may follow until the remaining parathyroid glands resume function. Congenital lack of parathyroid tissue and idiopathic hypoparathyroidism are causes of hypoparathyroidism in infants and children. Autoimmune processes may also target and damage the parathyroid glands.

Clinical manifestations. Clinical manifestations of hypoparathyroidism occur as a result of low serum calcium levels. The manifestations of acute hypocalcemia include circumoral numbness, paresthesias of the distal extremities, muscle cramps, fatigue, neuromuscular irritability, anxiety, nonspecific electroencephalographic changes, and prolongation of Q-T intervals on the electrocardiogram. Severe manifestations of hypocalcemia include carpopedal spasm, laryngospasm, and seizures. Neuromuscular irritability can be elicited by a positive Chvostek sign (ipsilateral contraction of the facial muscles that occurs from tapping the facial nerve anterior to the ear) or Trousseau sign (carpal spasm produced by pressure ischemia of the nerves in the upper arm during inflation of a blood pressure cuff for 3 to 5 minutes above the systolic blood pressure).

The serum calcium level is low, and the phosphorous level is elevated. Antibodies to the parathyroid gland are present if an autoimmune mechanism is operant.

Treatment. Emergency treatment with IV calcium is needed if an individual presents in acute hypocalcemic crisis (tetany, laryngospasm, convulsions). Calcitriol, an activated form of vitamin D, may be useful. Long-term treatment includes administration of an oral calcium supplement with vitamin D.

KEY POINTS

- Parathyroid hormone (PTH) is an important regulator of serum calcium levels. Low serum levels of ionized calcium are a potent stimulus for PTH release. PTH increases calcium absorption from the GI tract in concert with vitamin D, resorption of calcium and phosphate from bones, and reabsorption of calcium from the urine filtrate. PTH also increases the excretion of phosphate by the kidney. Disorders of PTH secretion are manifested as alterations in serum Ca^{2+} levels.
- Hyperparathyroidism may be idiopathic or may be due to a parathyroid adenoma. Its manifestations result from high serum calcium levels and bone demineralization. High serum calcium levels decrease neuromuscular excitability. Treatment entails removing the abnormal glands. Adequate hydration may help prevent the formation of kidney stones.
- Hypoparathyroidism may be idiopathic, autoimmune, or secondary to surgical removal of the parathyroid gland. The manifestations result from low serum calcium levels, which increase neuromuscular excitability. Paresthesias, cramps, spasms, tetany, and seizures may result. Elicitation of Chvostek and Trousseau signs indicates neuromuscular hyperexcitability. Treatment entails calcium (and vitamin D) supplementation rather than PTH replacement.

ANTIDIURETIC HORMONE DISORDERS

ADH (vasopressin) is secreted from the posterior pituitary gland in response to changes in blood osmolality. The details of ADH synthesis, regulation, and activity are described in Chapter 39. Concepts related to physiologic effects on renal water handling can be found in Chapter 26; Chapter 24 includes details of body fluid regulation and osmolality.

Diabetes Insipidus

Etiology and pathogenesis. Diabetes insipidus (DI) is a disorder of insufficient ADH activity characterized by excessive loss of water in the urine. ADH acts directly on the renal collecting ducts and distal tubules, increasing membrane permeability to and reabsorption of water. Damage to the ADH-producing cells in the hypothalamus can occur with closed head trauma, intracranial tumors, and neurosurgery. Some pharmacologic agents can lead to abnormalities in ADH secretion (Box 40.4). For example, the diuresis that follows alcohol ingestion occurs because of decreased ADH secretion.

DI is a term meaning *a large diuresis of inappropriately dilute urine.* In adults with DI, 30% of cases are idiopathic, 20% are caused by the surgical treatment of brain tumors, 16% result from nonsurgical brain trauma, 25% are secondary to brain tumors, and 9% follow a

BOX 40.4 Agents That Cause Alterations in Antidiuretic Hormone Secretion

Agents That Enhance Release	Agents That Suppress Release
β-Adrenergic agents	α-Adrenergic agents
Barbiturates	Alcohol
Carbamazepine	Phenytoin
Clofibrate	
Cyclophosphamide	
Histamine	
CO_2	
Morphine and narcotic analogs	
Nicotine	
Prostaglandin E_2	
Vincristine	

hypophysectomy. ADH deficiency may be accompanied by other hypothalamic-pituitary hormone deficiencies. Damage to the posterior pituitary gland may cause temporary or permanent deficiency of ADH. With insufficient amounts of ADH, urine cannot be concentrated and free water is lost, causing hyperosmolality and hypernatremia. This is called *central DI* because the ability to produce and release ADH from the pituitary gland is lost. *Nephrogenic DI* occurs when the kidney is unable to respond to ADH because of chronic renal disease, receptor defects, serum electrolyte abnormalities, or drugs (e.g., lithium). The clinical presentation of both forms of DI is similar, and measurement of serum ADH level may be helpful in determining the etiology.

Clinical manifestations and diagnosis. The development of polyuria (excessive urination) and polydipsia (excessive drinking) is the hallmark of DI. The patient may void as much as 15 L of urine daily. The specific gravity of the urine will be greatly decreased. If the thirst center of the hypothalamus is functional, the patient will consume up to 15 L of water to maintain osmolar balance. Symptoms persist at night (nocturia), interrupting normal sleep patterns. If the thirst center has been damaged, DI becomes a life-threatening illness because increased water losses from the kidneys (resulting from the absence of ADH) are not counteracted by increased thirst and fluid intake.

DI results in hypernatremia (water deficit) from loss of water without concurrent loss of sodium. Hypernatremia is associated with serum sodium concentrations in excess of 145 mEq/L and indicates a body water deficit relative to sodium. Signs and symptoms include thirst, dry mucous membranes, poor skin turgor, decreased saliva and sweat production, disorientation, lethargy, and seizures. The early neurologic symptoms are thought to be due to shrinkage and dehydration of neuronal cells, which are more sensitive to osmolality changes than other cell types.

Most sudden, critical presentations are straightforward, with documented hypotonic polyuria, hypernatremia, and hypertonicity indicating a defect in secretion of ADH. Individuals presenting with the sudden onset of polyuria and polydipsia should undergo laboratory studies, including tests for glucose, urine and serum electrolytes, serum creatinine, and blood urea nitrogen (BUN) levels. The results of these tests should exclude diabetes mellitus and kidney disease as the basis for the presenting complaints. A comparison of serum and urine osmolality is needed, as is a urine specific gravity measurement. Dilute urine in the presence of water deficit and hypernatremia along with abnormally low serum ADH levels are diagnostic of central DI.

A water deprivation test may be used to confirm the diagnosis. Water intake is restricted, and the urine osmolality is measured hourly. When a plateau in urine osmolality is reached, vasopressin is administered. With central DI, urine concentration increases after vasopressin administration. Polyuria and polydipsia also resolve. In the case of nephrogenic DI, little or no response to vasopressin occurs.

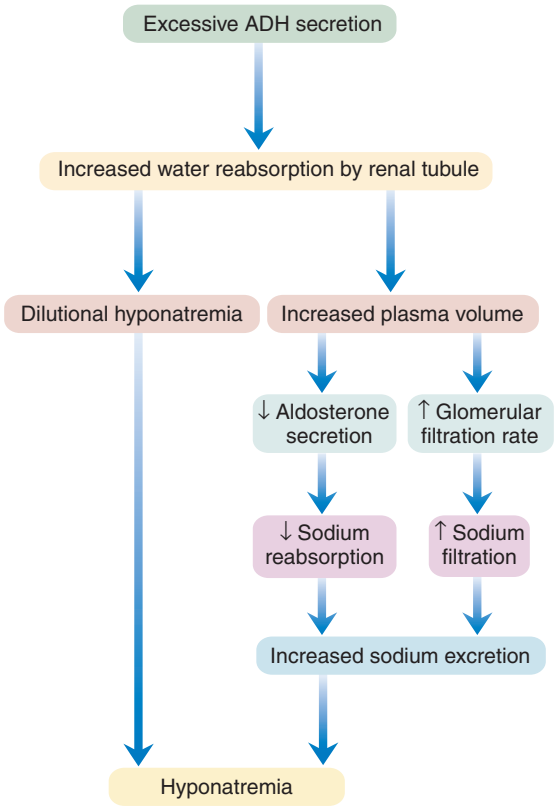


FIG 40.13 Syndrome of inappropriate antidiuretic hormone (SIADH) secretion leads to hyponatremia by two mechanisms: (1) dilution of plasma and (2) increased excretion of sodium by the kidneys. Sodium excretion is increased because of the expanded plasma volume, which enhances sodium filtration and reduces sodium reabsorption. *ADH,* Antidiuretic hormone.

Treatment. Daily replacement of ADH is needed for the management of DI. DDAVP (1-deamino-8-d-arginine vasopressin), a synthetic analog of ADH, can be given to replace vasopressin deficiency. Free access to fluids is necessary, and home testing of urine specific gravity may be useful for some individuals to allow them to adjust their dose independently.

Syndrome of Inappropriate Antidiuretic Hormone Secretion

Etiology and pathogenesis. Inappropriate secretion of excessive amounts of ADH is referred to as *syndrome of inappropriate antidiuretic hormone* (SIADH). The excessive ADH is from ectopic production and has been noted in association with several types of tumors, the most common of which are primary lung malignancies. Nonmalignant lung disorders are also capable of ADH synthesis, or stimulation of central ADH production, especially pulmonary tuberculosis. Drug-induced ADH secretion occurs with the administration of a number of medications, including carbamazepine, morphine, and barbiturates (see Box 40.4).

SIADH results in hyponatremia when free water is inappropriately conserved and "dilutes" the serum to a sodium concentration below the normal range (<135 mEq/L). In hyponatremia there is an excess of water relative to solute. Cells swell, and the effects of cellular swelling on neurons can be profound. Adrenal insufficiency and hypothyroidism may

also cause increased ADH secretion and hyponatremia. Both of these hormonal deficiencies must be excluded before SIADH is diagnosed.

Clinical manifestations. Clinical manifestations are due to the hypotonicity of body fluids. SIADH is characterized by hyponatremia. Urine osmolality is inappropriately high because of increased water reabsorption in the renal tubules and collecting ducts. Serum osmolality is low because of dilution by the reabsorbed water (Fig. 40.13). The symptoms of SIADH include weakness, muscle cramps, nausea and vomiting, postural blood pressure changes, poor skin turgor, fatigue, anorexia, and lethargy. In very severe cases, confusion, hemiparesis (motor weakness on one side of the body), seizures, and coma may occur. Laboratory findings include low serum sodium, hematocrit, and BUN levels as a result of dilution of the extracellular fluid.

Treatment. Free water restriction is implemented for individuals with SIADH. Water restriction should result in a slow, steady rise in serum sodium levels and osmolality. If severe symptoms develop, IV administration of saline, combined with diuretic therapy, may cause loss of free water. Hyponatremia should be corrected slowly to avoid rapid changes in brain cell volume. If hyponatremia is persistent, drugs such as lithium may be used to block the effects of ADH.

KEY POINTS

- Antidiuretic hormone (ADH) secretion is regulated primarily by osmoreceptors in the hypothalamus that respond to changes in extracellular osmolality. An increase in serum osmolality stimulates secretion of ADH. Renal distal and collecting tubules respond to ADH by becoming more permeable to water. In the presence of ADH, water is reabsorbed from the urine filtrate, resulting in a concentrated urine.
- Central diabetes insipidus (DI) is due to lack of production of ADH by the hypothalamus or release by the posterior pituitary gland. Central DI may be idiopathic or related to brain surgery, trauma, or tumor.
- Nephrogenic DI is caused by lack of renal collecting tubule responsiveness to ADH. Nephrogenic DI may be caused by receptor abnormalities, renal disease, medications, or electrolyte imbalance.
- Most commonly, DI causes polydipsia accompanied by thirst, polyuria, increased serum sodium level, and increased osmolality. Increased osmolality may cause cellular shrinkage with neurologic signs and symptoms. The diagnosis is confirmed when dilute urine is formed during water deprivation, which is promptly corrected with administration of vasopressin.
- DI is treated with ADH hormone replacement therapy and fluid therapy.
- Syndrome of inappropriate antidiuretic hormone (SIADH) is associated with pulmonary tumors, central nervous system disease, and certain drugs. Excess ADH stimulates the renal tubules to reabsorb water despite decreased blood osmolality.
- Clinical manifestations of hyponatremia are associated with cellular swelling and neurologic dysfunction (e.g., confusion, coma). Water restriction and diuretic administration may be used to manage hyponatremia. Detection and management of the underlying cause are paramount.

SUMMARY

Endocrine disorders present as hyperfunction or hypofunction of hormone actions. An understanding of the usual actions of hormones is useful in predicting the signs and symptoms that will be apparent with excesses and deficits (see Chapter 39). Several important hormone systems are controlled by the hypothalamic-pituitary system, and disorders may occur from intrinsic defects in the target gland (primary) or abnormalities in pituitary secretion of trophic hormones (secondary). Diagnosis relies on laboratory evaluation of pituitary gland and target gland hormone levels because the signs and symptoms are similar, regardless of primary or secondary etiology. The etiologies of endocrine disorders are also similar, regardless of the particular gland involved, and include tumors and autoimmune disorders, as well as destruction, suppression, removal, or inadequate development of the gland. Treatment strategies are few and for hyperfunction include surgical removal, ablation, or drugs to block hormone synthesis. Replacement therapy is available for most endocrine deficiency disorders and is tailored to mimic normal secretion as much as possible.

RESOURCES

Agrawal P, et al: Congenital hypothyroidism. *Indian J Endocrinol Metab* 19(2):221–227, 2015.

Ball SG: Vasopressin and disorders of water balance: the physiology and pathophysiology of vasopressin. *Ann Clin Biochem* 44(Pt 5):417–431, 2007.

Chiha M, Samarasinghe S, Kabaker AS: Thyroid storm: an updated review. *J Intensive Care Med* 30(3):131–140, 2015.

Maitra A: The endocrine system. In Kumar V, et al, editors: *Robbins and Cotran pathologic basis of disease*, ed 9, Philadelphia, 2015, Saunders, pp 1073–1140.

Nayak B, Burman K: Thyrotoxicosis and thyroid storm. *Endocrinol Metab Clin North Am* 35(4):663–686, 2006.

White BA, Porterfield SP, editors: *Endocrine and reproductive physiology*, ed 4, Philadelphia, 2013, Mosby.

White B: The hypothalamus and pituitary gland. In Koeppen BM, Stanton BM, editors: *Berne and Levy physiology*, ed 6, Philadelphia, 2010, Mosby, pp 706–724.

41

Diabetes Mellitus

Benjamin J. Miller

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- Which hormones are involved in the regulation of serum glucose level, and under what physiologic conditions would each be secreted?
- What are the differentiating characteristics of type 1 and type 2 diabetes?
- How do the pathophysiologic processes differ among the various types of diabetes?
- What clinical findings are associated with hyperglycemia, and how do they differ from those of hypoglycemia?
- How is diabetes mellitus diagnosed, monitored, and managed?
- What are the acute and chronic complications of diabetes mellitus?

CHAPTER OUTLINE

Regulation of Glucose Metabolism, 816
Hormonal Regulation, 816
Neural Regulation, 817
Exercise, 818
Stress, 818
Glucose Intolerance Disorders, 820
Classification of Glucose Intolerance Disorders, 820
Prediabetes, 820
Impaired Glucose Tolerance and Impaired Fasting Glucose Tolerance, 820
Diabetes Mellitus, 821
Type 1 Diabetes Mellitus, 821
Type 2 Diabetes Mellitus, 823
Other Specific Types of Diabetes, 824
Gestational Diabetes Mellitus, 824
Screening for Diabetes, 824
Clinical Manifestations and Complications, 825
Acute Hyperglycemia, 825
Diabetic Ketoacidosis, 825
Nonketotic Hyperglycemic Hyperosmolar Syndrome, 826
Chronic Hyperglycemia, 826
Vascular Complications, 826
Macrovascular Complications, 826
Microvascular Complications, 826
Neuropathic Complications, 827
Complications in Pregnancy, 827
Treatment and Education, 827
Nutrition, 828
Obesity and Eating Disorders, 828
Exercise, 829
Pharmacologic Agents, 829
Oral Antidiabetic Agents, 829
Incretin Enhancers, Incretins, and Amylins, 830
Insulin, 830
Stress Management, 831
Assessment of Efficacy, 831
Pediatric Considerations, 833
Goals of Therapy, 834
Acute Complications, 834
Chronic Complications, 834
Treatment, 834
Geriatric Considerations, 834
Goals of Therapy, 835
Acute Complications, 835
Chronic Complications, 835
Treatment, 835

The public health impact of diabetes mellitus is enormous. According to the Centers for Disease Control (CDC) (2017) more than 30 million persons (over 9% of the population) have diabetes mellitus, although it is estimated that only 72% are aware of their diagnosis. Should this trend continue unabated, worldwide the number of persons with diabetes will rise to 439 million by 2030. The annual cost of diabetes to the U.S. medical care system was estimated to be $245 billion in 2012 (41% increase from 2007), with almost half of the total cost attributed to inpatient diabetes care. The American Diabetes Association (2016) estimates the cost may be closer to $322 billion per year. Diabetes mellitus is the seventh leading cause of death and is a major cause of disability in the United States, as well as a major risk for heart disease, end-stage renal disease, blindness, amputation, and complications of pregnancy. The disease disproportionately affects non-Caucasian and elderly individuals.

REGULATION OF GLUCOSE METABOLISM

Because diabetes mellitus affects the utilization of all energy nutrients, it is helpful to review energy nutrient metabolism to understand the disease process. The energy requirements of humans are predominantly met by glucose and fats. Produced from endogenous glycogen stores in the muscles and liver or manufactured from such substrates as amino acids and lactate, glucose is supplied to the bloodstream from the gastrointestinal tract and liver. Glucose is typically present in greater quantities in extracellular fluid than within cells.

Cells are variously permeable to glucose, and the diffusion of glucose into them is accomplished by glucose transporters (GLUT 1 to 4) specific to each tissue. GLUT 1 to 3 transporters are insulin independent; they remain in the plasma membrane whether or not insulin is present. GLUT 1 is the major glucose transporter at the blood–brain barrier, and GLUT 3 is the dominant glucose transport molecule for neurons. GLUT 2 is the primary glucose transporter in the liver and is present in small quantities in the pancreatic β cells. GLUT 1 and GLUT 3 are the predominant glucose transport molecules in the pancreatic β cells.

GLUT 4, found in muscle and adipose cells, is insulin dependent. In the absence of insulin GLUT 4 is sequestered in vesicles located within the cell. When insulin binds to insulin receptors, an intracellular signaling cascade occurs that causes the vesicles with GLUT 4 in their membranes to move (translocate) to the plasma membrane, enabling glucose entry into the muscle cell or adipocyte. When insulin no longer binds to its receptor, GLUT 4 is removed from the plasma membrane.

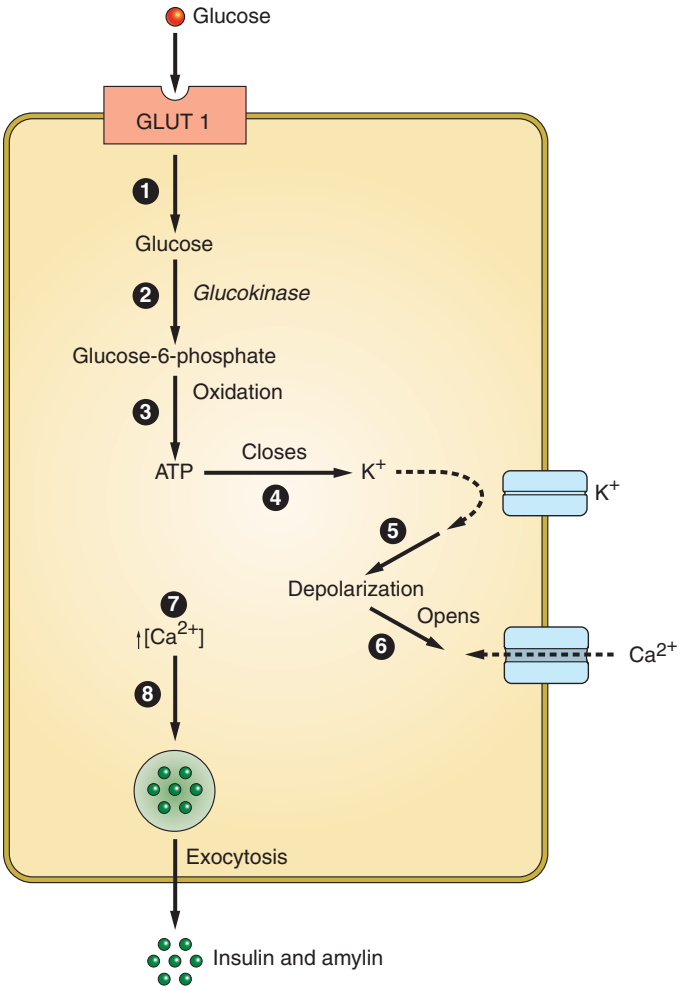


FIG 41.1 Processes of glucose-stimulated exocytosis of insulin from β cells of the pancreas. Glucose enters the β cells by facilitated diffusion through GLUT 1 in the plasma membrane. Glucose triggers production of ATP, which closes ATP-sensitive K+ channels and promotes depolarization of the cell. Depolarization triggers opening of voltage-sensitive Ca^{2+} channels, allowing calcium ions to enter the cell. Calcium ions interact with release-site proteins and trigger exocytosis of stored insulin and amylin.

Hormonal Regulation

Protein and fat metabolism is regulated by the anabolic effects of insulin. Insulin is synthesized in the pancreas by the β cells of the islets of Langerhans. The islets are groups of cells dispersed throughout the pancreas. Within the islets can be found β cells that produce insulin in the form of proinsulin, α cells that produce glucagon, δ cells that produce somatostatin, and F cells that produce pancreatic polypeptide. The primary stimulus for release of insulin from the pancreatic β cells is glucose. Glucose enters the β cells by facilitated diffusion through GLUT 1 and GLUT 3 carriers in the plasma membrane (Fig. 41.1). The concentration of glucose in the extracellular fluids determines how much enters the cell. Glucose within the β cell triggers a cascade of events that results in exocytosis of vesicles containing insulin (see Fig. 41.1). Proinsulin is produced and packaged into vesicles along with enzymes that cleave proinsulin into insulin and C-peptide (Fig. 41.2). Insulin binds to its receptor on insulin-sensitive cells and triggers glucose uptake through GLUT 4 carriers (Fig. 41.3). These carriers are sequestered within the cell when insulin levels are low and then sent to the plasma membrane to transport glucose when insulin levels are higher. Insulin mediates other effects besides glucose uptake. Insulin appears to increase the uptake and to decrease the release of amino acids by skeletal muscle, thus inducing protein synthesis and preventing muscle breakdown. The amount of stored fats in the form of triglyceride is potentiated by the action of insulin in preventing fat breakdown and inducing lipid formation. Insulin also appears to have a role in growth by stimulating the secretion of insulin-like growth factor 1 (somatomedin).

Normal glucose metabolism is usually described in reference to the fed and fasting (or absorptive and postabsorptive, respectively) states. The fed state occurs after ingestion of a meal and is characterized by utilization and storage of ingested energy nutrients. The fasting state is characterized by utilization of stored nutrients for the energy needs of the body.

In the fed state, glucose from ingested food provides the primary energy source (Fig. 41.4). The postprandial rise in blood glucose level and the presence of certain gastrointestinal hormones stimulate the production of insulin. Initial stimulation produces a brief rise in insulin secretion, termed the *first phase.* The continued presence of increased concentration of glucose produces the second phase of insulin secretion, a state characterized by insulin synthesis. Amylin is a peptide hormone produced by pancreatic β cells and cosecreted with insulin. Amylin acts on the area postrema (AP) in the brain to inhibit gastric emptying,

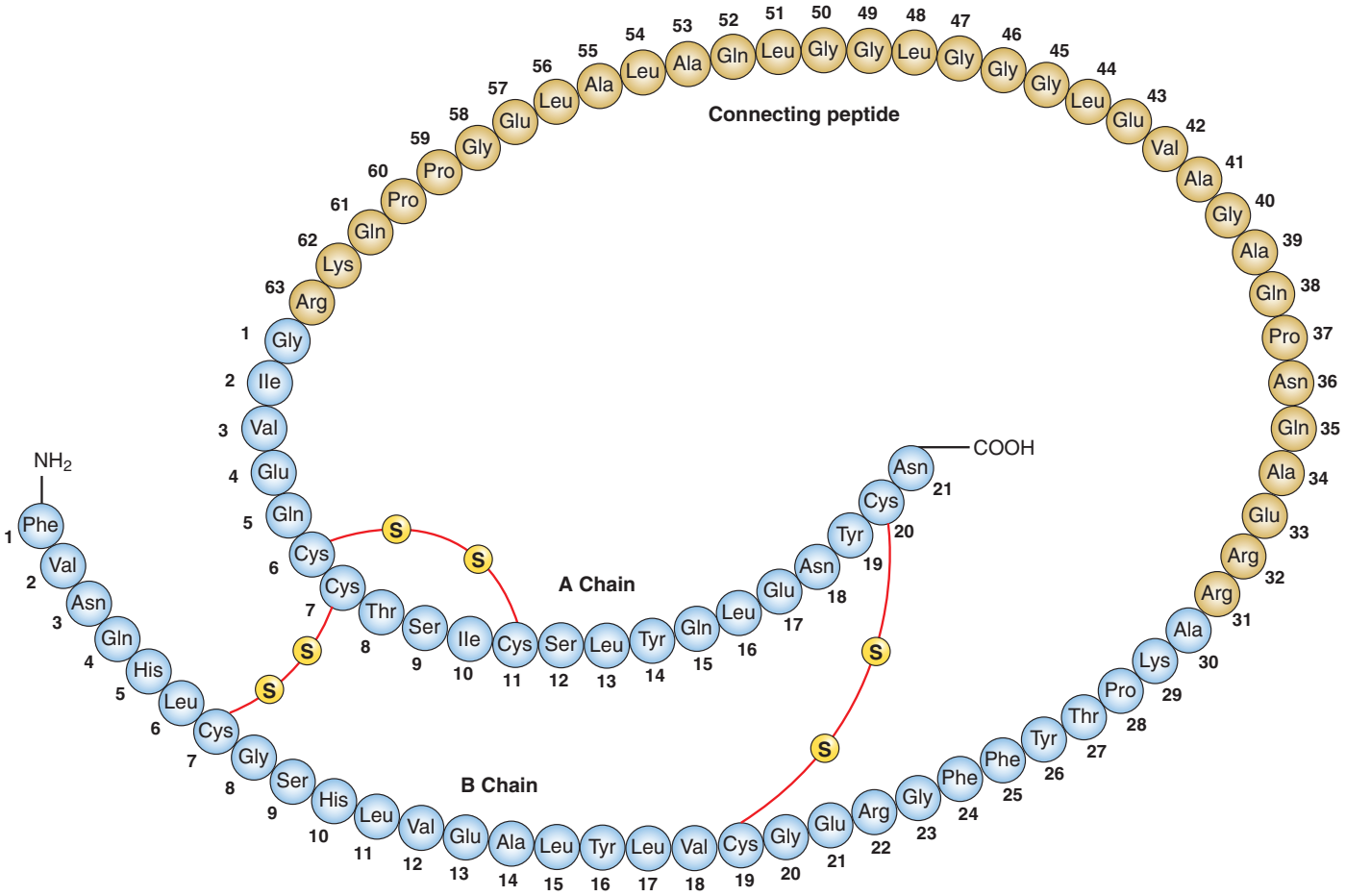


FIG 41.2 Structure of proinsulin. The blue-colored amino acids represent the insulin that is released when the C-peptide (connecting peptide) segment is cleaved.

induce satiety, and prevent postprandial spikes in blood glucose levels. Suppression of glucagon release by amylin is from a paracrine effect within the pancreatic islets and does not require any participation by the AP.

The ingestion of nutrients stimulates the release of incretin hormones, which include glucose-dependent insulinotropic polypeptide (GIP) and glucagon-like peptide 1 (GLP-1) from cells in the intestine. Both hormones stimulate production of insulin in the presence of glucose, promote proliferation of β cells, and inhibit apoptosis. The presence of these hormones and their effect on blood glucose level is known as the *incretin effect.* With parenteral nutrition, the incretin effect is not observed. In addition, GLP-1 delays gastric emptying, inhibits glucagon production, and increases satiety.

The presence of insulin stimulates the diffusion of glucose into adipose and muscle tissue and inhibits the production of glucose by the liver. After diffusion into the cell, glucose may be oxidized for the energy needs of the cell, a process termed *glycolysis.* Most ingested glucose is utilized in *glycogenesis* (production of glycogen in the muscle and the liver) (see the Metabolic Syndrome section in Chapter 42).

In the fasting state, glucose is produced by *glycogenolysis* (breakdown of stored glycogen) in the liver and muscles and by *gluconeogenesis* (production of glucose from amino acids and other substrates) in the liver (Fig. 41.5). Insulin levels, no longer stimulated by an influx of ingested glucose, fall to a basal level. The catabolic effects of the absence of insulin are evident in the stimulation of glycogenolysis and are accompanied by a rise in glucagon levels. If insulin is the hormone that dominates the fed state, glucagon dominates the fasting state. Glucagon-stimulated glycogenolysis and gluconeogenesis are responsible for up to 75% of glucose production in the fasting state. The primary source of energy to muscle tissue in the fasting state is free fatty acids produced by lipolysis (breakdown of fat from adipose tissue). Lipolysis is stimulated by the decline in plasma insulin levels.

Other hormones referred to as *counterregulatory hormones* have a role in glucose metabolism in the fasting state. Corticosteroids stimulate gluconeogenesis and counteract the hypoglycemic action of insulin. Growth hormone increases peripheral insulin resistance and prevents insulin from suppressing hepatic glucose production. Catecholamines augment glucose production by prompting hepatic glycogenolysis and gluconeogenesis.

Neural Regulation

There is a strong connection between the neural regulatory pathways and the enteric function of digestion, motility, secretion, absorption, and defense. Neural influences from the sympathetic and parasympathetic nervous system are directly involved with carbohydrate metabolism and glucose utilization. There are glucose-sensitive receptors in the brain, mouth, and pancreatic β cells and also in the hepatic portal vein. Once food is placed in the mouth, there is stimulation of the parasympathetic nervous system with stimulation of the β cells for insulin release. This is referred to as *first-phase insulin release.* Glucose-sensitive cells are located in many areas of the brain. These are activated by a decline in glucose levels (glucose-inhibited neurons [GI neurons]) or by a rise in glucose concentrations (glucose-excited neurons [GE neurons]). Under the control of the vagus pathway, the parasympathetic nervous

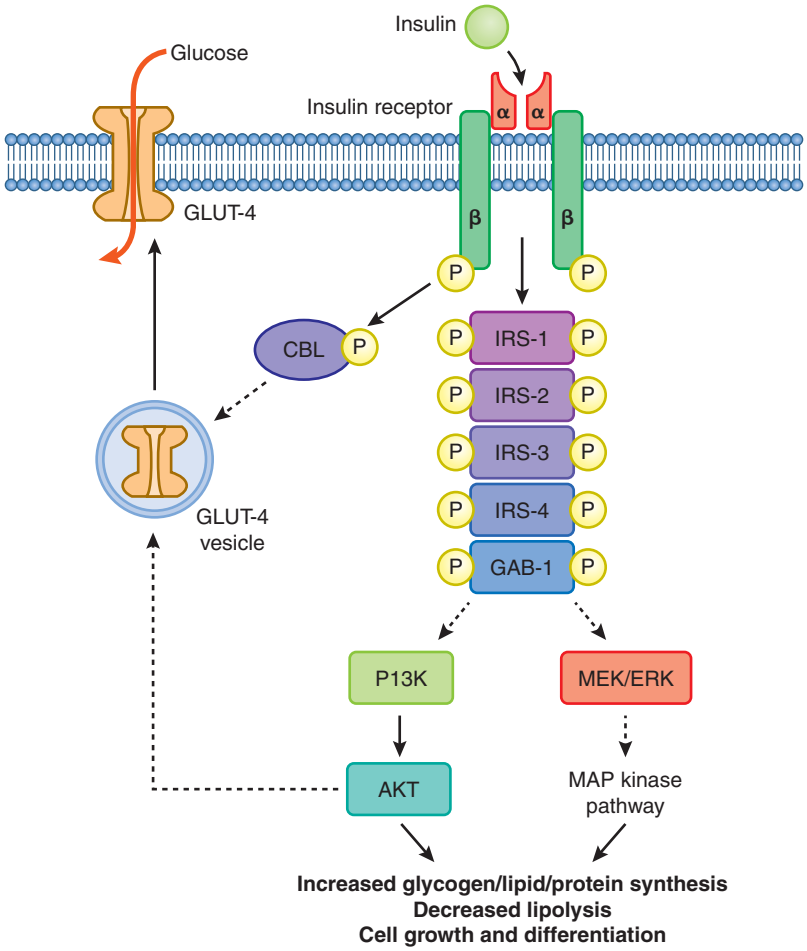


FIG 41.3 The insulin receptor is a protein kinase receptor that triggers an enzyme cascade within the cell. One effect of insulin binding to its receptor is the translocation of sequestered GLUT 4 transporters to the cell surface. The GLUT 4 carriers are passive and transport glucose down its concentration gradient. *Dotted lines* are indirect pathways of activation. *AKT,* Protein kinase B; *CBL,* Casitas B lineage proto-oncogene; *GAB,* alpha-1-B glycoprotein associated binding protein; *IRS,* insulin receptor substrates; *MAP,* mitogen activated protein; *MEK/ERK,* extracellular-regulated kinase; *PI3K,* phosphatidylinositol 3-kinase.

system not only stimulates the release of insulin but also can influence the secretion activity and β-cell mass.

Glucagon is predominantly regulated by the sympathetic nervous system in response to hypoglycemia. The hepatoportal vein contains glucose-sensitive nerve fibers; when stimulated they release norepinephrine and, along with epinephrine released from the adrenals, activate α cells to release glucagon.

Exercise

Increasing activity requires increased fuel for muscle tissue. At the onset of exercise, insulin levels drop and glucagon and catecholamine levels initially rise and increase the production of free fatty acids, the primary energy source of resting muscle (Fig. 41.6). Falling insulin levels and increased glucagon levels stimulate glycogenolysis. Under the influence of catecholamines, muscle tissue shifts from using primarily fatty acids for fuel to using stored glycogen. The relative absence of insulin and the increased production of glucagon also stimulate hepatic glycogenolysis. Glucose released by the liver increasingly meets the energy needs of muscle tissue while exercise continues. After 10 to 40 minutes of exercise, blood glucose use by muscle tissue increases 7 to 20 times. The interactions of hormones thus produce the mixture of glucose and free fatty acids used by muscle tissue during exercise.

Muscle tissue is affected not only by the influence of hormones but also by exercise itself. The resulting increase in insulin sensitivity can last as long as 16 hours. Thus in normal metabolism, increased insulin sensitivity allows normal blood glucose values in the presence of lower levels of circulating insulin.

Stress

During stress such as injury, illness, and pain, stress hormones, including corticosteroids and catecholamines, interact to ensure continuous supplies of glucose. Corticosteroids increase the production of glucose in the liver and elevate the production of glucagon. Glucocorticoids also decrease the utilization of glucose by muscle tissue by diminishing the effect of insulin on glucose transporters and by generating a decline in the number of insulin receptors and their function.

Catecholamines increase plasma glucagon levels, increase glucose production by the liver, and decrease the use of glucose by muscle and fat tissue. The production of fatty acids that is triggered by the action of catecholamines further inhibits glucose uptake in the periphery. The series of events produced by traumatic stress is referred to as *stress hyperglycemia.*

Psychological stress can produce metabolic changes comparable with those of physical stress. Deterioration in metabolic control has been

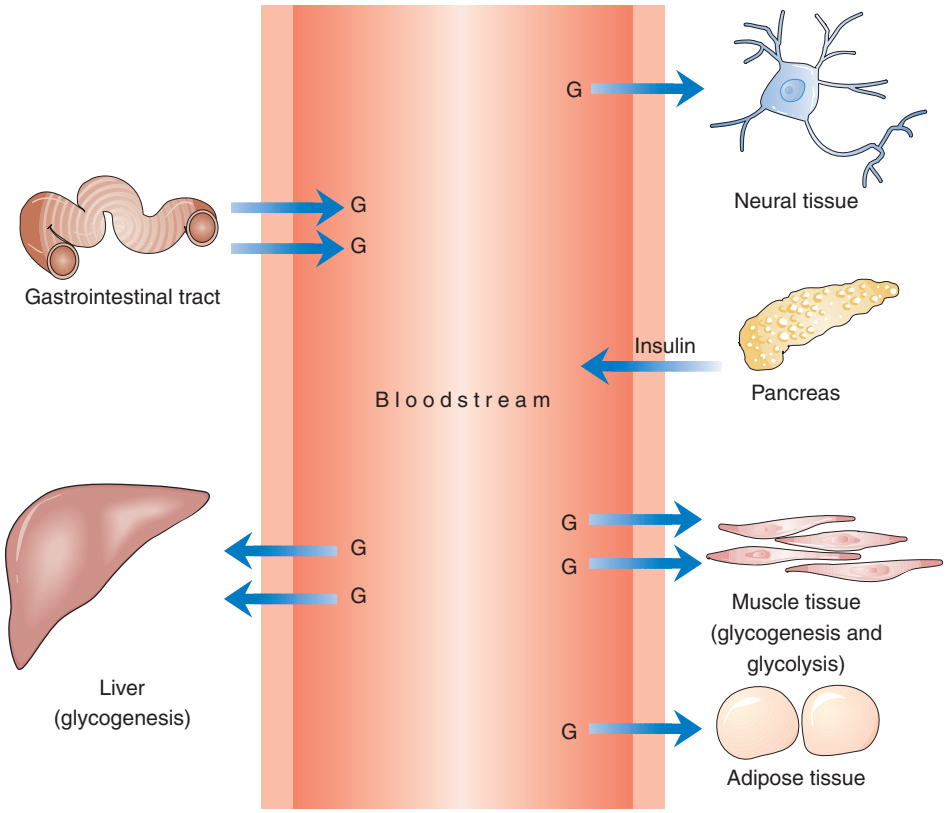


FIG 41.4 Energy metabolism in the fed state. *G,* Glucose.

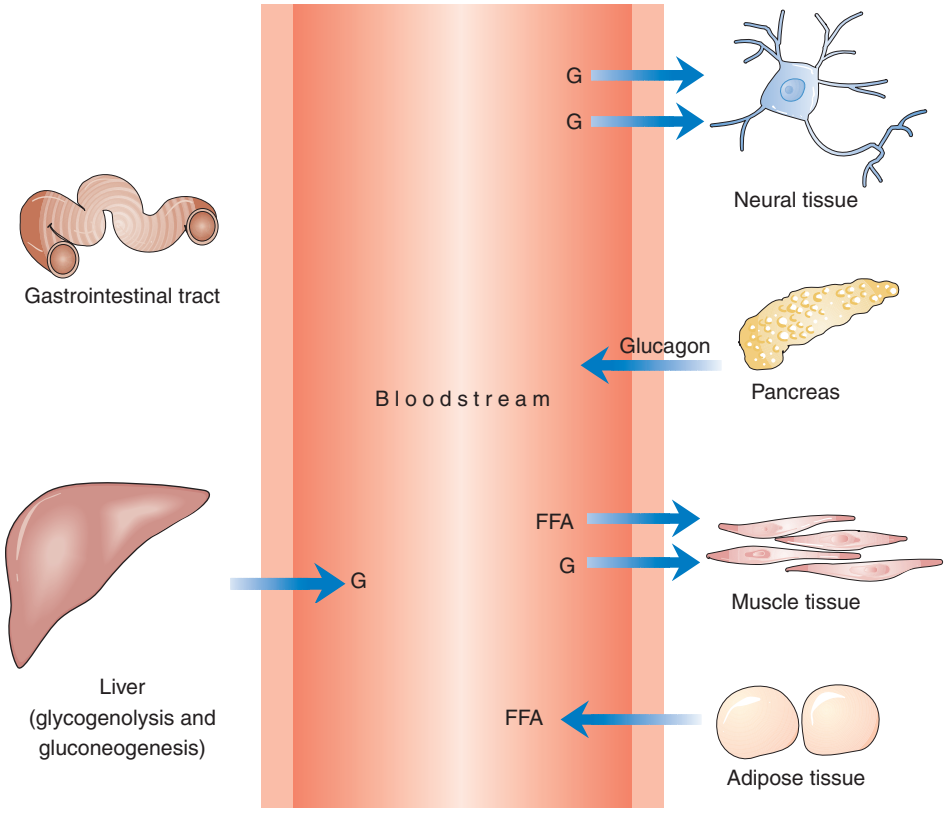


FIG 41.5 Energy metabolism in the fasting state. *FFA,* Free fatty acid(s); *G,* glucose.

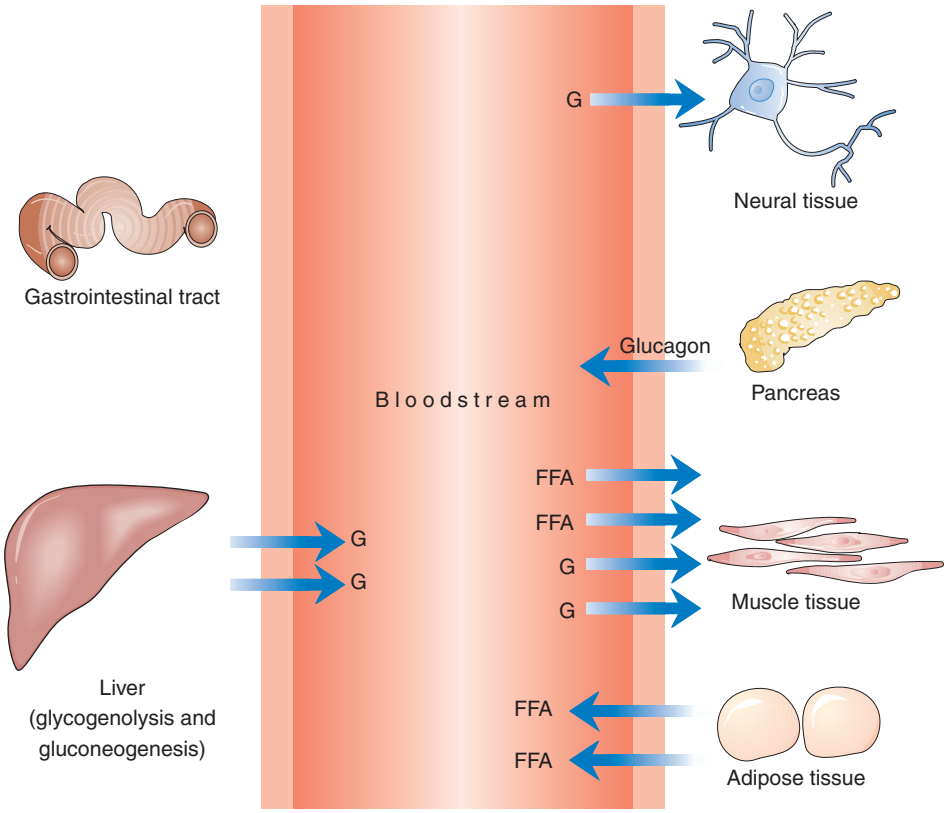


FIG 41.6 Energy metabolism during exercise. *FFA,* Free fatty acid(s); *G,* glucose.

noted in stressed diabetic subjects. However, the response is by no means universal. An increase in blood glucose levels has frequently been observed during acute stress reflecting responses to psychological stress (e.g., disordered eating).

KEY POINTS

- Plasma membrane permeability to glucose is determined by the type and density of glucose transport proteins in the membrane. In some tissues, particularly muscle and fat, the density of facilitative glucose transporters is regulated by insulin. Insulin binding to receptors on the cell surface results in translocation of glucose transporters to the cell surface. Glucose enters the cell passively by facilitated diffusion. Neurons, endothelial cells, and erythrocytes have glucose transporters that do not require insulin.
- The metabolic effects of insulin include enhancing protein synthesis and inhibiting gluconeogenesis, enhancing fat deposition and inhibiting lipolysis, and stimulating cellular growth by enhancing somatomedin secretion. Insulin is synthesized in pancreatic β cells as proinsulin. Proinsulin is stored in granules, where it is cleaved into insulin and C-peptide. A postprandial rise in the levels of glucose and other substrates stimulates the release of insulin into the bloodstream. During fasting, when blood glucose levels fall, the decrease in insulin production and the increase in glucagon secretion lead to lipolysis, glycogenolysis, and gluconeogenesis.
- Exercise has complex effects on glucose metabolism. The decrease in production of insulin and the increase in secretion of glucagon and catecholamines lead to elevated blood glucose levels. However, exercising muscle has increased insulin sensitivity, which facilitates glucose uptake.
- A number of hormones released during stress increase blood glucose levels and oppose the effects of insulin. Catecholamines, glucocorticoids, and glucagon may precipitate stress hyperglycemia.

GLUCOSE INTOLERANCE DISORDERS

Classification of Glucose Intolerance Disorders

Diabetes mellitus is not a single disease entity; as many as 30 different disorders may be called diabetes. Criteria for diagnosing the different conditions, all associated with glucose intolerance, were established by the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus in 2011.

Classifications include two prediabetes classes and four clinical classes (Box 41.1). Prediabetes classes are impaired glucose tolerance and impaired fasting glucose tolerance. The four clinical classes are type 1 diabetes mellitus, type 2 diabetes mellitus, other specific types of diabetes mellitus, and gestational diabetes mellitus (GDM).

Prediabetes

Impaired Glucose Tolerance and Impaired Fasting Glucose Tolerance

Guidelines for diagnosing the categories of impaired glucose tolerance (IGT) and impaired fasting glucose tolerance (IFG) are listed in Box 41.2. IGT and IFG are intermediate stages between normal glucose metabolism and the onset of diabetes. The pathophysiology of prediabetes is complex with a distinct relationship of elevated glucose levels and the development of insulin resistance. Glucose release is stimulated by the mass of metabolically active tissues, including fat-free mass and fat mass. The signals to stimulate gluconeogenesis are greater than the counterregulatory effects, resulting in hepatic insulin resistance.

When glucose levels remain high, the fat-free tissues such as muscle become supersaturated with glucose and start to down-regulate the glucose transporters, accelerating systemic insulin resistance.

BOX 41.1 Classifications of Glucose Metabolism Disorders

Prediabetes (Increased Risk for Diabetes Mellitus)
Impaired fasting glucose levels
Impaired glucose tolerance

Type 1 Diabetes Mellitus
Type 1A: Immune mediated
- Polygenic
- Monogenic
- Latent autoimmune diabetes in adults

Type 1B: Idiopathic

Type 2 Diabetes Mellitus

Other Specific Types of Diabetes

1. Genetic defects of β-cell function
   a. Chromosome 12, *HNF1A* (formerly *MODY3*)
   b. Chromosome 7, glucokinase (formerly *MODY2*)
   c. Chromosome 20, *HNF4A* (formerly *MODY1*)
   d. Chromosome 13, insulin promoter factor 1 (IPF-1, *MODY4*)
   e. Chromosome 17, *HNF-1β (MODY5)*
   f. Chromosome 2, *NeuroD1 (MODY6)*
   g. Mitochondrial DNA
   h. Others
2. Genetic defects in insulin action
   a. Type A insulin resistance
   b. Leprechaunism
   c. Rabson–Mendenhall syndrome
   d. Lipoatrophic diabetes
   e. Others
3. Diseases of the exocrine pancreas
   a. Trauma/pancreatectomy
   b. Neoplasia
   c. Cystic fibrosis
   d. Hemochromatosis
   e. Fibrocalculous pancreatopathy
   f. Pancreatitis
   g. Others
4. Endocrinopathies
   a. Acromegaly
   b. Cushing syndrome
   c. Glucagonoma
   d. Pheochromocytoma
   e. Hyperthyroidism
   f. Somatostatinoma
   g. Aldosteronoma
   h. Others
5. Drug or chemical induced
   a. Pyriminil (Vacor)
   b. Pentamidine
   c. Nicotinic acid
   d. Glucocorticoids
   e. Thyroid hormone
   f. Diazoxide
   g. β-Adrenergic agonists
   h. Thiazides
   i. Phenytoin (Dilantin)
   j. Interferon-α
   k. Others
6. Infections
   a. Congenital rubella
   b. Cytomegalovirus
   c. Others
7. Uncommon forms of immune-mediated diabetes
   a. Stiff-man syndrome
   b. Anti–insulin receptor antibodies
   c. Others
8. Other genetic syndromes sometimes associated with diabetes
   a. Down syndrome
   b. Klinefelter syndrome
   c. Turner syndrome
   d. Friedreich ataxia
   e. Huntington chorea
   f. Laurence–Moon–Biedl syndrome
   g. Myotonic dystrophy
   h. Porphyria
   i. Prader–Willi syndrome
   j. Others

Gestational Diabetes Mellitus

Data from American Diabetes Association: Standards of medical care for diabetes, *Diabetes Care* 2016;39(Suppl 1):S13–S23.
Dib S, Gomes M: Etiopathogenesis of type 1 diabetes mellitus: prognostic factors for the evolution of residual beta cell function, *Diabet Metab Syndrome* 2009;1(1):25.

Diabetes Mellitus

Type 1 Diabetes Mellitus

Type 1 diabetes mellitus is, by definition, characterized by destruction of the β cells of the pancreas. Type 1 diabetes can occur at any age, but peaks at the ages of 2, 4 to 6, and 10 to 14 years of age. Type 1 diabetes accounts for 10% of all diabetes and affects 1.25 million people in the United States and approximately 10 to 20 million people globally. The incidence is 1 in every 300 to 600 children and adolescents. Caucasian populations are more susceptible to type 1 diabetes mellitus than are African American, Hispanic, Asian, or Native American populations. Little gender difference is noted in the incidence of type 1 diabetes mellitus in children younger than age 15. However, more men than women are affected in the population.

Etiology. The two forms of type 1 diabetes are type 1A immune-mediated diabetes, which is the most common, and type 1B idiopathic, which is rare. Immune-mediated type 1A can be further delineated into three subcategories. Polygenic type 1 diabetes involves two or more genetic loci and accounts for 80% to 90% of type 1 cases. Monogenic causes of type 1 diabetes are rare and associated with IPEX syndrome (*i*mmune dysfunction, *p*olyendocrinopathy, *e*nteropathy, *X*-linked). The final subgroup of type 1 diabetes is latent autoimmune diabetes in adults (LADA). LADA is linked to the development of T-cell reactivity to islet antigens and autoantibodies to glutamic acid decarboxylase 65 (GADA65). LADA accounts for 2% to 12% of all cases of diabetes and is typically diagnosed after the age of 35. Often initially misdiagnosed as type 2 diabetes, LADA involves destruction of the pancreatic β cells, resulting in insulinopenia.

Type 1A diabetes is the result of an autoimmune attack on the β cells of the pancreas. A strong association with the presence of a gene or genes in the major histocompatibility complex (MHC) on chromosome 6 has been observed. Genes in the MHC are responsible for the creation of cell-surface proteins (human leukocyte antigens, or HLAs) influencing the lymphocytes to stimulate or suppress antibody production. Recent

BOX 41.2

Diagnosis of Prediabetes

Fasting plasma glucose 100 to 125 mg/dL (IFG)

2-hr plasma glucose in the 75-gm oral glucose tolerance test (OGTT) 140 to 199 mg/dL (IGT)

HbA_{1C}: 5.7%–-6.4%

Diagnosis of Diabetes

FPG ≥126 mg/dL. Fasting is defined as no caloric intake for at least 8 hours.*

OR

Symptoms of hyperglycemia and a casual plasma glucose ≥200 mg/dL. *Casual* is defined as any time of day without regard to time since last meal. The classic symptoms of hyperglycemia include polyuria, polydipsia, and unexplained weight loss.

OR

2-hour plasma glucose ≥200 mg/dL during an OGTT. The test should be performed as described by the World Health Organization, using a glucose load containing the equivalent of 75 g of anhydrous glucose dissolved in water.*

OR

HbA_{1C} greater than 6.5%. The test should be performed in a laboratory using a method that is NGSP certified and standardized to the DCCT assay.

From American Diabetes Association: Diagnosis and classification of diabetes mellitus, *Diabetes Care* 2012;35(Suppl 1):S64–S71, doi: 10.2337/dc12-s064.

DCCT, Diabetes Control and Complications Trial; *FPG,* fasting plasma glucose; *IFG,* impaired fasting glucose; *IGT,* impaired glucose tolerance; *NGSP,* National Glycohemoglobin Standardization Program; *OGTT,* oral glucose tolerance test.

*In the absence of unequivocal hyperglycemia, criteria 1 to 3 should be confirmed by repeat testing on a different day.

evidence demonstrates that two primary loci (DR and DQ) confer a genetic predisposition, whereas other loci may have a protective effect on the development of type 1 diabetes mellitus.

Viral infection or exposure to a toxic agent may be the responsible environmental influence for triggering the autoimmune process in susceptible individuals. The immune system activation is a complex process of antigen recognition (please see Chapter 9 for a detailed explanation).

The etiologic progression of type 1B diabetes mellitus is not known. Idiopathic diabetes is associated with β-cell destruction without autoimmune markers or HLA association.

Pathogenesis and clinical manifestations. Type 1A diabetes is the result of destruction of the pancreatic β cells. The process is mediated by macrophages and T lymphocytes with detectable autoantibodies to various β cells. The T lymphocytes infiltrate the islets and destroy the β cells through the secretion of cytokines (CD4 cells) and direct cytotoxic action (CD8 cells). The preclinical β-cell autoimmunity is variable and precedes the clinical diagnosis. Specific antibodies that develop against GADA65, insulinoma-antigen 2, or insulin frequently appear early in the onset of immune-mediated diabetes. Antibodies may be present for as long as 13 years before diagnosis, and the order of antibody appearance is not significant; however, the presence of multiple antibodies is highly predictive of type 1A diabetes. The presence of these antibodies results in the destruction of the pancreatic β cells. The presence of hyperglycemia indicates that autoimmune destruction of β cells has reached the point at which insulin secretion is inadequate.

Type 1A and 1B diabetes mellitus are characterized by an absolute insulin deficiency, and thus glucose cannot enter muscle and adipose tissue (Fig. 41.7). Production of glucose by the liver is no longer opposed by insulin. Overproduction of glucagon by pancreatic α cells stimulates

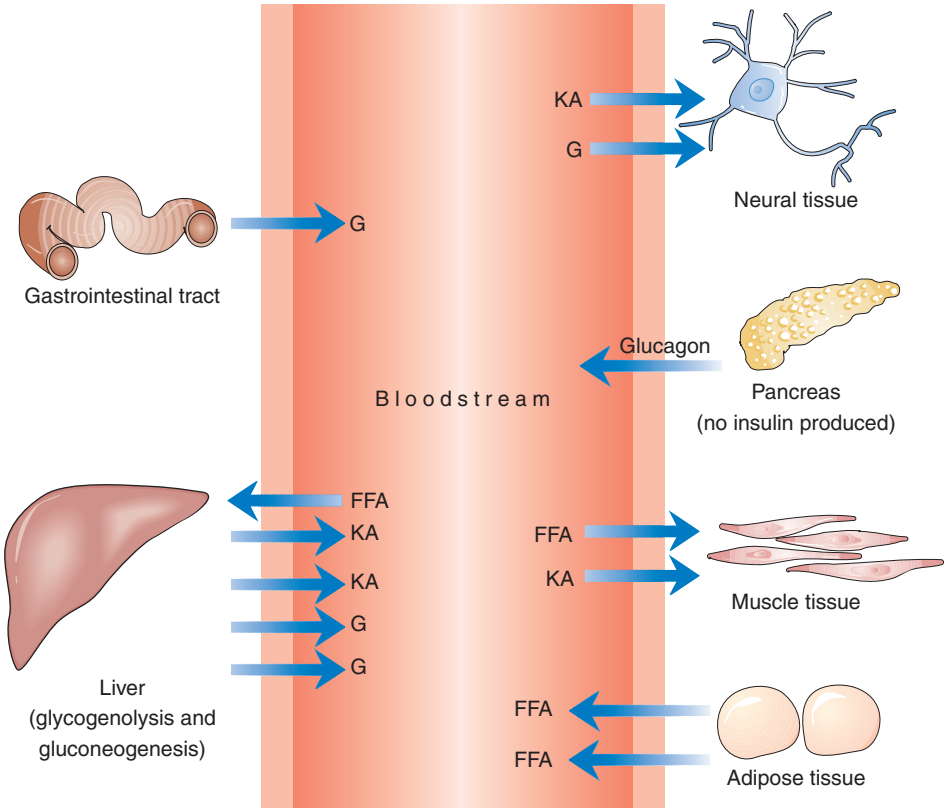


FIG 41.7 Pathophysiology of energy metabolism in type 1 diabetes mellitus. *FFA,* Free fatty acid(s); *G,* glucose; *KA,* ketoacid(s).

glycogenolysis and gluconeogenesis. Plasma blood glucose levels rise. When the maximal tubular absorptive capacity of the kidney is exceeded, glucose is lost in the urine, and the resulting glycosuria and osmotic fluid loss eventually lead to profound hypovolemia. Tissues dependent on insulin for glucose transport do not have glucose available as a substrate. Neural tissue in the brain responds to this emergency by promoting eating behavior. The increased thirst (polydipsia), increased urination (polyuria), and increased hunger (polyphagia) resulting from the aforementioned processes are the classic symptoms of diabetes.

Type 2 Diabetes Mellitus

Etiology. Individuals with type 2 diabetes mellitus are resistant to the action of insulin on peripheral tissues. Type 2 diabetes mellitus affects 90% to 95% of individuals with diabetes in the United States. Non-Caucasian and elderly populations are disproportionately affected. The prevalence of diabetes is 7.1% in the non-Hispanic white population, 12.6% in non-Hispanic blacks, and 11.8% in Hispanic/Latino Americans. The age-adjusted prevalence of diabetes in Native American populations ranges from 5.5% among Alaska Native adults to 33.5% in American Indians in southern Arizona. Overall, close to 26.9% of American individuals older than 65 years have diabetes mellitus. In people under the age of 20, diabetes affects 25.6 million people in the United States, or 11.3% of the population. The greatest at-risk population is adolescent Native Americans.

Risk factors include aging and a sedentary lifestyle, but the most powerful predictor is obesity. Excessive abdominal (visceral) fat introduces a greater threat of diabetes mellitus (and cardiovascular disease) than does lower-body obesity. The initial symptoms of polydipsia, polyuria, polyphagia, and weight loss may be subtle or absent in type 2 diabetic patients.

Epidemiologic studies indicate a strong genetic component, but no specific HLA type has been identified. When one identical twin is affected, there is close to a 100% chance the other twin will be affected.

Pathogenesis and clinical manifestations. Type 2 diabetes is characterized by a relative lack of insulin. The processes instrumental in producing the relative lack of insulin are insulin resistance and β-cell dysfunction (Fig. 41.8).

The insulin resistance of type 2 diabetes mellitus is defined as a requirement for more insulin for the same biological action, along with lowered glucose utilization at all levels of insulin concentration. A decreased number of insulin receptors and such postreceptor defects as decreased action of glucose transporters are associated with insulin resistance. Impaired glycogen synthesis may also be a factor in the development of type 2 diabetes. Impaired production of insulin by the pancreatic β cells that intensifies as the disease progresses is also present in type 2 diabetes mellitus. Individuals with this condition ultimately have an absent first-phase insulin response and a diminished second-phase response.

Basal insulin secretion may be higher than normal in type 2 diabetes. Glucagon secretion is increased absolutely or relatively (relative to insulin levels). The incretin effect, stimulation of insulin secretion by GLP-1 and GIP, is diminished in type 2 diabetes. Individuals with this condition may be predominantly insulin resistant or predominantly insulin deficient.

Type 2 diabetes mellitus is a progressive disease characterized by the development of insulin resistance, at first compensated for by increased insulin production and hyperinsulinemia. Decompensation occurs as the impaired β cells are unable to produce sufficient insulin to overcome insulin resistance. Insulin levels, however, remain elevated above normal until later in the progression of the disease. Relatively decreased insulin levels, continued insulin resistance, and hyperglucagonemia result in the hyperglycemia of diabetes. Hyperglycemia itself may then increase insulin resistance and further diminish insulin secretion. The latter process has been termed *glucose toxicity.*

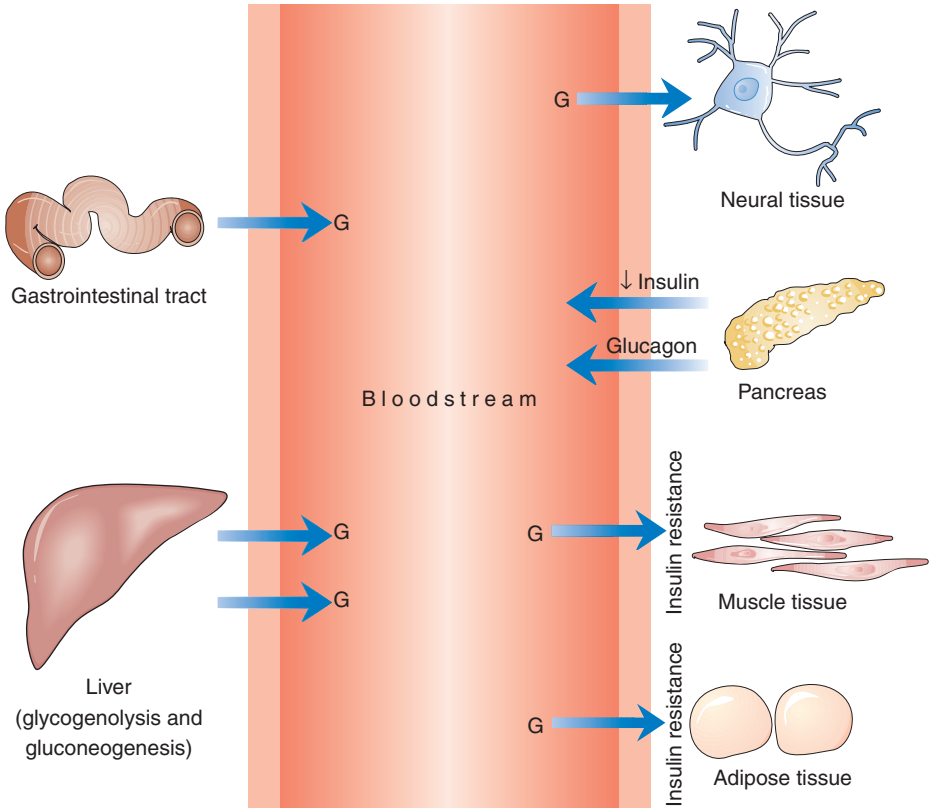


FIG 41.8 Pathophysiology of energy metabolism in type 2 diabetes mellitus. *G,* Glucose.

Other Specific Types of Diabetes

- **Genetic defects of β cells:** The genetic defects of β cells follow an autosomal-dominant pattern of inheritance and are characterized by a defect in the production of insulin. Affected individuals are identified before they reach 25 years of age, The disorder is also referred to as *mature-onset diabetes of the young* (MODY). MODY is a monogenetic disorder affecting approximately 1% of people with diabetes and tends to be seen in families of people with the condition. MODY is caused from a mutation in a single gene and passed on to offspring. Six genes have been identified that cause MODY: *HNF1A, GCK, HNF1B, HNF4A, IPF1,* and *NEUROD1*. Because of the genetic variation, some people with MODY are treated with insulin, whereas other people respond to diet, exercise, and oral agents.
- **Genetic defects in insulin action:** The disorders listed in Box 41.1 result in relatively rare, genetically determined defects of the insulin receptor.
- **Diseases of the exocrine pancreas:** Diseases of the pancreas can affect the insulin-producing capability of the organ (see Box 41.1).
- **Endocrinopathies:** Excessive production of insulin antagonists (e.g., cortisol, growth hormone, glucagon, and epinephrine) affects glucose metabolism (see Box 41.1).
- **Drug- or chemical-induced diabetes:** Many chemicals can affect the ability of the pancreas to produce insulin (see Box 41.1).
- **Infections:** Destruction of the β cells of the pancreas has been linked to various infectious agents (see Box 41.1).
- **Uncommon forms of immune-mediated diabetes:** Certain autoimmune disorders are linked to glucose intolerance (see Box 41.1).
- **Other genetic syndromes sometimes associated with diabetes:** Certain genetic syndromes are linked to glucose intolerance (see Box 41.1).

Gestational Diabetes Mellitus

GDM is, by definition, a disorder of glucose intolerance of variable severity with onset or first recognition during pregnancy. Approximately 4% (may range from 1% to 14% depending on population) of pregnancies are affected by GDM. GDM represents 90% of all pregnancies complicated by diabetes.

Etiology. In its pathophysiologic characteristics, GDM closely resembles type 2 diabetes mellitus. As in type 2 diabetes, tissue insulin resistance is present during normal pregnancy. Insulin resistance in normal pregnancy is most likely precipitated by the presence of placental hormones: human chorionic somatomammotropin, estrogen, and cortisol. The weight gain of pregnancy is also responsible for an increase in insulin resistance. During pregnancy, women require two to three times as much insulin as they do in the nonpregnant state. Women with gestational diabetes are unable to produce sufficient insulin to meet their needs during pregnancy.

Risk factors for GDM include severe obesity, history of gestational diabetes, previous offspring weighing more than 9 lb at birth, presence of glycosuria, or a strong family history of type 2 diabetes. High-risk individuals should be screened as soon as possible after confirmation of pregnancy. Because insulin needs rise sharply in the twenty-fourth to twenty-eighth weeks of pregnancy, it is recommended that all pregnant women older than 25 years be screened for gestational diabetes during the twenty-fourth to twenty-eighth weeks of gestation with either a one-step or a two-step screening strategy. Younger pregnant women who are obese, have a first-degree relative with diabetes, are members of an ethnic/racial group with a high prevalence of diabetes (e.g., African American, Hispanic, Asian, Native American), have a history of abnormal glucose tolerance, or have a history of poor obstetric outcome should also be screened (Table 41.1). Untreated gestational diabetes can result in metabolic abnormalities and stillbirth. However, the most common complications are macrosomia and neonatal hypoglycemia. Macrosomia (birth weight greater than 4000 g or >90% for gestational age) is a result of increased glucose, free fatty acids, and amino acids delivered to the fetus. Neonatal hypoglycemia is due to increased production of insulin by the fetal pancreas in response to the chronic stimulation of hyperglycemia while in utero.

TABLE 41.1 Diagnosis of GDM With a One-Step or Two-Step Test*

	One Step	TWO STEP	
	75-g Glucose Load (Fasting)	50-g Glucose Load (Nonfasting)	100-g Glucose Load (Fasting)
Fasting	92 mg/dL	NA	95 mg/dL
1 hour	180 mg/dL	≥140 mg/dL	180 mg/dL
2 hours	153 mg/dL	NA	155 mg/dL
3 hours	NA	NA	140 mg/dL

Data from American Diabetes Association: Diagnosis and classification of diabetes mellitus, *Diabetes Care* 2015;38(Suppl 1):S8–S16.
*Two or more blood glucose level determinations must be equal to or greater than these values to establish the diagnosis. The test should be done in the morning after an overnight fast of between 8 and 14 hours and after at least 3 days of unrestricted diet (150 g of carbohydrate per day) and unlimited physical activity. The subject should remain seated and should not smoke throughout the test.

Treatment. Management of GDM includes education regarding appropriate dietary choices, implementation of an exercise regimen, and observation of blood glucose and urine ketone levels. If hyperglycemia persists, insulin therapy should be initiated. Only glyburide does not cross the placenta to cause fetal hypoglycemia; this sulfonylurea may be used in GDM.

Glucose tolerance will return to normal after parturition in 97% of women with GDM. Women with this condition have a markedly increased risk for the development of type 2 diabetes mellitus or impairment in glucose tolerance later in life, especially within the first 5 years postpartum. In subsequent pregnancies, recurrence varies between 30% and 84% in women with a history of GDM.

Screening for Diabetes

Because of the high prevalence of undiagnosed type 2 diabetes mellitus, current recommendations are to screen all adults older than age 45 for diabetes at least every 3 years. There is some controversy regarding screening individuals with asymptomatic disease. Individuals who are obese with risk factors should be screened at more frequent intervals (Box 41.3).

Screening for gestational diabetes is discussed in the Gestational Diabetes Mellitus section. It is not recommended that routine screening for type 1 diabetes mellitus be performed. No agreement has been reached on the blood level of immune markers that represents a risk for type 1 diabetes or on treatment after identification of the presence of such markers.

KEY POINTS

- Diabetes mellitus is an endocrine disorder diagnosed by the presence of chronic hyperglycemia. Diabetes may be diagnosed based on the fasting (8-hour caloric fast) plasma glucose level ≥126 mg/dL (7.0 mmol/L) or the 2-hour oral glucose tolerance test ≥200 mg/dL (11.1 mmol/L) repeated on separate days. In the presence of symptoms of hyperglycemia, a single

random plasma glucose level greater than 200 mg/dL would confirm a diagnosis of diabetes. The glycosylated hemoglobin level (HbA_{1C}) may be used to confirm a diagnosis of diabetes when the HbA_{1C} ≥6.5% (48 mmol/mol). The test should be performed in a laboratory using a method that is certified and standardized.

- The classification of diabetes mellitus includes two broad categories: (1) actual glucose intolerance and (2) risk of glucose intolerance. Disorders of actual glucose intolerance include type 1, type 2, other specific types, and gestational diabetes. Prediabetes categories include individuals with impaired glucose tolerance and those with impaired fasting glucose tolerance. Individuals at risk for glucose intolerance include those with a history of glucose intolerance and those with a positive family history, obesity, or other risk factors.
- Persons with type 1 diabetes have an absolute insulin deficiency caused by pancreatic β-cell failure. Immune-mediated type 1A diabetes is associated with a specific HLA genetic makeup and is autoimmune. Idiopathic type 1B diabetes is not an autoimmune process. Type 1 diabetes may affect people of any age. Classic manifestations include polyuria, polydipsia, polyphagia, and weight loss.
- Persons with type 2 diabetes have a relative insulin deficiency caused by decreased tissue sensitivity and decreased responsiveness to insulin. A decreased number of insulin receptors or abnormal translocation of glucose transporters is suspected. As the disease progresses, pancreatic insulin production may become impaired. Obesity, female gender, family history, older age, and lack of exercise are risk factors.
- Gestational diabetes is a disorder of glucose intolerance that is diagnosed during pregnancy. Placental hormones and weight gain are contributing factors. High infant birth weight and neonatal hypoglycemia are common complications. Gestational diabetes is a risk factor for the later development of type 2 diabetes.

BOX 41.3 Screening for Diabetes

Testing should be considered in all adults who are overweight (BMI ≥25 kg/m^2) and have additional risk factors:

- Physical inactivity
- First-degree relative with diabetes
- Members of a high-risk ethnic population (e.g., African American, Latino, Native American, Asian American, and Pacific Islander)
- Women who delivered a baby weighing >9 lb or were diagnosed with GDM
- Hypertension (≥140/90 mm Hg) or receiving therapy for hypertension
- HDL cholesterol level <35 mg/dL and/or triglyceride level >250 mg/dL
- Women with polycystic ovarian syndrome
- IGT or IFG on previous testing
- Other clinical conditions associated with insulin resistance (e.g., severe obesity and acanthosis nigricans)
- History of CVD

In the absence of these criteria, testing for prediabetes and diabetes should begin at age 45 years.

If results are normal, testing should be repeated at least at 3-year intervals, with consideration of more frequent testing depending on initial results and risk status.

Data from American Diabetes Association: Standards of medical care in diabetes, *Clin Diabetes* 2016;34(1):3–21.

CVD, Coronary vascular disease; *HDL,* high-density lipoprotein; *IFG,* impaired fasting glucose; *IGT,* impaired glucose tolerance.

CLINICAL MANIFESTATIONS AND COMPLICATIONS

Acute Hyperglycemia

Etiology. Acute hyperglycemia is most commonly caused by alterations in nutrition, inactivity, inadequate use of antidiabetic medications, or any combination of these factors. Persistent fasting hyperglycemia can occasionally be attributed to the dawn phenomenon, which is a rise in blood glucose concentration in the early morning hours attributed to increased growth hormone, cortisol, glucagon, and epinephrine release.

Complications. A primary concern both of health care providers and of individuals with diabetes is avoiding the acute and chronic complications. Acute complications of diabetes include the signs and symptoms of hyperglycemia—polydipsia, polyphagia, and polyuria—and concomitant metabolic and fluid problems. Prolonged insulinopenia can result in ketoacidosis and nonketotic hyperglycemic coma with the accompanying more severe electrolyte and fluid derangements.

Acute complications of diabetes also include infections, most commonly of the skin, urinary tract, and vagina. Infections that particularly affect elderly diabetic patients include malignant otitis externa, necrotizing fasciitis, and persistent candidal infections. Tuberculosis infection and reactivation can be a particular problem in diabetic residents of extended care facilities.

Nausea, fatigue, and a generally decreased sense of well-being frequently accompany hyperglycemia. Blurred vision is a common short-term problem of acute hyperglycemia. These symptoms can be quite distressing and uncomfortable. Acute complications are directly linked to hyperglycemia and recede as euglycemia is approached.

Diabetic Ketoacidosis

Continued insulin deficiency and other hormonal influences (increased levels of catecholamines, cortisol, glucagon, and growth hormone, in part caused by hypovolemia, physical stress, or insulin deficiency itself) lead to lipolysis in body tissues. As the catabolic process continues, metabolism of fats stored in adipose tissue leads to the production of fatty acids. The resulting fatty acids undergo transformation to ketoacids in the liver. Hepatic gluconeogenesis in response to tissue glucose deprivation is also responsible for the increased production of ketoacids. Under normal circumstances, ketoacids can be used by neural and muscle tissue in energy metabolism. When the normal pathway is saturated, the pH falls (6.8 to 7.3) and ketone bodies are present in the urine, thus sharply increasing osmotic fluid loss. Metabolic acidosis ensues as the bicarbonate concentration decreases, and diabetic ketoacidosis results. In response to the metabolic acidosis, extracellular hydrogen ions are transported intracellularly. Physiologically, potassium is constantly leaking into the vascular space through diffusion and is transported intracellularly via the sodium–potassium pump. This active transport mechanism requires insulin, and in the presence of ketoacidosis, the absence of insulin results in hyperkalemia. Serum potassium levels rise (transient hyperkalemia) and excess potassium is excreted into the urine, eventually leading to a net potassium loss. Losses of sodium, magnesium, and phosphorus also occur as the amount of total body water decreases. Serum levels of the ions may be normal or elevated due to hypovolemia. Hypovolemia and dehydration also account for increased values of the following: hematocrit, hemoglobin (Hb), protein, white blood cell count, creatinine, and serum osmolality (Table 41.2). Lactic acidosis, or an excessive amount of lactate, a product of glucose metabolism, can also be present because of hypovolemia and possibly reduced uptake of lactate by the liver as a result of acidosis. Hypovolemia and muscle catabolism are present in persons with ketoacidosis and often at diagnosis of type 1 diabetes. Hypovolemic shock can lead to death if the patient is not promptly treated.

TABLE 41.2 Diabetic Ketoacidosis and Nonketotic Hyperglycemic Hyperosmolar Syndrome

Parameter	Diabetic Ketoacidosis	Nonketotic Hyperglycemic Hyperosmolar Syndrome
Glucose	>300 mg/dL	>600 mg/dL
Urine ketones	Moderate to high	None
pH	6.8 to 7.3	Normal
Na^+, K^+	Low, normal, or high	Low, normal, or high
Hct, Hb, protein, WBC count, Cr, BUN, serum osmolality	High	High

BUN, Blood urea nitrogen; *Cr,* creatinine; *Hb,* hemoglobin; *Hct,* hematocrit; *WBC,* white blood cell.

Respiratory compensation for the metabolic acidosis in the form of deep, labored respirations that are "fruity" in odor (Kussmaul respirations) results in lowered P_{CO_2} values (compensatory respiratory alkalosis). Ketoacidosis may be the initial symptom of a new diagnosis of type 1 diabetes. Other factors that may precipitate ketoacidosis are intercurrent illness and inadequate treatment. The need for hospitalization for diabetic ketoacidosis is greatest for individuals less than 45 years of age (39.7 per 1000 individuals with diabetes).

Nonketotic Hyperglycemic Hyperosmolar Syndrome

The relative lack of insulin seen in individuals with type 2 diabetes mellitus leads to similar but not identical sequelae as the absolute lack of insulin of type 1 diabetes mellitus. The initial symptoms of polyuria, polydipsia, and polyphagia may be present, possibly in a more subtle form. Ketoacidosis is an uncommon occurrence in type 2 diabetes. The presence of endogenous insulin in type 2 diabetes suppresses the lipolysis that leads to the production of ketone bodies and subsequently ketoacidosis. More common in type 2 diabetes mellitus, especially in older individuals, is nonketotic hyperglycemic hyperosmolar syndrome (NHHS), characterized by severe hyperglycemia with no or slight ketosis and striking dehydration. NHHS is more likely to occur in institutionalized patients, especially patients unable to recognize or respond appropriately to thirst. Diabetic ketoacidosis and NHHS can be life-threatening events, especially in the elderly. Seventy percent of all cases of NHHS coma occur in individuals older than 64 years. The mortality rate is approximately 11%.

Chronic Hyperglycemia

Chronic complications associated with diabetes are extensive and are generally placed into two categories: vascular and neuropathic. The vascular complications are further subdivided into macrovascular and microvascular components. Microvascular complications affect the capillaries, and macrovascular complications involve damage to large vessels.

Vascular Complications

Macrovascular Complications

Macrovascular complications of diabetes mellitus are defined as damage to the large blood vessels providing circulation to the brain, heart, and extremities. Although atherosclerosis is an age-dependent process, the presence of diabetes results in accelerated atherosclerosis. These complications include cardiovascular disease and stroke (38.1% of all individuals with diabetes mellitus over the age of 35), as well as peripheral arterial disease. The presence of heart and blood vessel disease is increased twofold to fourfold in individuals with diabetes and is responsible for more than half of all deaths. Ischemic cerebrovascular accidents (strokes) are more prevalent in individuals with diabetes and are associated with poorer outcomes. Cerebrovascular accidents are responsible for almost 1% of hospitalizations of individuals with diabetes mellitus.

Several studies have investigated whether intensive glycemic control (HbA_{1C} <6.5%) would result in a decrease in cerebrovascular events, cardiovascular events, and all-cause mortality. The ACCORD study was the only study to suggest an increase in overall mortality but a significant reduction in nonfatal coronary events. Other studies have demonstrated improved overall mortality with aggressive glycemic control (mean HbA_{1C} of 6.5%) and a 14% reduction in cardiovascular events for a 1% decrease in HbA_{1C}. Despite the reduced mortality, the risk of severe hypoglycemic events significantly increased.

Diabetes is an independent risk factor for coronary artery disease. However, several important risk factors—dyslipidemia, hypertension, and impaired fibrinolysis—are present in uncontrolled diabetes and decrease with improved blood glucose level control. The latter risk factor may be linked to the presence of the compensatory hyperinsulinemia of type 2 diabetes mellitus. Reduction of insulin resistance by such hygienic measures as caloric restriction and exercise, and possibly by pharmacologic means, may be of principal importance in reducing the incidence of macrovascular complications. Conventional measures of risk reduction, including measures to control dyslipidemia and hypertension, continue to be considered essential. Improved glycemic control is thought to affect the microvasculature rather than the macrovascular complications.

Microvascular Complications

The microvascular complications of diabetes—retinopathy and nephropathy—are thought to result from abnormal thickening of the basement membrane in capillaries. Capillary basement membrane thickening has been shown to increase with the length of time after diagnosis and with persistent hyperglycemia.

Hyperglycemia has been shown to disrupt platelet function and growth of the basement membrane. The presence of proteins altered by high glucose levels (advanced glycation end products) is also believed to play a part in the pathogenesis of microvascular complications. Thickening of capillary basement membranes has been shown to decrease with improved glycemic control. Other risk factors for microvascular disease include hypertension and smoking.

Retinopathy is evident in less than 10% of those diagnosed with diabetes for less than 5 years. However, more than 50% of patients diagnosed with diabetes for more than 20 years are found to be suffering with some degree of retinopathy. Because of the prevalence of long-standing undiagnosed diabetes mellitus, as many as 21% of individuals with newly diagnosed type 2 diabetes are affected by retinopathy.

Retinopathy is the primary cause of new cases of blindness in adults in the United States. The incidence of diabetic retinopathy appears to correlate with the duration of diabetes. Retinopathy is a progressive disease involving three stages: background retinopathy, preproliferative retinopathy, and proliferative retinopathy. Background retinopathy is characterized by microaneurysms and small hemorrhages in the retinal capillaries. Background retinopathy usually does not affect visual acuity. Preproliferative and proliferative retinopathies involve further damage to retinal capillaries, with the latter condition characterized by capillary neovascularization. The small new capillaries are particularly prone to hemorrhage. Proliferative retinopathy is managed with laser photocoagulation.

Nephropathy affects 20% to 40% of individuals with type 1 diabetes. Fewer individuals with type 2 progress to end-stage renal disease. Diabetic

nephropathy accounts for 40% of cases of end-stage renal disease. Ethnic origin is a risk factor in the development of diabetic nephropathy, with African American, Hispanic, and Native American diabetic individuals experiencing an increased rate of end-stage renal disease compared with Caucasians. The characteristic lesion of diabetic nephropathy is glomerulosclerosis, or thickening and hardening of the basement membrane of capillaries in the glomeruli. Filtration, an essential component of kidney function, occurs in the glomerulus. The first stage of diabetic nephropathy is an increase in the glomerular flow rate, or the rate of blood flow through the glomerulus. This increased flow rate leads to hyperfiltration in the glomerulus, or a rise in the rate at which blood is filtered. The mechanisms leading to the increase in glomerular flow rate are unclear but are evidently related to poor glycemic control. As hyperfiltration progresses, the glomeruli become damaged. The resulting glomerulosclerosis leads to blockage and leaking of the capillaries. Protein is characteristically seen in the urine, at first in small amounts (microalbuminuria) and then grossly. As diabetic nephropathy advances, the glomerular filtration rate drops and chronic kidney disease ensues. Hypertension is an important contributing factor to diabetic nephropathy. Management of hypertension with medications that inhibit angiotensin-converting enzyme has been shown to reduce the rate of chronic kidney disease, end-stage renal disease, and mortality.

Neuropathic Complications

Diabetic neuropathy produces symptoms in 60% to 70% of individuals with diabetes and is responsible for 6.8 per 1000 diabetic population hospitalizations. Neuropathic complications are divided into autonomic dysfunction and sensory dysfunction. Autonomic complications include gastrointestinal disturbances, bladder dysfunction, tachycardia, postural hypotension, and sexual dysfunction. Approximately 35% to 50% of men with diabetes experience erectile dysfunction. Sensory disturbances include carpal tunnel syndrome and paresthesias or lack of sensation in the feet and lower legs. Neuropathy is largely responsible for the increased risk of serious foot problems in individuals with diabetes. The rate of lower extremity amputation in individuals with diabetes is 15 to 40 times higher than in nondiabetic individuals. More than 60% of all nontraumatic amputations in the United States are performed on individuals with diabetes.

Diabetes in humans and experimentally induced diabetes in animals are associated with decreased levels of myoinositol in peripheral nerves. Myoinositol is a cell membrane component normally found in abundance in nerve tissue. Several theoretical explanations have been proposed for the myoinositol link to neuropathy. Glucose appears to compete with myoinositol in transport into the cell. Degradation of glucose to sorbitol and fructose (the polyol pathway) occurs in the nerves in the presence of hyperglycemia and insulinopenia. Increased activity of the polyol pathway also appears to be linked to reduced amounts of myoinositol in the peripheral nerves. Focal ischemic lesions of the nerves may also have a role in diabetic neuropathy. Pathologic findings include degeneration or loss of nerve fibers resulting in decreased nerve function.

Glycemic control has been shown to improve nerve function in animals and in humans and to decrease perceived pain. In addition to hyperglycemia, hypertriglyceridemia, obesity, smoking, and hypertension are modifiable vascular risk factors for the development of neuropathy. Other risk factors associated with neuropathy include male gender, white race, older age, and possibly height >175.5 cm. In addition, the elderly, non-Hispanic blacks, and Mexican Americans are disproportionately affected. Strong evidence linking prolonged hyperglycemia to neuropathy and the microvascular complications of diabetes was provided by the Diabetes Control and Complications Trial, a 9-year multicenter prospective study designed to examine the effect of intensive insulin therapy on the development of retinopathy. Renal and neurologic indices were also examined. Subjects in this trial were individuals with type 1 diabetes and no or mild retinopathy at baseline. The experimental group was intensively treated with three or more insulin injections daily or with insulin delivered by a pump. The incidence of initial retinopathy was reduced by 76%, and progression of existing retinopathy was reduced by 54% in the experimental group. Indices of beginning nephropathy such as microalbuminuria and albuminuria were reduced in the experimental group by 39% and 54%, respectively. Symptomatic neuropathy was reduced by 60% in the experimental group. Dilemmas presented by the Diabetes Control and Complications Trial include the presence of a significant increase in severe hypoglycemia in the experimental group and some question about the validity of extrapolating all results to individuals with type 2 diabetes. A large study on the effect of lowering blood glucose level in type 2 diabetes has also yielded data on decreased morbidity and mortality with improved glycemic control (United Kingdom Prospective Diabetes Study).

Complications in Pregnancy

Pregnancy in women with type 1 diabetes has been complicated by an increased risk for perinatal infant mortality and congenital anomalies. Metabolic control during pregnancy reduces the risk of perinatal mortality to a rate approximating that of the general population. An increased rate of congenital malformations continues to attend pregnancies in women with type 1 diabetes. Because the affected organs develop early in the first trimester, excellent glycemic control before conception is recommended. Untreated maternal hyperglycemia can result in increased rates of macrosomia, shoulder dystocia, and preeclampsia, as well as death, fractures, and nerve palsies.

KEY POINTS

- The acute complications of diabetes are hyperosmolar coma, ketoacidosis, and infection. Hyperglycemia may be associated with nausea, fatigue, and blurred vision.
- Ketoacidosis occurs primarily in type 1 diabetes mellitus as a result of increased lipolysis and conversion to ketone bodies. Excessive ketones result in metabolic acidosis, which is recognized by a fall in pH and bicarbonate levels. Ketoacidosis may occur in patients with type 2 diabetes mellitus under severe stress, such as concomitant sepsis, stroke, or myocardial infarction. Acidosis-induced hyperkalemia and compensatory hyperventilation (Kussmaul respirations) resulting in reduced levels of arterial carbon dioxide are associated findings.
- Nonketotic hyperosmolar syndrome is more common in type 2 diabetes because endogenous insulin suppresses ketone formation and thus prevents ketoacidosis. Hyperglycemia may go untreated for a time and result in persistent glycosuria with osmotic diuresis. Dehydration may be manifested as high osmolality and hemoconcentration of erythrocytes, proteins, and creatinine.
- The chronic complications of hyperglycemia are primarily caused by vascular and neuropathic dysfunction. Individuals with diabetes are prone to vascular complications, including coronary artery disease, stroke, and peripheral arterial disease. These complications are related to dyslipidemia, hypertension, and impaired fibrinolysis. Retinopathy and nephropathy are thought to be due to hyperglycemia-induced thickening of retinal and glomerular basement membranes. Neuropathy is manifested as pain and loss of sensation. Excessive glucose is thought to interfere with myoinositol in neurons.

TREATMENT AND EDUCATION

The usual treatment goals in diabetes are achieving metabolic control of blood glucose levels and preventing acute and chronic complications. The American Diabetes Association recommends as goals a preprandial

blood glucose level between 70 and 130 mg/dL and a postprandial blood glucose level less than 180 mg/dL for adults with diabetes. A glycemic control algorithm to guide treatment is presented in Fig. 41.9. The goals of treatment are accomplished by diet, exercise, medication, and such hygiene practices as daily foot care and smoking cessation. Each treatment involves lifestyle changes that are difficult to accomplish initially and challenging to maintain. Treatment must be individualized to the type of diabetes and the unique traits of each patient.

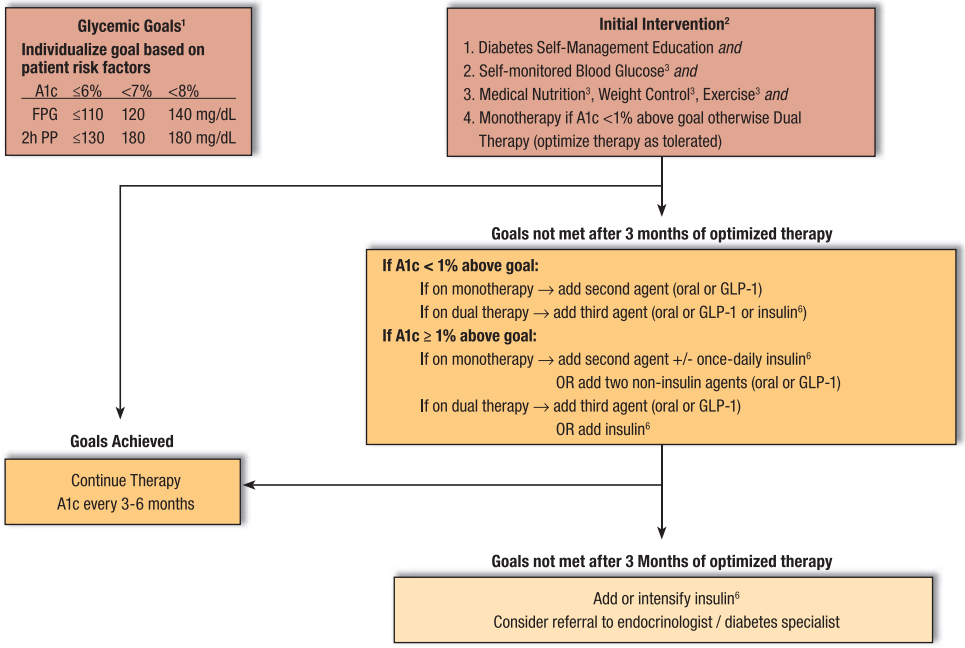
Nutrition

Nutrition has often been called the cornerstone of diabetes therapy. Ideas about the optimal dietary prescription have been far from constant throughout recorded history. From wheat, fruit, and beer in ancient Egypt through blood pudding and rancid meats in nineteenth-century France to more recent investigations of fiber and fat, nutritional controversies are far from over.

In a position statement, the American Diabetes Association has listed four recommendations for nutritional therapy in patients with diabetes (Box 41.4). Accomplishing the goals can involve changes in composition of the diet, meal patterns and timing, and caloric consumption. All the energy nutrients—carbohydrates, fats, and protein—have an essential role in optimal nutrition. Obesity and eating disorders such as bulimia have an important impact on nutritional status.

Obesity and Eating Disorders

Obesity is the strongest risk factor for type 2 diabetes, in addition to being a risk factor for cardiovascular disease in women. Obesity is defined as a body mass index (BMI) of greater than 30 kg/m^2. BMI is calculated by dividing body weight in kilograms by the square of the height in meters. Increased risk for health problems occurs at a BMI equal to or greater than 25.2 kg/m^2. A BMI of 18.5 to 24.9 kg/m^2 is considered normal. Weight management is a chronic and difficult



Recommended Options for Dual Therapy[4]

Metformin
- \+ TZD or DPP-4 or SU[5] or GLP-1 or Meglitinide or colesevelam

Recommended Options for Triple Therapy

Metformin
- \+ TZD or SU[5]
- \+ GLP-1 or DPP-4 or AGI or colesevelam

Metformin
- \+ TZD or DPP-4 or AGI or SU[5] or colesevelam
- \+ Insulin

Abbreviations

AGI	Alpha-glucosidase inhibitors
DPP-4	Dipeptidyl peptidase-4 Inhibitor
FPG	Fasting plasma glucose
GLP-1	Glucagon-like peptide-1 agonist
PP	Postprandial
SU	Sulfonylurea
TZD	Thiazolidinedione

Footnotes

1 **Intensify management if:** Absent/stable cardiovascular disease, mild-moderate microvascular complications, intact hypoglycemia awareness, infrequent hypoglycemic episodes, recently diagnosed diabetes. **Less intensive management if:** Evidence of advanced or poorly controlled cardiovascular and/or microvascular complications, hypoglycemia unawareness, vulnerable patient (ie, impaired cognition, dementia, fall history). Refer to TDC "A1c Goal" treatment strategy for further explanation. A1c is referenced to a non-diabetic range of 4-6% using a DCCT-based assay. ADA Clinical Practice Recommendations. *Diabetes Care* 2010;33(suppl 1):S19-20.

2 If initial A1c on presentation is ≥10%, consider the use of insulin, with or without oral agents, as the initial intervention (see Insulin Algorithm). Other agents may be introduced as glycemic control improves. If ketoacidosis or recent rapid weight loss, consider Type 1 diagnosis.

3 These interventions should be maintained life-long; (refer to Medical Nutrition, Weight Loss, and Exercise Algorithms).

4 Refer to the Diabetes Medications Supplement: Working Together to Manage Diabetes found in the Texas Diabetes Council's Diabetes Toolkit.

5 If a SU is selected, low dose glipizide ER or glimepiride are recommended because they have a lower incidence of hypoglycemia than glyburide.

6 Refer to Insulin Algorithm for Type 2 Diabetes Mellitus in Children and Adults / Initial Insulin Therapy for Type 2
Diabetes Mellitus in Children and Adults: A Simplified Approach

FIG 41.9 Algorithm to achieve glycemic control for type 2 diabetes mellitus in adults. *ER,* Extended release (Reprinted with permission from the Texas Diabetes Council, Texas Department of State Health Services. www.tdctoolkit.org.)

BOX 41.4 American Diabetes Association Recommendations for Nutritional Therapy

1. To promote and support healthful eating patterns, emphasizing a variety of nutrient-dense foods in appropriate portion sizes, in order to improve overall health and specifically
   a. Attain individualized glycemic blood pressure and lipid goals
   b. Achieve and maintain body weight goals
   c. Delay or prevent complications of diabetes
2. To address individual nutritional needs based on personal and cultural preferences, health literacy and numeracy, access to healthful food choices, willingness and ability to make behavioral changes, and barriers to change
3. To maintain the pleasure of eating by providing positive messages about food choices while limiting food choices only when indicated by scientific evidence
4. To provide the individual with diabetes with practical tools for day-to-day meal planning rather than focusing on individual macronutrients, micronutrients, or single foods

Data from American Diabetes Association: Nutrition recommendations and interventions for diabetes, *Diabetes Care* 2016;39(Suppl 1):S20–S30.

TABLE 41.3 Oral Antidiabetic Agents

Category	Class	Drugs
Sensitizers	Biguanides	Metformin
	Thiazolidinediones (TZDs)	Pioglitazone, rosiglitazone
Secretagogues	Sulfonylureas	First generation: carbutamide, chlorpropamide, tolbutamide, tolazamide
		Second generation: glyburide, glipizide, glimepiride
	Meglitinides	Nateglinide, repaglinide, mitiglinide
	DDP-4 inhibitors	Linagliptin, saxagliptin, sitagliptin, vildagliptin
Other	α-Glucosidase inhibitors	Acarbose, miglitol
	Gliflozins (SGLT-2 inhibitors)	Canagliflozin (Invokana), dapagliflozin (Farxiga), and empagliflozin (Jardiance).

problem for many individuals. No single strategy has been shown to be effective for all individuals. Current recommendations for management of obesity include the use of a nutritionally complete diet, a maintenance routine, and an exercise program. A modest weight loss of 2 to 8 kg may provide significant clinical benefits in those with type 2 diabetes. Consult Chapter 42 for additional nutritional risk factors related to obesity.

Eating disorders such as bulimia and anorexia may be more common in women with type 1 diabetes. Inducing weight loss by reducing insulin levels has also been noted. In order for clinicians to be aware of eating disorders in their patients, careful assessment is necessary.

Exercise

Exercise, one of the oldest treatments for diabetes, was prescribed in India in 500 BC. Exercise can have a role in both type 1 and type 2 diabetes to lower blood glucose levels and promote health maintenance. Exercise lowers such cardiovascular risk factors as high blood pressure and dyslipidemia, increases work capacity, reduces stress, prevents bone loss, and improves reaction time. Exercise may also be beneficial in weight reduction.

The effects of exercise on fuel utilization and insulin sensitivity are comparable to exercise-induced changes in normal metabolism. As glucose production increases and plasma insulin levels drop, the reduction in tissue insulin resistance can result in a net decline in blood glucose levels. Exercise has the potential to decrease insulin requirements in type 1 diabetes, decrease and possibly eliminate the need for pharmacologic agents in type 2 diabetes, and reduce the risk for heart disease in all persons with diabetes. It has been shown to prevent the onset of type 2 diabetes in persons who are genetically at risk.

Although the benefits of exercise are numerous, there are associated risks. Individuals with type 1 diabetes are at risk for hypoglycemia and ketoacidosis. Exercise in individuals with type 2 diabetes can be associated with hypoglycemia, cardiac dysfunction, orthopedic injury, and worsening of some complications.

When insulin or an oral hypoglycemic agent is used in the management of diabetes, the usual fuel metabolism of exercise is disturbed. Inappropriately high insulin levels result in decreased hepatic glucose production and increased tissue insulin sensitivity. The latter processes can lead to hypoglycemia. Replacement of expended glycogen and continued insulin sensitivity can result in hypoglycemia as long as 24 hours after activity.

Hyperglycemia and ketosis may be a result of exercise under insulinopenic conditions. Safeguards must be built into exercise programs to prevent hypoglycemia, ketoacidosis, injury, cardiac compromise, and exacerbation of diabetic complications.

The exercise prescription is as essential in diabetes management as the medication or nutrition prescription. The CDC recommends an exercise program of moderate intensity for at least 30 minutes, 5 or more days per week or vigorous intensity for 20 minutes, 3 or more days per week (with no more than 2 consecutive days without exercise). All individuals should be encouraged to reduce sedentary time by breaking up extended amounts of time (>90 min) spent sitting.

Exercise should be avoided if ketosis is present, because it can indicate an acute shortage of insulin. Exercise under the latter conditions can lead to increased hyperglycemia. Exercise when blood glucose values are greater than 250 mg/dL is safe if ketosis is not present.

Pharmacologic Agents

Oral Antidiabetic Agents

Along with diet and exercise as an initial treatment of type 2 diabetes, numerous medications may be used. A patient-centered approach should be used to guide the choice of pharmacologic agents. Considerations include efficacy, cost, potential side effects, weight, comorbidities, hypoglycemia risk, and patient preferences.

Current guidelines call for timely implementation and titration of oral agents as well as insulin to achieve euglycemia. Sulfonylurea drugs have been used in the management of type 2 diabetes for more than 40 years. These drugs have been joined by other agents with different mechanisms of action. Table 41.3 includes a list of common antidiabetic agents.

Metformin, classified as a biguanide, suppresses hepatic gluconeogenesis and enhances glucose uptake by peripheral tissues without causing hypoglycemia. Metformin, if not contraindicated and if tolerated, is the first-line drug for type 2 diabetes. Metformin has been used alone and in combination with other oral agents and is associated with improvement in dyslipidemia and weight loss. Side effects include nausea and diarrhea. Serious side effects of metformin include acute renal injury and metabolic acidosis. In a large review of several thousand patient-years, the incidence of metformin-associated lactic acidosis is rare and most commonly associated with acute kidney injury. Caution

should be employed when using metformin in patients with liver disease. Administration of radiographic iodine dyes has potential risk for causing acute renal injury. Patients who are taking metformin have an increased risk of developing metabolic acidosis. Depending on the creatinine clearance value, some patients should hold their metformin for 24 hours before and 48 hours after receiving IV contrast medium.

Sulfonylureas are often second-line antihyperglycmeic agents and exert their hypoglycemic effect by binding to the adenosine triphosphate (ATP)–dependent potassium channels on the cell membrane of β cells. This inhibits efflux of potassium, resulting in a rise in intracellular calcium concentration and the secretion of proinsulin. Sulfonylureas are effective in augmenting the action of insulin in glucose disposal, diminishing insulin clearance by the liver, and reducing hepatic glucose production. Because sulfonylureas are ineffective in the management of type 1 diabetes, stimulation of β cells is known to be a crucial factor in the action of these oral agents. Enhanced insulin secretion appears to be a short-term effect, possibly caused by a reduction in insulin requirements as a result of the other hypoglycemic activities of sulfonylurea agents.

The so-called *first-generation agents* (those formulated earliest) are tolbutamide, chlorpropamide, and tolazamide. Second-generation agents include glyburide, glipizide, and glimepiride. Use of first-generation agents is not recommended unless patients have a well-established history of previous use with good results. Second-generation agents are more potent than first-generation agents, possibly because of increased capacity to bind to the plasma membrane of the β cell. In addition, they are more predictable, have fewer side effects, and represent more convenient dosing.

The side effects of sulfonylureas include hypoglycemia, nausea, dizziness, headache, allergic reactions, and flushing with alcohol use (disulfiram effect). Sulfonylureas that are metabolized to inactive compounds by the liver are considered safer for use in renal disease. Duration of action is another safety issue. After insulin, sulfonylureas are the major cause of severe hypoglycemia caused by a drug. The longer-acting sulfonylureas, chlorpropamide and glyburide, are responsible for the majority of cases of severe hypoglycemia and fatal hypoglycemic coma.

Acarbose and miglitol, α-glucosidase inhibitors, diminish postprandial hyperglycemia by delaying carbohydrate absorption. They can be used alone and in combination with sulfonylurea drugs, metformin, or insulin. Side effects include symptoms related to decreased gastrointestinal absorption (e.g., flatulence and diarrhea). They do not cause hypoglycemia but can complicate its management if used in combination with a sulfonylurea. Sulfonylurea-induced hypoglycemia cannot be managed with sucrose when acarbose is being used because of drug-induced delayed sucrose absorption.

Thiazolidinedione drugs increase tissue sensitivity to insulin and inhibit hepatic gluconeogenesis. Thiazolidinediones are thought to decrease insulin resistance, increase glucose uptake, and redistribute fat. In addition, they are believed to preserve β-cell function, decrease vascular inflammation, and minimally decrease hepatic glucose production. Currently pioglitazone is the only commercially available thiazolidinedione; rosiglitazone is available only through selective distribution points because of the cardiovascular risks.

When prescribed along with an appropriate meal plan, oral antidiabetic agents can be very effective in the management of type 2 diabetes mellitus. Primary failure of the drug is considered to have occurred when initiation of oral agent therapy does not result in a significant decline in blood glucose levels. Primary failure can be due to misdiagnosis of type 1 diabetes mellitus or inadequate adherence to the diet and exercise regimen. Secondary failure, or hyperglycemia after an effective initial response to the drug, is often due to dietary nonadherence but may also be due to the progressive β-cell dysfunction of type 2 diabetes mellitus.

Incretin Enhancers, Incretins, and Amylins

A newer group of oral agents for management of diabetes are the incretins, incretin enhancers, and amylins. If noninsulin monotherapy at the maximum tolerated dose does not achieve or maintain the HbA_{1C} target over 3 months, then a GLP-1 receptor agonist may be added.

Dipeptidylpeptidase-4 (DPP-4) is a widely expressed enzyme that is responsible for rapidly degrading GLP-1 and GIP. Both of these substances are critical factors in the development and treatment of type 2 diabetes. DPP-4 inhibitors block the secretion of DPP-4, causing sustained action of incretin hormones that results in satiety, lower serum glucose levels, and enhanced β-cell mass and function. DPP-4 inhibitors prevent the release of glucagon from α cells through the release of incretin while increasing insulin secretion. Medications in this class include sitagliptin, saxagliptin, linagliptin, and vildagliptin.

The incretin class of medications focuses on GLP-1. GLP-1 mimetics delay gastric emptying, inhibit release of glucagon, and increase satiety. These drugs are injectable medications with the common side effects of nausea and hypoglycemia. Exenatide can be used alone or in combination with sulfonylureas, metformin, or thiazolidinediones. Medications in this class include exenatide, liraglutide, and lixisenatide.

Pramlintide is an amylin analog and can be used in conjunction with insulin for the management of glycemia. It may not be mixed with insulin and should be injected separately. Insulin requirements are usually decreased by 50% with pramlintide; in addition, minimum carbohydrate and caloric intake is required for its use. It is contraindicated in patients with hypoglycemia unawareness or gastroparesis. Both incretin and amylin mimetics have been effective in weight loss and lowering HbA_{1C} values when used in conjunction with metformin or a sulfonylurea (both) or insulin (pramlintide only).

Gliflozins are the newest class of antihyperglycemic agents that inhibit reabsorption of glucose in the kidney. In normal renal physiology, the kidneys filter glucose out of the blood then reabsorb glucose back into the blood through the sodium–glucose cotransportor 2 receptor (SGLT2). During states of hyperglycemia, the SGLT2 receptors are at capacity, and excessive glucose is excreted in the urine. SGLT2 inhibitors prevent renal reabsorption of glucose, resulting in increased renal excretion and thus reducing serum glycemia. Currently three SGLT2 inhibitors are approved in the United States: canagliflozin (Invokana), dapagliflozin (Farxiga), and empagliflozin (Jardiance). This group of medications has no effect on the pancreatic β cells and has a minimal risk of hypoglycemia.

Insulin

Insulin therapy is required in all persons with type 1 diabetes mellitus and in 35% of persons with type 2 diabetes mellitus. Persons with type 1 diabetes mellitus require replacement of the deficient hormone in a manner that most closely resembles normal physiologic mechanisms. The role of insulin in type 2 diabetes mellitus is more complex.

Because type 2 diabetes mellitus is a progressive disease, many (if not most) individuals will need insulin at some time, either because of increasing insulin resistance or because of β-cell dysfunction. Glucose toxicity, a phenomenon in which insulin resistance and decreased production of insulin are worsened by hyperglycemia, may respond to insulin therapy. Insulin may be necessary in type 2 diabetes intermittently during times of physiologic stress that increase insulin requirements (e.g., concomitant illness, surgery, inactivity, weight gain). However, with resolution of the stress, the individual may be able to discontinue insulin use.

Types of insulin are classified by a variety of features, but from a practical standpoint are grouped according to the duration of action: rapid acting, short acting, intermediate acting, and long acting

(Table 41.4). The most commonly used insulins in the ultra rapid-acting category are aspart, glulisine, and lispro. Regular insulin is considered short acting. Intermediate-acting agents include isophane. Glargine and detemir are in the long-acting, or basal insulin group. Advances have been made with the delivery of insulin, including Technosphere insulin (TI), which is a fast-acting form of insulin delivered via an inhaled powder. TI was approved for the management of type 1 diabetes mellitus with concomitant use of basal insulin, with an onset of action and peak of 30 and 53 minutes, respectively. This new form of insulin delivery has added benefits, including reduced social stigma and decreased pain from injections. Novel forms of insulin and incretins are being developed, including oral insulin tablets, oral insulin spray, and oral incretins.

Patterns of insulin use vary with the type of diabetes and the degree of desired metabolic control. For patients with type 1 diabetes mellitus, a minimum of two or more daily injections of rapid-acting and long-acting insulins has been used to control postprandial and fasting hyperglycemia. A popular regimen includes use of long-acting glargine with preprandial injections of a rapid-acting insulin.

In the management of type 2 diabetes, insulin is initiated early in the course of the disease to achieve and maintain glycemic control. It can be used concomitantly with other antidiabetic agents. The intensity of the regimen is based on the needs of the patient. A benchmark of an HbA_{1C} of 7% has been suggested as the target. In patients who have confounding variables, this target may be adjusted down to an HbA_{1C} of 6.5% or less or increased to 7.5% or more. Variables to consider are risks associated with hypoglycemia, disease duration, life expectancy, established vascular complications, patient attitudes, and patient resources.

Long-acting insulin or rapid-acting insulin can be used to improve fasting or postprandial glycemia, respectively, depending on the patient's needs. Normally, fasting glucose is targeted first; if HbA_{1C} goals are not met, prandial insulin is initiated. Mixed formulations are beneficial and require only one to two injections per day in type 2 diabetes.

The action of insulin is affected by many elements, including climate, alteration in blood flow, tobacco use, and the injection site. Insulin is absorbed most rapidly from the abdomen, less rapidly from the arm, and most slowly from the legs and buttocks. Insulin is absorbed more rapidly from areas that are exercised or massaged after injection.

Hypoglycemia complications. Hypoglycemia is the most common complication of hyperglycemic therapy and the most hazardous. Neural tissue depends on a constant supply of glucose for normal function. When insufficient food intake, unplanned activity, or an inappropriate insulin or sulfonylurea dose lowers the blood glucose concentration excessively, counterregulatory mechanisms are activated to ensure a continued supply of glucose to the brain. The counterregulatory mechanism that commences with the activation of the sympathetic nervous system is a response to hypoglycemia that results in the release of glucagon, corticosteroids, and growth hormones.

Symptoms of hypoglycemia produced by counterregulatory mechanisms include pallor, tremor, diaphoresis, palpitation, and anxiety. Neuroglycopenic symptoms noted in hypoglycemia are hunger, visual disturbance, weakness, paresthesias, confusion, agitation, coma, and death. In long-standing diabetes, neuropathy can alter the counterregulatory mechanisms. Hypoglycemic unawareness, in which the diabetic patient does not experience counterregulatory symptoms, can be the result.

Other complications of insulin therapy. Another typical complication of insulin therapy is lipodystrophy. Lipoatrophy has been linked to the use of insulin from animal and human sources and is manifested as hollows in the surface of the skin caused by the destruction of subcutaneous adipose tissue. The exact mechanism is not clear but is suggested to be derived from an immune-mediated response. Lipohypertrophy is characterized by an increase in subcutaneous tissue because of insulin-stimulated growth of adipose tissue at the injection sites. Avoiding repeated injections at the same site is recommended to prevent lipodystrophy.

An acute complication of insulin use can be insulin edema, or a localized or generalized accumulation of fluid. Weight gain can accompany initiation of insulin therapy, especially when glycemic control is improved. A third complicating factor in insulin therapy is insulin resistance. Insulin resistance is exacerbated by obesity and can necessitate the use of large insulin doses. An appropriate diet and exercise program is as important to insulin-treated individuals as it is to other patients with diabetes.

Stress Management

Living with diabetes can be stressful. The tasks of blood glucose monitoring, medication administration, meal planning, and implementing preventive care to avoid complications can be demanding. Fearing the onset of complications and their impact in addition to living with complications are parts of the psychological impact of diabetes. Depression is more likely to be diagnosed in individuals with diabetes and is correlated with deterioration of glycemic control. Stress management can have an important role in diabetes care by improving quality of life and reducing the possible impact of stress on glycemic control.

Assessment of Efficacy

Clinicians use several measures to determine the adequacy of glycemic control. One indirect but very useful indication of blood glucose levels is the level of glycated hemoglobin. Hemoglobin becomes glycated when glucose is nonenzymatically attached to one of its terminal amino acids. Four glycated hemoglobin products are formed: HbA_{1a1}, HbA_{1a2}, HbA_{1b}, and HbA_{1C}. The latter is produced in the largest quantity and is used in most assays.

Because erythrocytes are freely permeable to glucose, the quantities of glycated hemoglobin formed are proportional to the quantity of glucose in the blood plasma. Glycated hemoglobin values will reflect mean blood glucose levels for the life of the average erythrocyte (100 to 120 days). Highly significant correlations have been found between HbA_{1C} levels and mean blood glucose level. The presence of abnormal hemoglobins or hemolytic anemia can skew results. The normal value for HbA_{1C} varies with the laboratory technique, but is usually less than 7%. Results of the A1C-Derived Average Glucose Trial published in 2008 correlated an HbA_{1C} number to average glucose levels rather than a percentage. This change has little impact on clinical management; however, it decreases patient confusion regarding glycemic control.

HbA_{1C} values are used clinically to estimate long-term control and to establish and evaluate therapeutic goals. Values of less than 7% or as close to normal as possible without adverse effects are considered desirable. Depending on individual circumstances and life expectancy, less stringent goals may be appropriate. HbA_{1C} values cannot be used for daily management of therapy.

Assessment of glycemia on a daily basis was attempted in the past by the use of testing for glycosuria. However, the blood glucose level at which glucose is measurable in the urine, the glycemic threshold, varies from individual to individual, is usually unacceptably high, and cannot be used to establish the presence of hypoglycemia.

The First and Second Consensus Development Conference on Self-Monitoring of Blood Glucose, convened by the American Diabetes Association and other involved agencies, formulated several goals for the use of capillary blood glucose monitoring. The goals included use of capillary blood glucose monitoring to achieve and maintain a specific level of glycemic control, prevent and manage hypoglycemia, avoid severe hypoglycemia, adjust care in response to changes in lifestyle in

TABLE 41.4 Insulin and Other Injectable Antihyperglycemic Medications

Drug Class	Action	Brand Name	Generic Name	Appearance	RX Or OTC	Drug Type
Noninjectable insulin	Rapid	Insulin Human, Inhalation	Afrezza	Powder	Rx	Technospehere insulin inhalation
Insulin	Rapid	Humalog	Insulin lispro	Clear	Rx	Insulin analog
		NovoLog	Insulin aspart	Clear	Rx	Insulin analog
		Apidra	Insulin glulisine	Clear	Rx	Insulin analog
	Short	Humulin R Novolin R	Regular insulin; injectable	Clear	OTC	Regular insulin; injected
	Intermediate	Humulin N Novolin N	Isophane insulin	Cloudy	OTC	Isophane insulin
	Long	Levemir	Insulin detemir	Clear	Rx	Long-acting insulin analog
		Lantus	Insulin glargine	Clear	Rx	Long-acting insulin complex
	Combination products	Humulin 70/30 Novolin 70/30	70% isophane/30% regular	Cloudy	OTC	NPH and regular combination
		NovoLog 70/30	70% aspart protamine/30% aspart	Cloudy	Rx	NPH-like and rapid-acting combination
		Humalog mix 75/25	75% lispro protamine/25% lispro	Cloudy	Rx	NPH-like and rapid-acting combination
		Humalog mix 50/50	50% lispro protamine/50% lispro	Cloudy	Rx	NPH-like and rapid-acting
		Humulin 50/50	50% isophane/50% regular	Cloudy	OTC	NPH and regular combination
Noninsulin Injectables						
Incretin mimetic	Adjunct therapy	Byetta	Exenatide	Clear	Rx	GLP-1 analog
	Adjunct therapy	Victoza	Liraglutide	Clear	Rx	GLP-1 analog
Amylin analog		Symlin	Pramlintide		Rx	Amylin analog

individuals requiring pharmacologic therapy, and determine the need for initiating insulin therapy in women with GDM.

Capillary blood glucose monitoring has been shown to be an accurate reflection of venous blood glucose level when performed by health professionals and by individuals with diabetes. Accuracy can be affected by such performance errors as underloading or overloading of the strip, incorrect placement of the sample, and improper handling of the sample. Training improves performance.

Diabetic individuals using capillary blood glucose monitoring have frequently reported enthusiasm and increased insight into the relationship between blood glucose level and such factors as diet, exercise, and stress and have expressed increased feelings of well-being. Capillary blood glucose monitoring has been associated with improved glycemic control. Monitoring of capillary blood glucose levels is simply a feedback mechanism that provides immediate information on the effects of a change in therapy.

In the replacement of testing for glycosuria with capillary blood glucose monitoring, testing for ketonuria should not be neglected. The presence of ketones in the urine can be an indication of diabetic ketoacidosis and may be harmful to a developing fetus in pregnancy complicated by diabetes. All individuals with diabetes should be tested for ketonuria when blood glucose values are greater than 300 mg/dL, during concomitant illness, during pregnancy, and in the presence of symptoms of diabetic ketoacidosis (nausea, vomiting, abdominal pain).

KEY POINTS

- The mainstays of diabetic treatment are diet, exercise, and drug therapy. Education is an integral part of treatment, enabling individuals with diabetes to follow the diabetic regimen and avoid complications. The efficacy of therapy can be assessed by monitoring blood glucose and HbA_{1C} levels. Blood glucose monitoring is useful for assessing short-term efficacy. HbA_{1C} is a better measure of the long-term efficacy of therapy. A mean blood glucose level less than 170 mg/dL (HbA_{1C} level of 7% or less) is desirable.
- Exercise has several benefits for an individual with diabetes. Insulin requirements may be reduced, weight loss facilitated, and the risk of cardiovascular

Onset (in Hours Unless Noted)	Peak (Hours)	Duration (Hours)	Compatible Mixed With	Storage/ Expiration	Typical Dosing/Comments
15–30 min	1 hr	2.5 hr	NA	Refrigerate/ 10 days at room temp	Black box warning for people with asthma and COPD for risk of bronchospasm
15–30 min	1–2	3–4	NPH	Refrigerate/28 days at room temp	15 min before meal or immediately after meal
15–30 min	1–2	3–5	NPH	Refrigerate/28 days at room temp	5–10 min before meal
15–30 min	1–2	3–4	NPH	Refrigerate/28 days at room temp	15 min before or within 20 min after meal
0.5–1	2–3	3–6	NPH	Refrigerate/28 days at room temp	30 min before meal
2–4	4–6	8–12	Insulin analogs/ injectable regular insulin	Refrigerate/28 days at room temp	One, two, or three times daily
2	6–9	14–24	None	Refrigerate/42 days at room temp	Once or twice daily
4–5	Peakless	22–24	None	Refrigerate/28 days at room temp	Once daily, at same time each day
30 min	1.5–16	24	None	Refrigerate/28 days at room temp	30 min before meal
15 min	1–4	24	None	Refrigerate/28 days at room temp	15 min before meal
15 min	1–6.5	24	None	Refrigerate/28 days at room temp	15 min before meal
15–30 min	1–13	14–24	None	Refrigerate/28 days at room temp	15 min before meal
30 min	2–5.5	24	None	Refrigerate/28 days at room temp	30 min before meal
	2	10	None	Refrigerate/30 days after opening	Within 60 min before morning and evening meals *When initiating, taking close to meals may minimize side effects
	8–12	13+	None	Refrigerate/30 days after opening	Initial dose is 0.6 mg subQ daily × 1 week, then 1.2 mg subQ daily
	20 min	3	None	Refrigerate/28 days at room temp or refrigerated	Immediately before each major meal (≥250 kcal or 30 g of carbohydrate)

complications decreased. Exercise may precipitate hypoglycemia, so insulin injections or dietary intake may have to be adjusted. Oral antidiabetic agents may be successfully used in type 2 diabetes. The sulfonylureas exert their effects primarily by stimulating the release of endogenous insulin. They also reduce insulin degradation and suppress the release of glucose from the liver. Metformin suppresses hepatic gluconeogenesis and enhances glucose uptake by peripheral tissue; thiazolidinediones enhance glucose uptake by peripheral tissue; and acarbose delays absorption of ingested carbohydrate. Newer agents such as incretin mimetics, incretin enhancers, and amylin analogs will also be affected by exercise.

- Insulin replacement therapy is required in all patients with type 1 diabetes and in about one third of patients with type 2 diabetes mellitus. Insulin is classified according to its onset, peak, and duration of action. A combination of insulins may be given to produce optimal control. Long-acting insulin (glargine or detemir) and a rapid-acting insulin (aspart, lispro, glulisine) are commonly used to mimic basal and prandial insulin levels, respectively.
- Hypoglycemia is the most common complication of pharmacologic therapy. Symptoms are mediated primarily by activation of the sympathetic nervous system stress response. Secretion of catecholamines, glucagon, corticosteroids, and growth hormone rises in an attempt to increase blood glucose levels. Pallor, tremor, diaphoresis, weakness, and decreased consciousness are the usual manifestations of hypoglycemia.

PEDIATRIC CONSIDERATIONS

Diabetes has been diagnosed in approximately 176,500 children and adolescents younger than 20 years. One in every 400 to 600 children and adolescents has type 1 diabetes, and 18 new cases per population of 100,000 are diagnosed yearly. The overwhelming majority have type 1 diabetes, with an approximate 5% prevalence of genetic defects in β cells.

Type 1 diabetes is characterized by destruction of the β cells of the pancreas with resulting insulinopenia. Type 1 diabetes in children is often manifested acutely as diabetic ketoacidosis when insulin secretion falls below insulin needs. A condition termed the *honeymoon period* can develop if the diagnosis occurs during a time of increased insulin needs, such as during a viral illness. When the illness is resolved, insulin needs can fall below residual insulin production and normoglycemia results without the use of exogenous insulin. The honeymoon period rarely lasts more than 1 year and is usually shorter.

Goals of Therapy

Goals of therapy for children include achieving normal growth and development, avoiding acute and chronic complications of diabetes, addressing psychosocial issues, and educating children regarding self-care.

Acute Complications

Whenever children with type 1 diabetes mellitus experience hyperglycemia, the resulting glycosuria can precipitate dehydration. The threat of dehydration is especially severe during episodes of diabetic ketoacidosis. Supplemental fluids and insulin may be necessary during these times.

Diabetic ketoacidosis can occur when insulin administration is inadequate for needs. Diabetic ketoacidosis frequently accompanies the diagnosis of type 1 diabetes mellitus. Avoiding diabetic ketoacidosis involves knowledge of appropriate care during times of increased insulin need, such as during concomitant illness. To ensure early detection of incipient diabetic ketoacidosis, children and adolescents should be tested for ketonuria when the blood glucose concentration is greater than 240 mg/dL and during concomitant illness.

Hypoglycemia can be difficult to detect in very young children. Caregivers should be alert for behavioral changes such as lethargy, pallor, and sleep disturbances.

Chronic Complications

Chronic complications of diabetes are rarely manifested during adolescence. Screening for neuropathy and nephropathy and determinations of serum lipid levels should occur on a regular basis. Pregnant adolescent girls must be counseled on the importance of maintaining excellent metabolic control.

Treatment

Insulin requirements are approximately 1.0 unit/kg per day. An intensive regimen of at least three injections per day is recommended to prevent chronic complications. The administration of very small doses of insulin in infants and children may necessitate the use of a diluent.

Insulin needs increase during times of physiologic stress, such as during concomitant illness or puberty. Children who are inadequately treated with insulin will not grow or mature normally.

Children are usually able to begin administering insulin and performing capillary blood glucose monitoring with supervision when they are of school age. The age may vary with different children. When administering insulin, the abdomen is the least preferred site because of insufficient abdominal subcutaneous fat.

Diabetic teaching of such self-management skills as insulin injection, capillary blood glucose monitoring, and recognition and treatment of hypoglycemia should include all caregivers. Babysitters and teachers will need information on prevention, recognition, and management of hypoglycemia. All educational materials used with children should be age appropriate.

Children must be provided with a caloric intake adequate to meet needs for energy expenditure, growth, and maturation. Caloric intake is usually calculated as 1000 calories/day plus 100 added calories for each year until age 11 years. From ages 11 to 18 years, an additional 100 calories is added for girls and an additional 200 calories for boys. Growth should be plotted at each medical appointment to assess adequate nutrition and adequate insulinization. After the age of 2 years, recommended dietary guidelines for percentage of fat, carbohydrate, and protein intake are the same as those for adults.

Exercise is encouraged, with careful attention to adequate nutritional intake. Insulin doses may need to be adjusted to plan for unusual levels of activity, such as during long-distance bike riding or hiking.

Following a regimen designed to prevent acute and chronic complications of diabetes is difficult under the best of conditions. The goals of treatment are best accomplished when meals, medication, exercise, and blood glucose monitoring are consistent. Achieving consistency while also achieving developmental goals of separation and independence is very difficult. Peer pressure during adolescence can lead to poor adherence to therapeutic regimens. Concern regarding weight can lead to omission of insulin injections or the manifestation of eating disorders.

The child and family need support and counseling to develop effective strategies for achieving desired goals. Disturbed family functioning can have an impact on children and adolescents with diabetes and can lead to an increased frequency of hospitalization for diabetic ketoacidosis.

Genetic defects of the β cell are usually diagnosed in individuals younger than 25 years. These individuals are not likely to become ketotic. Management is identical to that of type 2 diabetes in young adults.

KEY POINTS

- Children and adolescents with diabetes overwhelmingly have type 1 diabetes mellitus (5% have genetic defects of the β cell). In type 1 diabetes, insulin is required at diagnosis or shortly thereafter.
- Goals of therapy for children include achieving normal growth and development, avoiding acute and chronic complications of diabetes, addressing psychosocial issues, and educating children regarding self-care.
- Acute complications of diabetes in children and adolescents include hyperglycemia leading to dehydration, possible diabetic ketoacidosis, and hypoglycemia. Hypoglycemia may manifest differently in children than in adults.
- Chronic complications in children and adolescents are rare. Screening for complications should nevertheless be initiated. Adolescent girls should be counseled on issues regarding diabetes and pregnancy.
- Insulin needs will vary according to growth stages, exercise, and concomitant illness. Regular capillary blood glucose monitoring is crucial for determining the efficacy of treatment. The age at which children can perform insulin measurement and administration, as well as capillary blood glucose monitoring independently, will vary. Nutritional needs are calculated at 1000 calories/day, with 100 added calories per year until age 11 and then an additional 100 calories for girls and an additional 200 calories for boys ages 11 to 18.
- Education in self-care activities should be appropriate for age and include other family members. Support both for the person with diabetes and for the family is important. Counseling may be helpful in some circumstances.

GERIATRIC CONSIDERATIONS

The prevalence of type 2 diabetes mellitus increases with age. Adults older than 60 years constitute nearly 50% of the diabetic population of the United States. The prevalence of diabetes in the elderly is almost 21% for individuals age 60 or older and more than doubles that of younger adults ages 40 to 59.

The increase in risk for type 2 diabetes in older adults is multifactorial. Aging often involves increased adiposity and a decrease in lean body mass and activity levels. The latter factors contribute to insulin resistance. Insulin secretion also diminishes with age. The risk of diabetes in the elderly is likewise increased by surgery, illness, and the use of such medications as steroids and diuretics.

Diabetes in the elderly can be difficult to diagnose because of fluctuating blood glucose values in response to food intake and activity and because of inconsistent or absent symptoms of hyperglycemia. Chronic complications of diabetes such as neuropathy and retinopathy are frequently present at diagnosis and indicate glucose intolerance of long duration.

Goals of Therapy

Goals of treatment for the elderly include prevention of acute complications, prevention and management of chronic complications, attention to psychosocial issues, and education regarding self-care.

Acute Complications

Uncontrolled hyperglycemia in the elderly may be asymptomatic or may produce such classic symptoms as polyuria and fatigue. Polydipsia is less common because of decreased thirst perception. Hyperglycemia can result in increased perception of pain and slowing of intellectual processes. The risk of infection is greater when blood glucose values are greater than 200 mg/dL. Such infectious disease processes as malignant otitis externa and reactivation of chronic tuberculosis are linked to hyperglycemia. Elderly individuals with diabetes are two times as likely to be hospitalized for kidney infections as elderly individuals without diabetes.

Chronic hyperglycemia can cause mild to moderate dehydration in the elderly that can be exacerbated by age-related changes in kidney function and water conservation. The resulting postural hypotension and electrolyte imbalances can increase the risk of falls.

Elderly people with type 2 diabetes mellitus are not prone to ketosis, but they are at risk for nonketotic hyperglycemic hyperosmolar coma. Particular risk factors include impaired thirst recognition, polypharmacy, dementia, and concurrent illness. Profound dehydration can occur and lead to a significant mortality rate for this complication (10% to 50%).

Older individuals with type 2 diabetes mellitus must be instructed in care during periods of concomitant illness and advised to perform capillary blood glucose monitoring on a regular basis to avoid undetected hyperglycemia.

Hypoglycemia can occur when an elderly individual with diabetes is treated with a sulfonylurea or insulin. Hypoglycemia can occur atypically with symptoms of lethargy or focal neurologic dysfunction. Elderly individuals may have age-related decreases in counterregulatory function or an inability to report hypoglycemic symptoms. Glycemic targets for these individuals may be higher. The risk of injury during a hypoglycemic episode warrants careful observation of blood glucose values and regular evaluation of treatment of all elderly individuals with diabetes.

Chronic Complications

The increased prevalence of heart and blood vessel disease, kidney disease, eye disease, and foot disease in patients with diabetes overlaps with the increased prevalence of these conditions in the general elderly population. Diabetes increases the incidence and severity of these diseases.

Aging-related changes can present a particular problem in the performance of diabetic foot care. Orthopedic deformity, loss of protective subcutaneous fat, and atherosclerotic changes are all common problems of the elderly. Inspection of the feet and nail care can be compromised by changes in visual acuity and joint function. Assistance with foot care is often necessary to minimize the risk of diabetic complications involving the feet.

Treatment

Oral antidiabetic agents must be chosen carefully to avoid age-related adverse effects. Oral agents with a shorter duration of action are preferable. Elderly individuals with diabetes should be evaluated for changes in renal and hepatic function. Oral agents metabolized in an impaired system should be avoided. Metformin is not appropriate for individuals with decreased liver and renal function because of the increased risk of lactic acidosis.

Insulin is safe to use with caution in the elderly. Thin elderly individuals can be highly sensitive to insulin and may require very small amounts to control hyperglycemia. Some elderly individuals are quite sensitive to regular insulin. Daily or twice-daily dosing of long-acting insulin may be preferable.

Age- or illness-related changes in vision, manual dexterity, and cognition can diminish the ability to measure and administer insulin. Magnification devices and the use of prefilled syringes can be of assistance.

Exercise is of benefit to individuals of all ages. Exercise plans must often be modified to account for orthopedic or other mobility problems. Armchair exercise can be an excellent way for elderly individuals to stay active.

Appropriate food intake can help control hyperglycemia and reduce the risk of chronic complications in the elderly, as in all individuals with diabetes. The quality and quantity of nutrients must be assessed carefully in the elderly. Elderly individuals may be obese, malnourished, or both. Increasing the nutritional value of a meal plan that is low in calories because of choice or a desire to lose weight is an important and difficult goal. Dental and other oral disease or dysfunction can have an impact on nutritional intake.

Education of the elderly in diabetic self-care practices can be challenged by age-related changes in vision, hearing, and cognition. Simple, clearly written educational materials and less complex therapeutic regimens can be of assistance. Caretakers and family members should be included in education sessions if possible. Family or other assistance can ensure safe performance of diabetic self-care activities while maintaining as much independence as possible.

KEY POINTS

- The increased prevalence of type 2 diabetes mellitus in the elderly is multifactorial and due to increased adiposity, decreased lean body mass, decreased activity levels, decreased insulin secretion, the hyperglycemic effect of certain medications, concurrent illness, and surgery. Varying blood glucose values can lead to difficulty in diagnosis.
- Goals of treatment for the elderly include prevention of acute complications, prevention and management of chronic complications, attention to psychosocial issues, and education regarding self-care.
- Acute complications of diabetes in the elderly include hyperglycemia, often asymptomatic, which can lead to dehydration; increased risk of infection; and nonketotic hyperglycemic hyperosmolar coma. Hypoglycemia can occur atypically and may lead to injury.
- Heart and blood vessel disease, foot problems, visual disabilities, and kidney disease have a significant presence in the aging population in general, as well as being chronic complications of diabetes. Avoiding foot problems can be particularly challenging given the frequent presence of orthopedic deformity and other common aging-related changes, as well as the decreased ability to perform appropriate foot care.
- Oral antidiabetic agents should be carefully chosen with consideration of renal and hepatic function. Short-acting agents are preferable. When insulin treatment is necessary, visual or orthopedic and other changes may hinder measurement of insulin. Adaptive devices can be helpful. Exercise should be encouraged and may have to be modified for people with limited mobility or other limiting factors. Meal planning for elderly individuals should emphasize appropriate amounts of foods with high nutritional value.
- Simple, clearly written educational material can be helpful for individuals with visual or cognitive impairments. Caretakers or family members should be included in education sessions if necessary.

SUMMARY

Diabetes mellitus, the most common endocrine disorder, affects millions of Americans. Diabetes is characterized and diagnosed by chronic hyperglycemia, the result of a relative or absolute deficiency of insulin; however, the metabolism of all energy nutrients is altered. Of the four clinical classes of diabetes, the most common are type 2 and type 1 diabetes. Type 2 diabetes is characterized by insulin resistance and a reduction in insulin production leading to a relative insulin deficiency. Type 1 diabetes is the result of destruction of the insulin-producing β cells of the pancreas because of an autoimmune or idiopathic process.

Sequelae of insulin deficiency include the acute and chronic complications of diabetes. Acute complications include diabetic ketoacidosis in type 1 diabetes and nonketotic hyperglycemic hyperosmolar coma in type 2 diabetes. Chronic complications include cardiovascular disease, retinopathy, nephropathy, and neuropathy.

The goals of treatment are glycemic control and prevention of complications. Treatment is individualized and encompasses an individualized diet, regular exercise, and appropriate use of medications such as oral antidiabetic agents, incretin and amylin mimetics, and insulin. The efficacy of treatment and the presence of complications of therapy are evaluated by capillary blood glucose monitoring. Patient education is an essential component in teaching skills associated with treatment.

Special considerations attend the treatment and education of individuals with diabetes in the pediatric and geriatric age groups. Children and adolescents require careful monitoring to adjust insulin levels for variations in maturation, exercise, and concomitant illness. The elderly may have chronic complications of diabetes or other impairments of mobility, vision, or cognition that affect treatment.

Educational materials must be appropriate for age in children and be accessible for elderly individuals with visual or cognitive impairments.

RESOURCES

Epidemiology of Diabetes

American Diabetes Association: Economic costs of diabetes in the U.S. in 2012. *Diabetes Care* 36:1–20, 2013.

American Diabetes Association: *Staggering cost of diabetes in America*. http://www.diabetes.org/diabetes-basics/statistics/infographics/adv-staggering-cost-of-diabetes.html. Retrieved May 28, 2016.

Centers for Disease Control and Prevention: *Diabetes report card, 2014*, Atlanta, 2015, U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.

Wild S, Roglic G, Green A, et al: Global prevalence of diabetes: estimates for the year 2000 and projections for the year 2030. *Diabetes Care* 27:1047–1053, 2004.

Physiology of Glucose Regulation

Augustin R: The protein family of glucose transport facilitators: it's not only about glucose after all. *IUBMB Life* 62(5):315–333, 2010. doi:10.1002/iub.315.

D'Alessio D: The role of dysregulated glucagon secretion in type 2 diabetes. *Diabetes Obes Metab* 13:126–132, 2011.

Drucker DJ: The biology of incretin hormones. *Cell Metab* 3:153–165, 2006.

Eisenbaeth GS, Polonsky KS, Buse JB: Diabetes mellitus. In Wilson RH, et al, editors: *Williams textbook of endocrinology*, ed 12, Philadelphia, 2012, Saunders.

Guillo C, Roper MG: Simultaneous capillary electrophoresis competitive immunoassay for insulin, glucagon, and islet amyloid polypeptide secretion from mouse islets of Langerhans. *J Chromatogr A* 1218(26):4059–4064, 2011.

Hartter E, Svoboda T, Ludvik B, et al: Basal and stimulated plasma levels of pancreatic amylin indicate its co-secretion with insulin in humans. *Diabetologia* 34:52–54, 1991.

Kovalaske M, Gandhi GY: Glycemic control in the medical intensive care unit. *J Diabetes Sci Technol* 3(6):1330–1341, 2009.

McCulloch LJ, et al: GLUT2 (SLC2A2) is not the principal glucose transporter in human pancreatic beta cells: implications for understanding genetic association signals at this locus. *Mol Genet Metab* 104(4):648–653, 2011. doi:10.1016/j.ymgme.2011.08.026.

Mourad FH, Saadé NE: Neural regulation of intestinal nutrient absorption. *Prog Neurobiol* 95(2):149–162, 2011.

Plum L, Belgardt BF, Brüning JC: Central insulin action in energy and glucose homeostasis. *J Clin Invest* 116(7):1761–1766, 2006.

Spiegel A, Carter-Su C, Taylor SI, et al: Mechanism of action of hormones that act at the cell surface. In Wilson RH, et al, editors: *Williams textbook of endocrinology*, ed 12, Philadelphia, 2012, Saunders.

Thorens B: Brain glucose sensing and neural regulation of insulin and glucagon secretion. *Diabetes Obes Metab* 13:82–88, 2011.

Thorens B, Mueckler M: Glucose transporters in the 21st century. *Am J Physiol Endocrinol Metab* 298(2):E141–E145, 2010. doi:10.1152/ajpendo.00712.2009.

Trevaskis JL, Parkes DG, Roth JD: Insights into amylin–leptin synergy. *Trends Endocrinol Metab* 21(8):473–479, 2010.

Pathophysiology of Diabetes Mellitus

American Diabetes Association: Classification and diagnosis of diabetes. *Diabetes Care* 38(Suppl 1):S8–S16, 2012.

Dib S, Gomes M: Etiopathogenesis of type 1 diabetes mellitus: prognostic factors for the evolution of residual beta cell function. *Diabetol Metab Syndr* 1(1):25, 2009.

Eisenbarth GS, Jeffrey J: The natural history of type 1A diabetes. *Arq Bras Endocrinol Metabol* 52:146–155, 2008.

Esienbarth SC, Homann D: *Primer: immunology and autoimmunity type 1 diabetes: molecular, cellular, and clinical immunology*, ed 3, Denver, 2011, Barbara Favis Center for Diabetes.

Ferrannini E, Gastaldelli A: Lozzo P: Pathophysiology of prediabetes. *Med Clin North Am* 95(2):327–339, 2011.

Green A: Descriptive epidemiology of type 1 diabetes in youth: incidence, mortality, prevalence, and secular trends. *Endocr Res* 33(1-2):1–15, 2008.

International Diabetes Federation: *Diabetes atlas*, ed 7, Brussels, 2015.

Nambam B, Aggarwal S, Jain A: Latent autoimmune diabetes in adults: a distinct but heterogeneous clinical entity. *World J Diabetes* 1(4):111–115, 2010.

Nolan CJ, Damm P, Prentki M: Type 2 diabetes across generations: from pathophysiology to prevention and management. *Lancet* 378(9786):169–181, 2011.

Rewers M, Norris J, Kretowski A: *Epidemiology of type 1 diabetes. Type 1 diabetes: cellular, molecular & clinical immunology*, ed 3, London, 2010, Oxford University Press. Retrieved from: www.ucdenver.edu/academics/colleges/medicalschool/centers/BarbaraDavis/OnlineBooks/Pages/Type1Diabetes.aspx.

Rolandsson O, Palmer JP: Latent autoimmune diabetes in adults (LADA) is dead: long live autoimmune diabetes. *Diabetologia* 53(7):1250–1253, 2010.

Stumvoll M, Goldstein BJ, van Haeften TW: Type 2 diabetes: principles of pathogenesis and therapy. *Lancet* 365(9467):1333–1346, 2005.

Diabetes Management and Outcomes

American Diabetes Association: Implications of the diabetes control and complications trial. *Diabetes Care* 26:25–27, 2003.

American Diabetes Association: Implications of the United Kingdom prospective diabetes study. *Diabetes Care* 26:28–32, 2003.

American Diabetes Association: Tests of glycemia in diabetes. *Diabetes Care* 26:106–108, 2003.
American Diabetes Association: Nephropathy in diabetes. *Diabetes Care* 27:S79–S83, 2004.
American Diabetes Association: National standards for diabetes self-management education. *Diabetes Care* 37(Suppl 1):S144–S153, 2014.
American Diabetes Association: Foundations of care: Education, nutrition, physical activity, smoking cessation, psychosocial care, and immunizations. *Diabetes Care* 38(Suppl 1):S20–S30, 2015.
American Diabetes Association: Glycemic targets. A position statement of the American Diabetes Association. *Diabetes Care* 38(Suppl 1):S22–S40, 2015.
American Diabetes Association: Standards of medical care in diabetes—2016; abridged for primary care providers. *Clinical Diabetes* 39(1):3–12, 2016.
American Diabetes Association: Standards of medical care in diabetes—2016. *Diabetes Care* 39(Suppl 1):S1–S112, 2016.
Bhartia M, Tahrani AA, Barnett AH: SGLT-2 inhibitors in development for Type 2 diaebtes. *Rev Diabet Stud* 8:348–354, 2011.
Buse JB, Polonsky KS, Purant CF: Type 2 diabetes mellitus. In Wilson RH, et al, editors: *Williams textbook of endocrinology*, ed 11, Philadelphia, 2008, Saunders.
Cahn A, Miccoli R, Dardano A, Del Prato S: New forms of insulin and insulin therapies for the treatment of type 2 diabetes. *Lancet Diabetes Endocrinol* 3(8):638–652, 2015.
Cheng YJ, Gregg EW, Kahn HS, et al: Peripheral insensate neuropathy—a tall problem for U.S. adults? *Am J Epidemiol* 164:873–880, 2006.
Coustand DR: Pharmacological management of gestational diabetes: an overview. *Diabetes Care* 30(S2):S206–S208, 2007.
Crowther CA, Hiller JE, Moss JR, et al: Effect of treatment of gestational diabetes mellitus on pregnancy outcomes. *N Engl J Med* 352:2477–2486, 2005.
Drucker DJ, Nauck MA: The incretin system: glucagon-like peptide-1 receptor agonists and dipeptidyl peptidase-4 inhibitors in type 2 diabetes. *Lancet* 368(9548):1696–1705, 2006.
Goldstein DE, Little RR, Lorenz RA, et al: Tests of glycemia in diabetes. *Diabetes Care* 27:1761–1773, 2004.
Gregg EW, Sorlie P, Paulose-Ram R, et al: Prevalence of lower-extremity disease in the U.S. adult population 40 years of age with and without diabetes. *Diabetes Care* 27:1591–1597, 2004.
Jacobson AM: The psychological care of patients with insulin-dependent diabetes mellitus. *N Engl J Med* 334:1249–1253, 1996.
Jenkins NT, Hagberg JM: Aerobic training effects on glucose tolerance in prediabetic and normoglycemic humans. *Med Sci Sports Exerc* 43(12):2231–2240, 2011.
Jones MC: Therapies for diabetes: pramlintide and exenatide. *Am Fam Physician* 75:1831–1835, 2007.
Kelly TN, Bazzano LA, Fonseca VA, et al: Systematic review: glucose control and cardiovascular disease in type 2 diabetes. *Ann Intern Med* 151(6):394–403, 2009.
Kim C, Berger DK, Chamany S: Recurrence of gestational diabetes mellitus: a system review. *Diabetes Care* 30:1314–1319, 2007.
Kim C, Newton KM, Knopp RH: Gestational diabetes and the incidence of type 2 diabetes—a systematic review. *Diabetes Care* 25:1862–1868, 2002.
Kitabchi AE, Umpierrez GE, Murphy MB, et al: Hyperglycemic crisis in adult patients with diabetes. *Diabetes Care* 29:2739–2748, 2006.
Kwik M, Seeho SKM, Smith C, et al: Outcomes of pregnancies affected by impaired glucose tolerance. *Diabetes Res Clin Pract* 77:263–268, 2007.
Kyriakos K: Incretin effect: GLP-1, GIP, DPP4. *Diabetes Res Clin Pract* 93(Suppl 1[0]):S32–S36, 2011.
Langley AK, Suffoletta TJ, Jennings HR: Dipeptidyl peptidase IV inhibitors and the incretin system in type 2 diabetes mellitus. *Pharmacotherapy* 27(8):1163–1180, 2007.
Lenters-Westra E, Slingerland R: Hemoglobin A1c determination in the A1C-derived average glucose (ADAG) study. *Clin Chem Lab Med* 46(11):1617–1623, 2008.
Linmans J, Spigt M, Deneer L, et al: Effect of lifestyle intervention for people with diabetes or prediabetes in real-world primary care: propensity score analysis. *BMC Fam Pract* 12(1):95, 2011.
MacIsaac RJ, Jerums G: Intensive glucose control and cardiovascular outcomes in type 2 diabetes. *Heart Lung Circ* 20(10):647–654, 2011.
Malin SK, Gerber R, Chipkin SR, Braun B: Independent and combined effects of exercise training and metformin on insulin sensitivity in individuals with prediabetes. *Diabetes Care* 35(1):131–136, 2012.
Mannucci E, Monami M, Lamanna C, et al: Prevention of cardiovascular disease through glycemic control in type 2 diabetes: a meta-analysis of randomized clinical trials. *Nutr Metab Cardiovasc Dis* 19(9):604–612, 2009.
National Institute of Diabetes and Kidney Disease: *Monogenic forms of diabetes: neonatal diabetes mellitus and maturity onset diabetes in the young*. NIH Publication No 07-6141; 2007.
Neumiller JJ: Differential chemistry (structure), mechanism of action, and pharmacology of GLP-1 receptor agonists and DPP-4 inhibitors. *J Am Pharm Assoc* 49(5 Suppl 1):S16–S29, 2009.
Nikoulina SE, Ciaraldi TP, Carter L, et al: Impaired muscle glycogen synthase in type 2 diabetes is associated with diminished phosphatidylinositol 3-kinase activation. *J Clin Endocrinol Metab* 86:4307–4314, 2001.
Rustad JK, Musselman DL, Nemeroff CB: The relationship of depression and diabetes: pathophysiological and treatment implications. *Psychoneuroendocrinology* 36(9):1276–1286, 2011. doi:10.1016/j.psyneuen.2011.03.005.
Senior PA: Type 2 diabetes, metformin and lactic acidosis—defining the risk and promoting safe practice. *Diabet Med* 29(2):161–163, 2012.
Shubrook JH, Bokaie BB, Adkins S: Empagliflozin in the treatement of type 2 diabetes: evidence to date. *Drug Des Devel Ther* 9:5793–5803, 2015.
Stacul F, van der Molen A, Reimer P, et al: Contrast induced nephropathy: updated ESUR Contrast Media Safety Committee guidelines. *Eur Radiol* 21(12):2527–2541, 2011.
Syed I: Glycated haemoglobin; past, present, and future: are we ready for the change? *J Pak Med Assoc* 61(4):383–388, 2011.
Tesfaye S, Chaturvedi N, Eaton SE, et al: Vascular risk factors and diabetic neuropathy. *N Engl J Med* 352:341–350, 2005.
The ACCORD Study Group: Long-term effects of intensive glucose lowering on cardiovascular outcomes. *N Engl J Med* 364(9):818–828, 2011.
Turnbull F, Abraira C, Anderson R, et al: Intensive glucose control and macrovascular outcomes in type 2 diabetes. *Diabetologia* 52(11):2288–2298, 2009.
Vijan S, Hayward RA: Treatment of hypertension in type 2 diabetes mellitus: blood pressure goals, choice of agents, and setting priorities in diabetic care. *Ann Intern Med* 138:593–602, 2003.

42

Nutritional and Metabolic Disorders

Brent A. Banasik

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- What are the primary nutrients and how are they metabolized by cells?
- What is basal metabolic rate and how is it affected by age, body type, and illness?
- How do genetics, epigenetics, and environment contribute to altered metabolism, metabolic syndrome, and obesity?
- What hormones are involved in the control of appetite and nutrient metabolism?
- How do starvation, physiologic stress, aging, and altered health states affect body metabolism?

CHAPTER OUTLINE

Metabolic Processes, 838
Anabolism and Catabolism, 839
Metabolic Rate, 839
Nutrient Metabolism, 839
Carbohydrates, 839
Lipids, 840
Proteins, 841
Regulation of Appetite and Nutrient Metabolism, 841
Role of Genetics, Epigenetics, and Environment, 841
Hormonal Regulation of Nutrient Intake and Appetite, 842
Hormonal Regulation of Nutrient Storage, Distribution, and Metabolism, 843
Insulin, 843
Glucagon, 843
Catecholamines, 843
Growth Hormone, 843
Cortisol, 843
Obesity and Metabolic Syndrome, 844
Obesity, 844
Metabolic Syndrome, 844
Metabolic Responses to Starvation and Physiologic Stress, 844
Starvation and Protein-Energy Malnutrition, 845
Physiologic Stress, 845
Nutritional Considerations for Aging and Altered Health States, 846
Aging, 846
Infection, Sepsis, and Fever, 846
Surgery, 848
Trauma, 848
Burns, 848
Cancer, 848
Immobility, 848

The human body is maintained by a complex interconnected network of biochemical reactions that control energy balance and synthesis of biomolecules to meet ever-changing physiologic needs. To fuel this metabolic system, adequate supplies of digestible foodstuffs with the appropriate organic chemical structure must be acquired on a regular basis. Adequate nutrition is needed for growth and metabolism, organ function, tissue repair, and response to infection. Excesses and deficits of certain nutrients, especially in conjunction with inactivity, can result in muscle wasting, excess adiposity, and glucose intolerance. Severely ill patients are at particular risk for morbidity and mortality from inadequate nutrition and altered metabolic function while hospitalized. This chapter reviews normal nutrient metabolism, neurohormonal regulation of food intake, and common disorders of nutrition and metabolism. Diabetes is discussed in Chapter 41.

METABOLIC PROCESSES

Metabolism is a dynamic phenomenon in biological systems involving the physical and chemical processes that produce and maintain biomolecules (anabolic) and transform molecules into energy (catabolic). Most metabolic pathways involve numerous sequential enzymatic steps and are organized together so that intermediates are transferred quickly to the next enzyme-active site to improve efficiency. The details of metabolic enzyme pathways can be found in

biochemistry texts. General concepts of anabolism and catabolism are reviewed next.

Anabolism and Catabolism

Anabolism refers to the constructive phase of metabolism and involves the synthesis of organic molecules by cells. More complex or larger molecules are built from simple ones, and in the process energy is consumed. Anabolism occurs continuously along with catabolism, but is more prominent during times of rest, healing, pregnancy, lactation, and growth. Hormones such as insulin and sex hormones may also trigger anabolism. Conversely, catabolism is the degradative phase of metabolism. Complex molecules are broken down into simpler substances, often with the concurrent release or production of chemical energy. During times of disease, stress, fever, or starvation or during the release of certain hormones such as thyroid hormone and cortisol, catabolism dominates the body's metabolic processes. The resultant tissue wasting may lead to cellular injury or death if excessive catabolism is left unregulated. Anabolism and catabolism occur simultaneously and are most efficient in the presence of molecular oxygen; therefore ischemia or hypoxia is associated with significant metabolic dysfunction that can lead to cellular injury or death (see Chapter 4).

The metabolic process requires nutrients in the form of carbohydrates, lipids, and proteins. Each of these three nutrients is an organic, carbon-containing substance that can be used to meet cellular energy requirements. Energy extracted by metabolic processes is used to create the energy currency of the body known as *adenosine triphosphate* (ATP) (see Chapter 3). The high-energy phosphate bonds in ATP are hydrolyzed to power the metabolic processes of the cell.

Energy can be measured in kilocalories (kcal); 1 kcal represents the amount of energy required to raise the temperature of 1 kg of water from 15°C to 16°C. Each gram of carbohydrate or protein provides 4 kcal of energy, whereas a gram of fat provides 9 kcal of energy. For comparison, a medium-sized baked potato without butter is about 200 kcal, whereas a glass of wine is about 100 kcal and a lean filet mignon (8 oz, 227 g) is about 400 kcal. The efficiency of nutrient energy capture from catabolism is approximately 40% with the remaining 60% producing heat. The energy released as heat is important for maintaining body temperature.

Metabolic Rate

The basal metabolic rate (BMR) refers to the rate of energy consumption at rest. It represents the energy used in maintaining basic body processes such as respiration, cellular metabolism, circulation, glandular activity, and the maintenance of body temperature. The body's BMR is determined by calculating oxygen use during a specific period. The normal range for BMR is generally between 0.8 and 1.43 kcal/min. Several factors that affect an individual's BMR are described in Table 42.1. Body stature and size affect BMR by the amount of heat lost from the body surface. Age is also an important determinant of BMR. A growing child's BMR is significantly higher than an adult's, primarily because of an increased rate of cellular activity, surface-to-volume ratio, and generation of new tissue. Conversely, as one ages, the BMR gradually declines by about 2% per decade. Body composition, determined by the amount of fat and lean tissue, also affects BMR. Women typically have a metabolic rate 5% to 10% less than that of men because of differences in body mass and composition. Pregnancy increases the BMR by about 20% to 28%, or 300 kcal/day, as a consequence of increased uterine and mammary gland size, fetal development, and additional cardiopulmonary workload. Other factors affecting BMR include nutritional status, sleep, fever, environmental temperature, and stress. Metabolic thermoregulation in newborns is described in the Pediatrics Consideration box.

Almost any alteration in the body's normal homeostatic state will alter its energy requirements and BMR. Many diseases are known to dramatically increase the body's energy requirements, including chronic obstructive and restrictive pulmonary diseases, fever, burns, cancer, diabetes, and hyperthyroidism.

TABLE 42.1 Examples of Factors Affecting Basal Metabolic Rate

Increasing Metabolism	Decreasing Metabolism
Childhood growth	Aging process
Exercise	End-stage illness
Sympathetic stimulation	Starvation
Shivering	Sleep
Fever	Tropical climates
Thyroid hormone	Calorie-restricted diets
Muscle mass	Reduced muscle mass
Pregnancy	
Stress	
Male sex hormone	

KEY POINTS

- Anabolism refers to energy-requiring processes involving synthesis of biomolecules. Catabolism refers to energy-producing processes during which biomolecules are broken down into simpler forms. Metabolism refers to the dynamic state of simultaneously occurring anabolism and catabolism.
- The basal metabolic rate (BMR) is the rate of energy utilization when the body is at rest. Examples of factors affecting the BMR include body size and composition, age, nutritional status, muscle mass, fever, stress, and pregnancy.

NUTRIENT METABOLISM

The primary nutrients for metabolism are complex carbohydrates, lipids, and proteins that are acquired through dietary intake. Complex nutrients must be enzymatically digested into small molecules for absorption through the gastrointestinal tract (see Chapter 35). These small molecules are metabolized to provide energy and building blocks for biosynthesis. The basic concepts for metabolism of carbohydrates, lipids, and proteins are presented next.

Carbohydrates

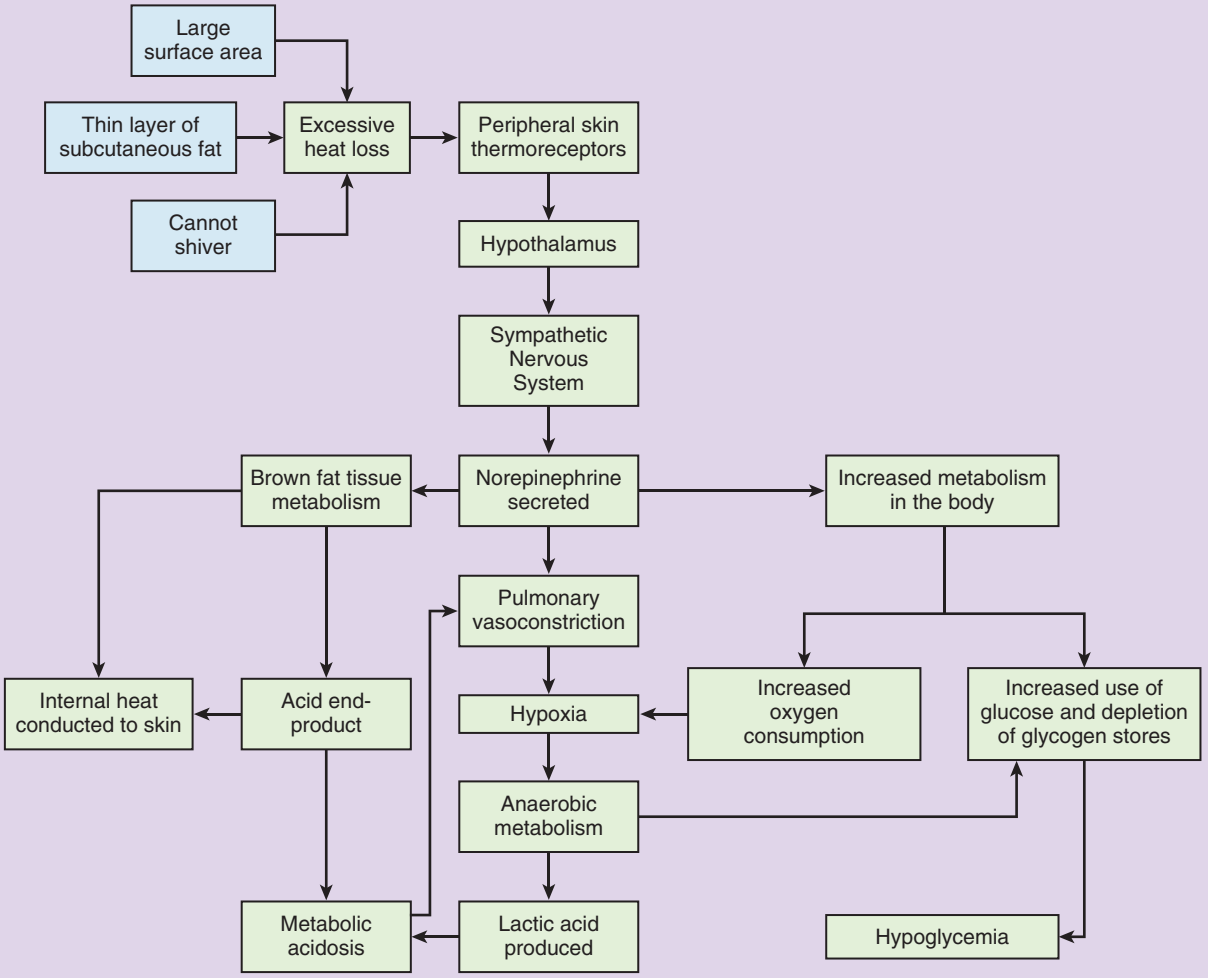
Carbohydrates are a major energy source for the body, providing approximately half of the dietary caloric intake. Dietary carbohydrates are classified into the three categories of monosaccharides (simple sugars), oligosaccharides (2 to 10 joined monosaccharide units), and polysaccharides (10 to 10,000 monosaccharide units). Each gram of carbohydrate provides 4 kcal of energy.

Dietary monosaccharides are the six-carbon sugars of glucose, mannose, fructose, and galactose. Glucose, the most physiologically important of the group, is the form of sugar normally found in the bloodstream. Glucose is derived from the catabolism of more complex carbohydrates during the process of digestion. Fructose and galactose are also eventually converted to glucose by the liver. Once in the bloodstream, glucose is transported throughout the body where it is converted to cellular energy (see Chapter 3), used to synthesize fatty acids, or stored in the liver and muscles as glycogen.

Gluconeogenesis refers to the process by which glucose is formed from noncarbohydrate sources, including amino acids or lactic acid supplied by muscle tissue and glycerol supplied from fat breakdown. The glucose made through this mechanism may be either stored in the liver as glycogen or released into the bloodstream. During periods of fasting, gluconeogenesis and glycogenolysis (glycogen breakdown) provide

PEDIATRIC CONSIDERATIONS

Thermoregulation in Newborns



Newborns are more susceptible to excessive heat loss because of several factors. Newborns have a large surface area of skin relative to their weight; they are born with a thin layer of subcutaneous fat, and they cannot shiver. These factors can lead to excessive heat loss if the newborn is not kept warm by other means, such as appropriate clothing.

When heat is lost, the peripheral skin thermoreceptors sense the heat change and signal the hypothalamus. The hypothalamus is responsible for coordinating temperature control in the body. The sympathetic nervous system is signaled by the hypothalamus to secrete norepinephrine. Norepinephrine increases the body's metabolic rate. Metabolism first occurs in the brown fat tissue of the newborn. Brown fat, or adipose, tissue is extremely vascular, has increased glycogen stores, and has more mitochondria within the tissue. As blood passes through the brown fat tissue it absorbs heat from the metabolic process. The heat is then conducted through the blood to the skin, warming the newborn. Brown fat tissue metabolism creates acidic end-products. Norepinephrine also increases general metabolism in the body, increasing consumption of oxygen and glucose and glycogen stores. In addition, norepinephrine secretion results in pulmonary vasoconstriction that can contribute to hypoxia. The lack of oxygen forces the body to use anaerobic metabolism to create energy. Anaerobic metabolism depletes glycogen stores faster than aerobic metabolism, contributing to hypoglycemia. Anaerobic metabolism also produces lactic acid, which, combined with the acidic end-products from brown fat metabolism, leads to metabolic acidosis. Metabolic acidosis in turn contributes to pulmonary vasoconstriction, which further propagates the cycle. It is critical to keep the newborn warm to minimize cold stress because the newborn is ill equipped for a significant stress response.

the necessary glucose to meet the metabolic requirements of the brain and other glucose-dependent tissues.

Lipids

Lipids or fats are the most concentrated form of energy. Fats supply 9 kcal of energy per gram. The average fatty acid can produce 146 ATP molecules through oxidative phosphorylation. The majority of dietary lipids are triglycerides and cholesterol. Triglycerides must be digested in the gastrointestinal tract into fatty acids and glycerol before they can be absorbed. The hydrocarbon tails of fatty acids contain 12 to 22 carbon atoms and may be *saturated* or *unsaturated*. The degree of hydrogen saturation is related to the number of double bonds between the carbon atoms in the chain. If a fatty acid chain contains all the hydrogen atoms possible with no double bonds, it is called a saturated fatty acid. Those fatty acids with one double bond are termed monounsaturated, and those with several double bonds are called polyunsaturated. Dietary recommendations suggest limiting the amount of saturated fats consumed because of associations with cardiovascular disease risk.

Whereas the human body is able to store approximately 24,000 kcal of protein and 800 kcal of carbohydrate as glycogen, it has essentially unlimited storage capacity for fats. Carbohydrates and amino acids not immediately used by the tissues are converted to fat and stored, along with ingested fat, as adipose tissue. During times of fasting, the body can retrieve stored fats to use for its energy needs. Tissues in the body, with the exception of brain cells, can effectively metabolize and use fatty acids as an energy source.

Dietary fats are synthesized into lipoproteins called *chylomicrons*, which consist predominantly of triglycerides. Chylomicrons are taken up by the liver and resynthesized into a variety of fats and lipoproteins. When released into the circulation, the fatty acids are quickly assimilated into tissue. Fatty acids are catabolized in tissues through a reaction known as *β oxidation* that removes two carbon units from the fatty acid chain and transfers it to coenzyme A (CoA) to form acetyl CoA. Acetyl CoA can be used to make ketone bodies, or it may enter the mitochondria and proceed through oxidative phosphorylation (see Chapter 3). During prolonged fasting, the ketone bodies can traverse the blood–brain barrier and provide energy to the brain. However, excessive acute ketogenesis is harmful and produces a condition known as *ketoacidosis*, a common complication of type 1 diabetes (see Chapter 41).

The liver is the major organ responsible for lipid metabolism and regulation of serum lipid levels. Increased mobilization of fatty acids from adipose tissue to the liver occurs in certain conditions, such as metabolic syndrome, diabetes mellitus, starvation, obesity, and critical illness.

Proteins

Proteins are composed of nitrogen, carbon, hydrogen, oxygen, and, occasionally, sulfur. When hydrolyzed, they yield amino acids. A total of 22 amino acids have been identified in protein, 8 of which are essential—meaning that they must be supplied through the diet. These include phenylalanine, valine, threonine, tryptophan, isoleucine, methionine, leucine, and lysine. The distinction between essential and nonessential amino acids is somewhat unclear, because some amino acids can be produced from others via transamination. Muscle tissue, skin, and hair are composed primarily of protein. Proteins also serve as the major enzymatic catalysts. Children, because of their rapid growth, require more protein per kilogram of body weight than adults. In addition, compared with adults, children need a larger percentage of their dietary intake of protein to contain essential amino acids.

Once ingested, proteins are broken down into amino acids or peptides and are absorbed through the intestinal lumen. They are then carried to the liver through the portal vein. The liver regulates protein metabolism through enzymatic breakdown of amino acids; formation of nonessential amino acids from simple precursors; and detoxification or elimination of ammonia, urea, uric acid, and other catabolic end products. Amino acids supplied in excess of metabolic requirements are degraded to by-products such as urea, uric acid, or creatinine, and the remaining carbon skeletons are converted to carbohydrate and fat or oxidized for energy. Of particular importance is the conversion of amino acids to ketoacids, which are carbohydrate-like in structure and created by removal of the amino group during deamination. These ketoacids may then enter the tricarboxylic acid (TCA) cycle, where they provide energy for liver metabolism, or they may be converted to fatty acids by the liver.

Protein metabolism can be measured in terms of nitrogen balance. If nitrogen (protein) intake approximates output, an equal nitrogen balance exists. If dietary intake of proteins exceeds output, a positive nitrogen balance occurs. Protein anabolism exceeds catabolism during periods of rapid growth, pregnancy, and the formation of new tissue. A negative nitrogen balance occurs when protein breakdown exceeds daily protein intake and synthesis. If the daily caloric intake is insufficient, the body catabolizes dietary and tissue protein for energy, as is the case after severe burns and during fever, illness, or stress.

KEY POINTS

- Metabolism of carbon-containing organic foodstuffs, in the form of carbohydrate, lipid, and protein, supplies energy to support the cell's energy-requiring processes and provides building blocks for the synthesis of cellular biomolecules.
- The main sources of cellular energy are glucose and fatty acids. Glucose is the primary energy source for the brain, although the brain can use ketone bodies. Ketone bodies are produced from fatty acids by the liver, particularly under conditions of decreased carbohydrate intake or fasting.
- Positive nitrogen balance occurs when dietary intake of protein exceeds output, and a negative nitrogen balance occurs when protein breakdown exceeds daily protein intake and synthesis. Proteins can be converted to glucose through the process of gluconeogenesis.

REGULATION OF APPETITE AND NUTRIENT METABOLISM

Genetics, epigenetics, culture, environment, and lifestyle interact in complex ways to regulate appetite, food intake, metabolism, and body composition. Hunger is a potent survival stimulus that drives an individual to find and consume food. Through much of human history, daily activity revolved around the quest to find, eat, and store enough food to ensure survival. In developed nations many now find abundant calorie-dense food to be readily available with little energy expenditure required to obtain it. Excessive intake paired with sedentary lifestyle has created an epidemic of obesity for these populations, whereas in many poor nations the struggle to find adequate nutrition continues and protein-calorie malnutrition is the norm. Even in developed nations, protein-calorie malnutrition is prevalent among the poor and the elderly. Although seemingly counterintuitive, many of the obese suffer from protein-energy malnutrition because of excessive consumption of inexpensive carbohydrates; they are unable to afford protein sources of food.

In modern society, the intake of food frequently is stimulated by situations not related to hunger. Stress, boredom, sleep deprivation, socialization, time of day, and the sight, smell, and taste of food can prompt eating, even when the individual has no sensation of hunger. These frequent cues to overeat may disrupt the normal physiologic controls that tend to match energy intake with energy expenditure, leading to obesity. Obesity-related conditions are a growing public health concern.

Role of Genetics, Epigenetics, and Environment

Genetic and epigenetic mechanisms shape metabolic activity and can contribute to pathophysiologic states. Genes establish the template for synthesis of cellular proteins and are inherited in predictable ways from parent to offspring (see Chapters 5 and 6). About 25% of the variance in body fat can be attributed to genetics. Epigenetic mechanisms are more variable and heavily influenced by environment, especially exposures during fetal development. Epigenetics involves several modifications of nuclear chromatin (DNA and histones), as well as posttranscriptional RNA-based gene regulation. These modifications can alter the activity of genes, thereby changing function even though the base sequence of the DNA is not altered. For example, during fetal development, it has been shown that pregnant mothers can transfer a particular phenotype

to their offspring (fetal programming). This fetal imprinting is the result of maternal metabolism, which is indirectly linked to maternal diet. Both maternal obesity and maternal malnutrition have been linked to the development of obesity in offspring. Maternal malnutrition during pregnancy may activate an epigenetic metabolic program of "thriftiness" in the fetus favoring excessive nutrient storage, slower basal metabolism, and a higher risk of obesity.

Besides the general effects of maternal diet on subsequent metabolic dysfunction in the offspring, specific nutritional deficiencies or excesses can also alter the epigenome. Dietary sources of methylating agents, such as bioavailable folic acid, methionine, choline, betaine, and homocysteine, may have persistent effects on the epigenetic methylation patterns of certain DNA sequences. Because all of a woman's eggs develop during her time in the womb, the effects of nutritional deficiencies in her mother may be epigenetically transmitted to grandchildren, even though the woman herself maintained normal nutrition during her pregnancies.

Epigenetic changes in metabolism can arise within a single generation and remain fixed there; alternatively, as in the case of maintenance methylation, the changes can be preserved and inherited. The role of environment is profoundly important in determining these epigenetic changes, particularly when exposure occurs during fetal development. The increasing rates of obesity in developed nations over the past several decades may be, in part, because of inherited epigenetic changes in metabolic efficiency.

Hormonal Regulation of Nutrient Intake and Appetite

A number of hormones are known to affect appetite, nutrient intake, and energy expenditure. Some of these are produced by the gastrointestinal tract (i.e., ghrelin) and some by adipose (fat) tissue (i.e., leptin). Appetite and hunger are regulated by specific nuclei within the hypothalamus. These neurons orchestrate the initiation of feeding behavior and the cessation of eating when full or satiated (Fig. 42.1). The hypothalamic nuclei are in turn influenced by a number of neuronal and hormonal inputs (Table 42.2). One of these is the hormone ghrelin produced by endocrine cells in the stomach and intestine. Ghrelin is released into the bloodstream when the stomach is empty and when blood glucose levels fall. Ghrelin potently stimulates eating behavior and tissue resistance to the effects of insulin, resulting in a rise in blood glucose levels. Excessive

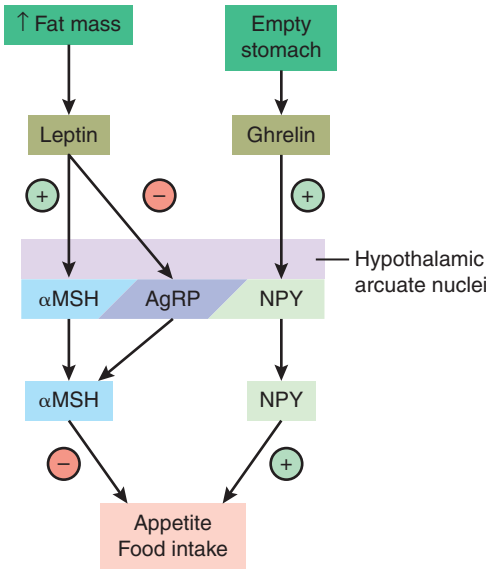


FIG 42.1 Hypothalamic control of eating behavior.

TABLE 42.2 Adipose and Stomach-Derived Hormones Involved in Appetite and Metabolism

Hormone/ Cytokine	Cell of Origin	Stimulus for Secretion	Primary Target	Actions
Leptin	Adipocyte	Increased adiposity	Hypothalamus	Decreases appetite Increases energy expenditure in adipose and nonadipose tissue Improves insulin sensitivity Allows for reproductive maturity
Adiponectin	Adipocyte	Weight loss (including surgical)	Muscle Liver Blood vessels and heart Macrophages	Activates oxidation of free fatty acids Antiinflammatory/antioxidant Improves insulin sensitivity Improves health of cardiovascular tissues and protects against cell death
Tumor necrosis factor-α	Adipocyte White adipose tissue Macrophage	Engorgement of adipocytes	Liver Muscle Adipocytes Other organs	Reduces adipocyte mass Proinflammatory Opposes insulin signaling Increases insulin resistance Atherogenic
Interleukin-6	White adipose tissue Macrophage	Other inflammatory cytokines	Liver Muscle Adipocytes Other organs	Opposes insulin signaling Increases insulin resistance Increases acute-phase protein production by liver Proinflammatory systemically
Ghrelin	P/D1 cells of stomach ε cells of islets	Empty stomach	Pituitary gland Hypothalamus	Increases growth hormone secretion Opposes glucoregulatory and liporegulatory actions of insulin Increases insulin resistance Increases appetite

From White BA, Porterfield SP: *Endocrine and reproductive physiology,* ed 4, Philadelphia, 2013, Mosby, p 68.

ghrelin signaling can lead to overeating; the disruption of ghrelin cells during procedures such as gastric bypass is thought to help patients reduce food intake postoperatively.

Leptin is the main hormonal signal indicating that adequate fat stores are present. Leptin is produced by adipose cells so that its level in the bloodstream reflects the relative fat mass of the body. Leptin thus provides a regulatory signal to the hypothalamus about the body's ability to support reproduction. When fat mass is significantly low, reproduction is inhibited. Another effect of leptin on the hypothalamus is to suppress appetite and reduce feeding behavior. Several other hormones are produced by adipose tissue and have effects on nutrient metabolism, including adiponectin and inflammatory cytokines (see Table 42.2). Dysregulation of these hormone systems is associated with obesity.

Hormonal Regulation of Nutrient Storage, Distribution, and Metabolism

Metabolism is dependent on energy balance, which is defined as the relationship between nutritional intake and expenditure. In general, metabolism is controlled by both the nervous system and the endocrine system. Several hormones are secreted from these systems that are triggered by the nutritional and energy status of the body, which may become rapidly altered during times of acute stress or chronic disease. Hormones involved in nutrient metabolism include insulin, glucagon, catecholamines, growth hormone, and cortisol. The effects of these hormones on carbohydrate, fat, and protein metabolism are summarized in Table 42.3. The primary role of insulin is to promote storage of nutrients when nutrients are abundant in the bloodstream. Insulin signaling is discussed in detail in Chapter 41. The primary metabolic role of the other hormones is to ensure adequate glucose availability for the brain by increasing blood glucose levels.

TABLE 42.3 Hormonal Actions on Carbohydrate, Lipid, and Protein Metabolism

Hormone	Actions
Insulin	Stimulates glucose uptake by liver, muscle, and fat cells
	Stimulates glycogenesis
	Inhibits gluconeogenesis
	Increases fatty acid uptake by fat cells
	Actively transports amino acids into cells
	Accelerates cellular protein synthesis
Glucagon	Stimulates glycogen breakdown
	Increases gluconeogenesis
	Promotes lipolysis in fat cells
	Stimulates protein breakdown into amino acids
	Increases amino acid transport into hepatic cells
Catecholamines	Maintains blood glucose level during stress
	Diminishes glucose uptake by cells
	Increases glycogen breakdown
	Increases fat mobilization
	Increases serum free fatty acid levels
Growth hormone	Stimulates glycogenolysis and raises blood glucose
	Inhibits uptake of glucose by insulin sensitive cells
	Stimulates lipolysis and release of free fatty acids
	Increases amino acid uptake into cells and protein synthesis
Cortisol	Stimulates gluconeogenesis and increases blood glucose
	Inhibits glucose uptake by insulin sensitive cells
	Increases fat cell membrane permeability
	Increases protein catabolism

Insulin

Insulin is the primary hormone that lowers blood glucose by increasing cellular uptake of glucose. Insulin, a peptide hormone synthesized by β cells in the pancreas, is secreted in response to increased blood glucose levels. Minutes after ingestion of a meal, insulin levels in the blood raise significantly, peak in 30 minutes, and plateau in about 3 hours. Between meals, when blood glucose levels tend to drop, insulin levels remain low.

Insulin promotes glucose uptake by the liver and increases the synthesis and storage of glycogen for later release. Gluconeogenesis and the breakdown of glycogen to form glucose (glycogenolysis) are inhibited by insulin. In addition to the liver, two other tissues are important targets for insulin signaling: muscle and adipose. These tissues also respond to insulin by taking up excess glucose and storing it as glycogen in muscle or transforming it into lipid for storage in fat cells. Insulin also increases the uptake of amino acids by cells and accelerates protein synthesis and growth in many tissues. Insulin is a fat-sparing hormone and decreases the activity of lipases in fat cells so that fat stores are not released. Thus insulin increases the fat mass of the body. Excessive exposure to insulin (endogenous or exogenous) can reduce insulin sensitivity and impair glucose uptake by these tissues, resulting in persistently elevated blood glucose levels (see Chapter 41).

Glucagon

Glucagon is a peptide hormone secreted mainly by α cells in the pancreas when glucose levels begin to drop. The two primary effects of glucagon are to promote the breakdown of liver glycogen with subsequent release of glucose into the bloodstream and to promote liver gluconeogenesis. Glucagon promotes gluconeogenesis by stimulating the breakdown of protein into amino acids and increasing their transport into hepatic cells for conversion into glucose. These actions tend to bring serum glucose levels back up to normal.

Catecholamines

Catecholamines (i.e., epinephrine and norepinephrine) serve a role in carbohydrate metabolism to maintain blood glucose levels during times of stress (see Chapter 2). During the stress response, catecholamines stimulate the conversion of glycogen to glucose in the muscles and liver. Epinephrine suppresses insulin release from the pancreas, thereby preventing glucose movement into muscle cells. Epinephrine and norepinephrine increase fat mobilization by stimulating the activity of hormone-sensitive lipase, thus increasing the serum free fatty acid level.

Growth Hormone

Growth hormone can have a significant impact on glucose regulation under circumstances of hypoglycemia and when it is hypersecreted (acromegaly). Hypoglycemia stimulates the release of growth hormone with a subsequent increase in gluconeogenesis in the liver and inhibition of glucose uptake by muscle cells. Chronically elevated serum growth hormone levels tend to persistently increase blood glucose levels and may contribute to the development of type 2 diabetes (diabetogenic hormone). Growth hormone also increases fatty acid mobilization from adipose tissue. In addition, it increases amino acid transport into cells and stimulates protein synthesis and growth in a variety of cell types.

Cortisol

Cortisol, a glucocorticoid, is an insulin antagonist and helps maintain serum glucose levels. During fasting, cortisol promotes gluconeogenesis and lipolysis. Cortisol is an important stress hormone and helps maintain adequate serum glucose for brain function. If deficiency occurs during stress or fasting, significant hypoglycemic reactions can occur. Cortisol

deficiency may be a significant cause of morbidity and mortality in critically ill patients (see Chapter 40).

KEY POINTS

- Metabolism is influenced by genetics, epigenetics, culture, lifestyle, and environment. About 25% of body fat accumulation may be attributed to genes, whereas complex interactions with epigenetics and environmental influences have a greater impact on body composition. Conditions during fetal development may significantly affect metabolism and body composition in adulthood.
- The primary hormonal regulators of appetite and eating behavior are ghrelin and leptin. Ghrelin is made in the gastrointestinal tract and stimulates appetite, whereas leptin is made by adipose cells and decreases eating behavior. However, eating is driven by a number of nonhunger-associated cues, such as stress, boredom, and socialization, that may dysregulate hormonal control.
- The primary hormonal regulators of nutrient metabolism are insulin, glucagon, catecholamines (i.e., epinephrine and norepinephrine), growth hormone, and cortisol,
- Insulin is secreted from pancreatic β cells in response to elevated serum glucose levels. Binding of insulin to receptors on target cells (liver, muscle, adipose tissue) facilitates the transport of glucose into cells and reduces blood glucose levels. Insulin inhibits lipolysis and gluconeogenesis.
- Glucagon is secreted from pancreatic α cells in response to low blood glucose levels. Glucagon promotes glycogenolysis and gluconeogenesis (from lactate, amino acids, and glycerol) by the liver, thereby increasing blood glucose levels.
- Catecholamines increase glycogenolysis and gluconeogenesis by the liver, thereby increasing blood glucose levels. Catecholamines also stimulate lipolysis in adipose cells by enhancing the action of hormone-sensitive lipase.
- Growth hormone increases blood glucose levels by inhibiting uptake by muscle cells and by stimulating gluconeogenesis in the liver. Growth hormone enhances the cellular uptake of amino acids and stimulates protein synthesis.
- Cortisol enhances the actions of glucagon and catecholamines and promotes glycogenolysis, gluconeogenesis, and lipolysis, thus raising blood levels of glucose and fatty acids.

OBESITY AND METABOLIC SYNDROME

Obesity

Obesity is defined as excessive accumulation of adipose tissue and is considered to be one of the leading contributors to preventable death in the United States. The body mass index (BMI) is calculated as weight (kg) divided by height squared (m^2). Overweight is defined as a BMI of 25 to 29.9 kg/m^2. Obesity is defined as a BMI greater than 30 kg/m^2, with morbid obesity 40 kg/m^2 or more. Obesity is a major health issue associated with higher rates of heart disease, kidney disease, hypertension, type 2 diabetes, polycystic ovary disease, and stroke. Obesity also creates challenges with activities of daily living from increased risk of osteoarthritis, chronic pain, gallbladder disease, sleep apnea, and psychosocial impairments. Approximately 20% of men and 25% of women in the United States meet the BMI criteria for obesity, and more than two-thirds of the population is overweight or obese.

The development of obesity involves complex interactions among a number of factors, including lifestyle choices, environment, epigenetics, and genetics. Endocrine causes of obesity are important and include hypothyroidism, hypercortisolism (Cushing syndrome), and hyperinsulinemia (metabolic syndrome and type 2 diabetes). However, most obese patients do not have a specific endocrine disorder, but rather an imbalance between caloric intake and energy expenditure. However, the problem is not a simple mathematical equation because of the influence of alterations in metabolic rate and hormonal regulation.

Once fat mass has accumulated, it tends to develop a new "set point" and resists reductions. It is well known that weight loss is notoriously difficult to maintain over time. Presumably regulatory hormones favor the reaccumulation of fat as soon as dietary intake is no longer restricted. The number of fat cells is thought to be more or less determined in childhood and adolescence, so particular attention to avoiding excessive caloric intake and accumulation of fat mass during that time is recommended.

Metabolic Syndrome

Metabolic syndrome presents with increased amounts of abdominal adiposity, increased levels of plasma free fatty acids, insulin resistance, low serum high-density lipoprotein concentrations, hypertriglyceridemia, and hypertension. The risk of metabolic syndrome increases with age and sedentary lifestyle. Abdominal fat accumulation is the most recognized symptom and, when combined with hypertension, hyperlipidemia, and insulin resistance, can significantly increase the risk for cardiovascular disease and diabetes.

Certain ethnic groups have a higher risk of developing metabolic syndrome when they are exposed to a Western diet and gain fat mass. Fat tends to accumulate in the visceral compartment to a greater extent in these groups. Native Americans and Hispanic Americans in particular exhibit a higher incidence of metabolic syndrome and a greater risk of type 2 diabetes than Caucasian Americans (see Chapter 41).

When excess fat is deposited in the subcutaneous tissue of the hips and legs, it does not confer the same risk of metabolic syndrome as visceral fat. Fatty acids released during lipolysis of visceral fat readily enter the portal vein and are directly delivered to the liver. Exposure of the liver to high free fatty acids is an important factor for developing insulin resistance, a key feature of metabolic syndrome. In addition, visceral fat produces proinflammatory cytokines that contribute to systemic inflammation and increased atherosclerotic cardiovascular disease. Metabolic syndrome frequently is a precursor to type 2 diabetes. Vigorous lifestyle management to increase activity and regulate dietary intake can delay or prevent progression, but is often difficult to maintain over time.

KEY POINTS

- Obesity is a complex disorder with multifactorial causes, including genetics, epigenetics, environmental influences, and hormonal dysregulation of appetite and energy expenditure.
- Obesity rates are increasing, with an estimated 20% of men and 25% of women in the United States having BMIs greater than 30 kg/m^2. An even larger number fall into the category of overweight with a BMI between 25 and 29.9 kg/m^2.
- Metabolic syndrome is characterized by excess abdominal adipose tissue, insulin resistance, hypertension, hyperglycemia, and dyslipidemia. It significantly increases the risk of cardiovascular disease and may progress to type 2 diabetes.

METABOLIC RESPONSES TO STARVATION AND PHYSIOLOGIC STRESS

An appropriate diet includes sufficient calories from carbohydrates, fats, and proteins; sufficient vitamins and minerals; and enough essential amino acids and fats to support biosynthesis. Many situations can lead

to dietary insufficiency, including poverty, war, displacement, chronic alcoholism, malabsorption syndromes, severe illness, and self-imposed restrictions. Metabolic responses to these conditions differ because of the severity of malnutrition, the duration of malnutrition, and concomitant conditions such as severe physiologic stress or trauma. Differences among metabolic responses to starvation and physiologic stress are described next.

Starvation and Protein-Energy Malnutrition

The metabolic response to starvation is different from the response to other forms of serious physiologic stress such as exhaustion or trauma. Starvation is a gradual process in which the metabolic rate decreases as storage carbohydrate reserves are metabolized. Liver glycogen stores are depleted within a few days, and the supply of glucose released into the bloodstream declines. When insulin levels decline during fasting, free fatty acids are released for energy use. Despite the available free fatty acids, protein is also used for energy by means of gluconeogenesis because certain tissues prefer glucose as an energy source. The use of protein for gluconeogenesis creates a negative nitrogen balance. However, as starvation continues, overall energy needs are reduced by slowing metabolism, and the tissues that usually require glucose for energy adapt by using ketone bodies produced from fat. Eventually lipolysis of stored fat provides the source of needed energy, and the use of protein as an energy source decreases. This change is an adaptive response through which the body strives to conserve lean body mass. Overall, the result is reduced depletion of the body's protein. Decreased serum glucose levels and urinary nitrogen excretion, along with elevation in the levels of ketone bodies and free fatty acids, characterize starvation. Fasting (voluntary starving) is not associated with a high mortality unless it is prolonged. Fig. 42.2 provides a summary of the physiologic effect of starvation.

Marasmus is a condition in children with prolonged malnutrition resulting in a body weight below 60% of normal for age and gender. Body muscle and fat are lost over time; however, visceral protein is conserved and serum albumin levels remain near normal (Fig. 42.3). When children eat primarily carbohydrates with little protein, a more severe form of malnutrition called *kwashiorkor* develops. Visceral protein is metabolized for energy and as it becomes depleted the serum albumin level drops. Low serum albumin results in generalized edema, especially of the face and abdomen, producing a bloated appearance (Fig. 42.4). Kwashiorkor resembles the metabolic response to severe physiologic stress described next.

Physiologic Stress

With physiologic stress from severe illness, conservation of lean body mass does not occur. The metabolic rate increases rather than decreases, and a high sustained rate of catabolism (breakdown of protein to meet energy needs) results. Amino acids are used as an energy source via hepatic gluconeogenesis using muscle stores of proteins. This adaptation quickly results in a negative nitrogen balance. The degree of hypermetabolism, hypercatabolism, and negative nitrogen balance associated with physiologic stress depends on the type, duration, and severity of the stressor present. Typically, the catabolic response to physiologic stress occurs in two phases: the immediate phase, lasting 5 to 8 days, and the subsequent adaptive phase.

The *immediate phase* of catabolism is characterized by increased sympathetic nervous system stimulation with release of glucagon, glucocorticoids, and catecholamines. The associated reduction of insulin activity causes a pseudodiabetic state. Hyperglycemia develops from decreased levels of circulating insulin and decreased utilization of glucose by muscle and other tissues (insulin resistance). Cells thus deprived of glucose uptake switch to oxidation of proteins for energy and to provide

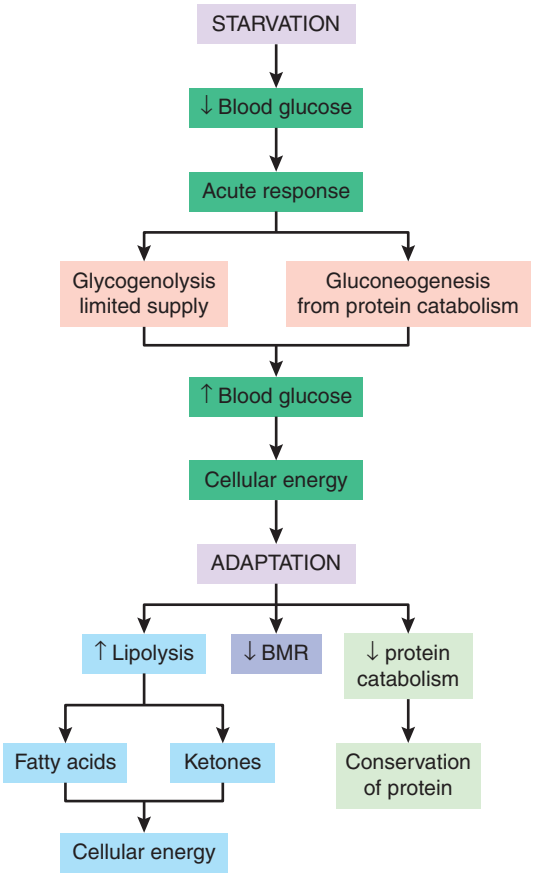


FIG 42.2 Physiologic responses to starvation.

FIG 42.3 Marasmus. This child is emaciated, and there is both hyperpigmentation and desquamation of the skin. (From Bologni JL, et. al: *Dermatology essentials.* Philadelphia, 2014, Saunders.)

FIG 42.4 Kwashiorkor, a particular manifestation of severe protein-energy malnutrition in some developing countries, where infants are weaned late from the breast and the young child's diet is high in starch. There is edema around the eyes and feet and leg, hyperkeratosis, and depigmentation of the skin and redness of the hair. (From Lissauer T, Carroll W: *Illustrated textbook of pediatrics*. Philadelphia, 2017, Saunders.)

amino acids for synthesis of acute-phase proteins by the liver (e.g., C-reactive protein, complement factors, ferritin, ceruloplasmin, amyloid A). As the amino acids are mobilized to meet energy needs, alanine can be used as a carbon source in hepatic gluconeogenesis. During this acute phase, adipose is not well utilized as an energy source. The nutritional result of the immediate phase of stress on the body is hyperglycemia, negative nitrogen balance, and retention of fluid and sodium. This protective mechanism uses skeletal muscle to meet energy requirements and protects the rest of the body's tissue from breaking down during periods of high-energy need. An overall loss of nitrogen and other electrolytes, including magnesium, phosphorus, and zinc, may occur.

In the *adaptive phase*, the body begins to use ketone bodies and reducing power from the oxidation of fatty acids, thus limiting protein catabolism. As the sympathetic nervous system response diminishes, insulin resistance decreases and glucose utilization improves. The overall result is an improvement in negative nitrogen balance. This phase is similar to the response of the body during starvation, when fat is used to meet energy requirements. Severe physiologic stress and chronic conditions such as diabetes, liver disease, or renal disease restrict the body's ability to move into the adaptive phase during physiologic stress. Nutrients are used more efficiently by the body's tissues during the adaptive phase than during the immediate phase. It is in this phase that nutrition can have a vital role in recovery. The combination of starvation and physiologic stress significantly increases the risk for morbidity and mortality.

A number of nutrient deficiencies can accompany severe illness and subsequently impair energy production, immune processes, and healing. Refeeding after a period of nutrient deprivation in patients with severe illnesses must be undertaken with caution to prevent a sudden increase in metabolism that could deplete phosphate stores or place undue burden on a failing heart or pulmonary system. For example, patients with significant pulmonary disease may need a lower-carbohydrate, higher-lipid diet to reduce the amount of carbon dioxide production. Otherwise, they may have difficulty managing the respiratory effort needed to remove the excess CO_2. Careful replacement of nutrients for energy production, as well as vitamins and minerals, is necessary. The major vitamins and minerals and their associated deficiency disorders are shown in Table 42.4.

KEY POINTS

- Starvation may be fairly well tolerated for a time because skeletal muscle and adipose tissue are used for energy and metabolic rate is gradually reduced. When this occurs in children, it is called *marasmus*. Prolonged protein-calorie malnutrition, especially when concomitant with illness, can result in kwashiorkor with the depletion of visceral protein. The reduction in serum albumin results in generalized edema and a bloated appearance.
- Acute physiologic stress results in activation of the sympathetic nervous system. The immediate phase is characterized by a high metabolic rate, sustained catabolism, hyperglycemia, and salt and water retention. The sympathetic response promotes the use of protein stores for gluconeogenesis, which results in a negative nitrogen balance. Fat stores are poorly utilized.
- After 5 to 7 days of acute physiologic stress, the body may enter an adaptive phase that more closely resembles the normal response to starvation. Ketones and fatty acids from the lipolysis of fat stores are used for energy, and body proteins are conserved. Glucose utilization improves and hyperglycemia resolves. Aldosterone secretion diminishes and edema resolves. During the adaptive phase, nutrients supplied to the body are used more efficiently than during the immediate phase.

NUTRITIONAL CONSIDERATIONS FOR AGING AND ALTERED HEALTH STATES

Aging

There is little difference in the ability of healthy people, young or old, to metabolize glucose and little difference in insulin secretion by the β cells in the pancreas. What does appear to occur with the aging process is a change in tissue sensitivity to insulin. Many possible causes for this phenomenon have been proposed, such as altered carbohydrate intake, decreased muscle mass, and lowered activity levels. The elderly have higher levels of circulating serum proinsulin, and there tends to be a decrease in insulin clearance, leading to insulinemia. The aging process may alter insulin receptors and postreceptor signaling mechanisms.

The aging process may also affect lipid metabolism as proportionate body fat increases. Although caloric intake generally decreases, a concurrent loss of lean body mass and a decline in energy expenditure begin in middle age and continue through life. A decline in the resting metabolic rate also occurs with the aging process.

A decrease in the quantity of skeletal muscle normally occurs with aging. Although this decrease in muscle is associated with factors such as physical inactivity and a decrease in the number of neurons to muscle cells, endocrine factors also influence the loss of muscle mass. The decreased growth hormone secretion noted in elderly individuals leads to decreased protein synthesis and a decline in insulin-like growth factor 1, a condition called *somatopause*.

Infection, Sepsis, and Fever

Malnutrition contributes to infections, in part, because many immune responses rely on adequate protein availability to provide amino acids

TABLE 42.4 Vitamins and Minerals: Major Functions and Deficiency Syndromes

	Functions		Deficiency Syndromes
Fat-Soluble Vitamins			
Vitamin A	A component of visual pigment Maintenance of specialized epithelia Maintenance of resistance to infection		Night blindness, xerophthalmia, blindness Squamous metaplasia Vulnerability to infection, particularly measles
Vitamin D	Facilitates intestinal absorption of calcium and phosphorus and mineralization of bone		Rickets in children Osteomalacia in adults
Vitamin E	Major antioxidant; scavenges free radicals		Spinocerebellar degeneration
Vitamin K	Cofactor in hepatic carboxylation of procoagulants—factors II (prothrombin), VII, IX, and X; and protein C and protein S		Bleeding diathesis
Water-Soluble Vitamins			
Vitamin B_1 (thiamine)	As pyrophosphate, is coenzyme in decarboxylation reactions		Dry and wet beriberi, Wernicke syndrome, Korsakoff syndrome
Vitamin B_2 (riboflavin)	Converted to coenzymes flavin mononucleotide and flavin adenine dinucleotide, cofactors for many enzymes in intermediary metabolism		Ariboflavinosis, cheilosis, stomatitis, glossitis, dermatitis, corneal vascularization
Niacin	Incorporated into nicotinamide adenine dinucleotide (NAD) and NAD phosphate, involved in a variety of redox reactions		Pellagra—"three Ds": dementia, dermatitis, diarrhea
Vitamin B_6 (pyridoxine)	Derivatives serve as coenzymes in many intermediary reactions		Cheilosis, glossitis, dermatitis, peripheral neuropathy Maintenance of myelinization of spinal cord tracts
Vitamin B_{12}	Required for normal folate metabolism and DNA synthesis		Megaloblastic pernicious anemia and degeneration of posterolateral spinal cord tracts
Vitamin C	Serves in many oxidation-reduction (redox) reactions and hydroxylation of collagen		Scurvy
Folate	Essential for transfer and use of one-carbon units in DNA synthesis		Megaloblastic anemia, neural tube defects
Pantothenic acid	Incorporated in coenzyme A		No nonexperimental syndrome recognized
Biotin	Cofactor in carboxylation reactions		No clearly defined clinical syndrome
Minerals			
Iron	Essential component of hemoglobin as well as several iron-containing metalloenzymes	Inadequate diet Chronic blood loss	Hypochromic microcytic anemia
Iodine	Component of thyroid hormone	Inadequate supply in food and water	Goiter and hypothyroidism
Copper	Component of cytochrome *c* oxidase, dopamine β-hydroxylase, tyrosinase, lysyl oxidase, and unknown enzymes involved in cross-linking collagen	Inadequate supplementation in artificial diet Interference with absorption	Muscle weakness Neurologic defects Abnormal collagen cross-linking
Fluoride	Mechanism unknown	Inadequate supply in soil and water Inadequate supplementation	Dental caries
Selenium	Component of glutathione peroxidase Antioxidant with vitamin E	Inadequate amounts in soil and water	Myopathy Cardiomyopathy (Keshan disease)
Zinc	Component of enzymes, principally oxidases	Inadequate supplementation in artificial diets Interference with absorption by other dietary constituents Inborn error of metabolism	Rash around eyes, mouth, nose, and anus called *acrodermatitis enteropathica* Anorexia and diarrhea Growth retardation in children Depressed mental function Depressed wound healing and immune response Impaired night vision Infertility

Adapted from Kumar V, et al: *Robbins and Cotran pathologic basis of disease*, ed 9, Philadelphia, 2015, Saunders, pp 433, 442.

for synthesis of antibodies, complement factors, and other acute-phase proteins. The immune system is regulated by a complex system of cytokines and receptors, all of which require proteins for biosynthesis. The presence of infection can also contribute to malnutrition through inhibition of appetite and stimulation of metabolic rate. Fever is a common symptom accompanying infection. Fever increases metabolic needs by 7% for each 1°F increase (13% for each 1°C increase). Energy requirements can increase by 40% when a high fever (above 104°F) is present. The metabolic response to fever is both anabolic and catabolic, which greatly increases nutrient requirements.

Although catabolism may be detrimental in some aspects, it is also a protective mechanism that provides needed substrates for activation of the immune response to infection. Nutritional support is important to provide substrates (amino acids) for these protective mechanisms. Part of the body's metabolic response to infection is to increase the amount of available glucose in the blood by releasing stress hormones. Sometimes the metabolic demand is too great for the body to manage; for example, sepsis can increase energy expenditure 20% to 60% above basal energy requirements. Nutritional support is needed to supply additional energy and the necessary substrates so that body stores are not excessively depleted.

Surgery

Adequate nutrition before and after surgery promotes wound healing, prevents infection, and decreases complications and mortality. A patient should be in the best nutritional condition possible before surgery. A common cause of protein-energy malnutrition in postoperative patients is starvation. The combination of poor presurgical nutrition and postoperative starvation because of nausea, vomiting, sedation, or mechanical ventilation may increase complications after surgery and delay healing.

Nutritional needs in the postoperative period depend on the extent and type of surgery, as well as the presurgical nutritional status. The postoperative energy requirement can increase from 10% to 35% above BMR. In addition, nitrogen loss through wounds can be large and create a greater need for increased protein intake. Protein intake sufficient to replace losses and promote anabolism will be required, together with nonprotein calories for energy requirements. As with every patient, individual assessment and determination of exact needs are required.

Trauma

The general catabolic response to stress is seen frequently in trauma patients. Energy expenditure is increased by 15% to 30%. Increased carbohydrate intake will be needed, but patients must be observed for complications of high carbohydrate intake, such as glucose intolerance. Nitrogen loss secondary to catabolism and to cellular damage can be high. Circulating stress hormones in the immediate phase have an antiinsulin effect that decreases glucose utilization; therefore gluconeogenesis is increased to meet energy needs. As with other stress states, the catabolism of protein provides a source of amino acids for acute-phase protein synthesis by the liver (see Chapter 9). Because trauma is a sudden stress, catabolism predominates in the early period, resulting in excessive negative nitrogen balance and significant loss of skeletal muscle. Posttrauma rehabilitation is more difficult when there has been significant skeletal muscle wasting.

Burns

A major burn is an extreme physiologic stressor that results in significant hypermetabolism. In addition, the destruction of skin increases energy expenditure through evaporative heat loss (see Chapter 54 for a discussion of burns). The energy needs of a burn patient increase 50% to 100% from the BMR. Because of individual variations such as preburn nutritional status, the presence of other physiologic stresses, activity level, stage of burn, and patient age, determination of individual nutrition needs may be difficult.

As with other stressors, negative nitrogen balance is increased by catabolism and by the use of amino acids to form stress proteins. In addition, burn wounds directly contribute to protein loss because soluble proteins leak from the wound and proteolysis is activated. The effective utilization of available energy sources in the immediate postburn phase may be impaired. A nutritional consultation is recommended.

Cancer

The nutritional effects of cancer can be severe and result in what is commonly termed cancer cachexia (see Chapter 7). Cachexia is associated with the end stage of cancer but can also develop earlier. The cause of cachexia is inadequate nutritional intake relative to energy requirements and increased metabolism of tumor cells. It results in significant weight loss and muscle weakness. A major contributor to cachexia is anorexia associated with the malignancy and with treatment. Sensory alterations such as changes in smell or taste may be associated with cancer treatment and contribute significantly to malnutrition. Because both nutrient intake and substrate metabolism are altered, nutritional support is difficult to achieve and frequently ineffective in reversing the existing cachexia.

Immobility

The main nutritional effect of immobility is loss of calcium from nonstressed bone, a process that can elevate serum calcium and phosphorus levels. This demineralization is best managed with weight-bearing exercise as early as possible rather than calcium supplementation. Calcium supplementation during immobility may increase the risk of developing renal calculi. A physical therapist should be consulted early to assist in prevention of demineralization. Because negative calcium balance can increase in a catabolic state, serum calcium levels must be monitored and abnormalities treated.

A second effect of immobilization is nitrogen loss while tissue mass is decreased from disuse atrophy. This loss can total 2 to 3 g/kg per day and require up to 10 to 15 g of protein to replenish the daily loss, which further emphasizes the need for early physical therapy and aggressive range-of-motion exercises.

KEY POINTS

- A decline in the resting metabolic rate occurs with aging. This change is related to several factors, such as reduced lean body mass, reduced lipogenic enzyme response to glucose, and excessive fat mass.
- Infection is associated with fever and an increased metabolic rate. For each 1°F increase in body temperature, metabolic needs increase 7%. The synthesis of acute-phase proteins and immune factors requires sufficient amino acid substrates.
- A major nutritional problem in postoperative patients is starvation because of nausea, decreased alertness, or swallowing impairment. In addition, nitrogen loss through wounds may be significant.
- Major trauma is associated with a 15% to 30% increase in energy expenditure. Glucose utilization is maintained. The release of stress hormones can lead to problems with hyperglycemia.
- Major burns are extreme physiologic stressors that result in an increase in energy expenditure of 50% to 100% above baseline. Protein loss from burned areas is high.
- Cancer cachexia is a result of several factors, including anorexia, poor intake, and preferential nutrient utilization by tumor cells.
- Immobility is associated with muscle atrophy and bone demineralization.

SUMMARY

Metabolism is a dynamic and continuous process affecting every organ and physiologic process in the human system. The building phase of anabolic metabolism occurs concurrently with the energy-consuming and destructive phase of catabolic metabolism. Phases of metabolism either release or require energy in the form of ATP. The rate at which metabolism occurs in the resting human system is referred to as the BMR, and the process releases both heat and energy.

The regulation of nutrient intake and metabolism is complex, with genetics, epigenetics, culture, environment, and lifestyle playing a part. In developed nations, the development of excess body fat and a sedentary lifestyle are contributing to the development of many obesity-associated disorders, such as type 2 diabetes, cardiovascular disease, and mobility limitations. Malnutrition is still a significant problem in many parts of the world and in poor populations and the elderly in the developed world.

The endocrine system greatly affects metabolism. Only one hormone, insulin, is known to significantly lower serum glucose levels by decreasing liver glucose production and promoting the transfer of glucose into cells. Although each works in a unique manner, growth hormone, cortisol, epinephrine, and glucagon all act in concert to maintain or raise blood glucose levels.

Physiologic stress is accompanied by changes in metabolism that alter nutrient utilization and increase nutrient requirements. The degree to which these changes occur varies with the type and severity of the particular stress. If the patient is not provided with adequate nutrition when one or more stressors are present, the hypermetabolism, hypercatabolism, and negative nitrogen balance associated with the physiologic stress will have detrimental effects on recovery.

Health care professionals must be aware of the impact of stressors on the nutritional status of the body, as well as the impact of nutrition on the well-being of body systems. If this point is well understood, appropriate interventions can be taken to prevent some of the complications that can develop when nutritional support is inadequate. Most well-nourished patients can tolerate a short period of inadequate intake without untoward effects. However, critically ill patients require early nutritional support because of the magnitude and intensity of the stressors.

RESOURCES

A Report from the American Heart Association, 2015. Available at http://my.americanheart.org/statements.

American Heart Association: *Heart Disease and Stroke Statistics—2015 Update*.

Gade W, et al: Beyond obesity: the diagnosis and pathophysiology of metabolic syndrome. *Clin Lab Sci* 23(1):51–61, 2010.

Guyton AC, Hall JE: *Textbook of medical physiology*, ed 13, Philadelphia, 2015, Saunders.

Kumar V, et al: *Robbins and Cotran pathologic basis of disease*, 9th ed, Philadelphia, 2015, Saunders.

The National Health and Nutrition Examination Survey (NHANES): *Analytic and reporting guidelines*, 2015. Available at www.cdc.gov/nchs/nhanes.htm.

U.S. Department of Health and Human Services and U.S. Department of Agriculture. *2015 – 2020 Dietary Guidelines for Americans*. 8th ed. December 2015. Available at http://health.gov/dietaryguidelines/2015/guidelines/.

White BA, Porterfield SP: *Endocrine and reproductive physiology*, 4th ed, Philadelphia, 2013, Mosby.

World Health Organization: *Protein and amino acid requirements in human nutrition*, Geneva, 2007, WHO Press.

43

Structure and Function of the Nervous System

Jacquelyn L. Banasik

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- How do the central nervous system (CNS), peripheral nervous system, and autonomic nervous system interrelate?
- How is the CNS protected and supported?
- What structures are located in each of the four principal areas of the brain: cerebrum, diencephalon, cerebellum, and brainstem?
- What neurologic functions have been mapped to particular locations in the brain?
- How do the properties of neuronal action potentials and neuronal communication through synapses relate to the functions of the nervous system?
- How is the somatotopic organization of sensory receptors, sensory pathways, and muscle control maintained in the CNS?
- How is voluntary muscle activity initiated and executed?
- How do the properties of the mind, including thought, memory, learning, consciousness, and sleep, relate to the physiologic substance of the nervous system?

CHAPTER OUTLINE

STRUCTURAL ORGANIZATION, 851
Central Nervous System, 851
Support and Protection of the Central Nervous System, 851
The Brain, 853
Cerebrum, 853
Diencephalon, 856
Cerebellum, 860
Brainstem, 860
The Spinal Cord, 861
Peripheral Nervous System, 863
Cranial Nerves, 863
Spinal Nerves, 864
Autonomic Nervous System, 867
NEURONAL STRUCTURE AND FUNCTION, 868
Neurons and Supportive Cells, 868
Neurons, 868
Glia, 868
Neuronal Communication, 872
Membrane Potentials, 872
Synaptic Transmission, 874
Neurotransmitters, 875
Neuronal Circuits, 880
Neural Development, Aging, and Injury, 880
Development, 880
Aging, 881
Injury, 881
SENSORY FUNCTION, 882
Sensory Receptors, 882
Sensory Pathways, 883
Somatosensory Cortex, 883
MOTOR FUNCTION, 883
Motor Neurons, 884
Spinal Reflexes, 884
Central Control of Motor Function, 885
CONSCIOUSNESS, MEMORY, AND SLEEP, 886
Consciousness and Memory, 886
Sleep, 888

The nervous system is a complex network of neurons and supportive cells that enables rapid communication between sensory receptors, central processing neurons, and functional responses. Much has been discovered about the mechanisms of sensory input and motor output, but the physiologic bases of thought, consciousness, emotion, and learning remain elusive. The idea that the mind is within the biological realm has been generally accepted, and the effects of mind-altering drugs on emotions, appetite, sleep, thought, and sensory perception have long been recognized. Research continues to reveal the great complexity of neurologic function. A bewildering array of neurotransmitter-signaling molecules and an even greater number of neurotransmitter receptors have been identified. Recently the long-held notion that neurons cannot regenerate in the mature brain has been disproved, and neuronal stem cells have been identified in certain areas. Each discovery brings us closer to understanding neural physiologic processes and gives hope for finding effective therapies for the devastating diseases that affect them. This chapter provides an overview of neural structure and function and is the basis for understanding the neurologic disorders in the chapters that follow.

STRUCTURAL ORGANIZATION

The nervous system is traditionally divided into three principal anatomic units: the central nervous system (CNS), the peripheral nervous system (PNS), and the autonomic nervous system (ANS). These systems are not anatomically or functionally distinct, and they work together as an integrated whole. Therefore when function, rather than anatomy, is the topic of concern, the nervous system is more conveniently divided into the sensory, motor, and higher brain functions. This chapter begins with a review of the major anatomic features of the nervous system; then addresses neurologic function at the cellular and synaptic level; and concludes with a summary of sensory, motor, and cognitive functions.

CENTRAL NERVOUS SYSTEM

The CNS includes the brain and spinal cord. Its primary functions are receiving and processing sensory information and creating appropriate responses to be relayed to muscles and glands. It is the site of emotion, memory, cognition, and learning. The CNS is bathed in cerebrospinal fluid (CSF) and shielded from the periphery by the blood–brain barrier (BBB). The CNS interacts with the neurons of the PNS through synapses in the spinal cord and cranial nerve ganglia.

Support and Protection of the Central Nervous System

Nervous tissue has the consistency of gelatin, so measures to support and protect its fragile structure are necessary. In addition, the CNS must be shielded from circulating substances that would interfere with neurotransmission. These protective functions are provided by the skull and vertebral column, meninges, CSF, and BBB. The bony structures of the skull and vertebral column encase the brain and cord and protect them from external trauma, whereas the CSF and meninges provide buoyancy and shock-absorbing capacity.

The meninges are composed of three layers that serve to suspend and maintain the shape and position of the nervous tissue during head and body movements. The brain is suspended within layers of meninges that are fixed to the skull. In this manner, the brain turns with the movement of the skull. The CSF circulates within the subarachnoid space, giving buoyancy to the brain and making an average 1500-gram brain mass resistant to distortion, which could occur from gravity alone were it not for the buoyancy effect. The three meningeal layers are the *dura mater, arachnoid,* and *pia mater* (Fig. 43.1).

The dura mater, the outermost meningeal layer, is a thick, tough, collagenous membrane. It is composed of two layers, one contiguous with the periosteum of the skull, and the other, which is adherent to the first, covering the surface of the brain. The tough dura protects the soft tissue of the brain. Support and stability are also provided by dural septa that invaginate into the cranial cavity. The falx cerebri is a thin wall of dura that folds down the cortical midline, separating the two hemispheres. The tentorium cerebelli is a septum that separates the cerebellum and brainstem from the rest of the cerebrum. The dural septa fix the brain in place by their tentlike structure and limit its movement within the skull. Venous sinuses that collect venous blood from cerebral veins are located between the two layers of the dura at the base of the septum.

Beneath and continuous with the dura is the arachnoid layer. The spaces between the dura and the skull and between the dura mater and the arachnoid are potential spaces. Only in the presence of pathologic processes, notably epidural and subdural hemorrhages, do these spaces become evident (see Chapter 44). Unlike the dura mater, the arachnoid is a thin, delicate membrane. It is semitransparent and weblike in appearance, hence its name. Strands of collagenous connective tissue called *trabeculae* extend from the arachnoid layer down to the pia mater, forming a subarachnoid space. The CSF flows in this space.

The pia mater, the third meningeal layer, is also very thin. However, unlike the other meningeal layers, the pia is attached to the brain and closely follows its contours over every sulcus and into every gyrus. Consequently, the subarachnoid space between the arachnoid and the pia mater is not evenly distributed. The arachnoid meshes with the pia via the trabeculae in such a subtle manner that it is often difficult to differentiate one from the other. Consequently, the two layers together are often referred to as the *leptomeninges.*

The meninges that cover and provide protection to the spinal cord are similar to those of the brain, with a few variations Fig. 43.2). The spinal dura has no periosteal layer, so it is a single rather than a double layer. It is continuous with the foramen magnum at the base of the skull and is separated from the spinal vertebral periosteum by an epidural space. Thus in the spinal cord, the epidural space is a true space, unlike its counterpart in the cranium, which is only a potential space. Within this space lie fatty connective tissue and a vertebral venous plexus.

The spinal arachnoid, much like that covering the cerebrum, is closely adherent to the spinal dura. Between the arachnoid layer and the pial lining is the CSF-filled subarachnoid space. The spinal meninges end at approximately the second sacral vertebra. However, the spinal cord ends between the first and second lumbar vertebrae (L1 to L2). This results in a large subarachnoid cistern, called the *lumbar cistern,* which is a favored place to obtain CSF samples (see Fig. 43.2). The spinal pia is much tougher and thicker than the cerebral pia. Projecting along the length of each side is the dentate ligament, which anchors the spinal cord to the arachnoid and through it to the dura. Another pial projection connects the tail of the spinal cord (the cauda equina) at level L1 to L2 to the caudal end of the spinal dural sheath, where it is tethered to the end of the vertebral column. This projection is called the *filum terminale.*

The majority of CSF is produced by the choroid plexus, located in the lateral and third ventricles of the brain, at a rate of approximately 500 mL/day. The composition of normal CSF is compared with plasma in Table 43.1. CSF is absorbed at about the same rate at which it is produced, so that only 150 to 175 mL is in circulation at any time. The large C-shaped lateral ventricles occupy the center of each hemisphere. They communicate with the third ventricle in the diencephalon by way of the intraventricular foramen. The third ventricle is linked to the fourth ventricle by way of the cerebral aqueduct, which lies between the pons and the medulla (Fig. 43.3). The CSF flows from the fourth

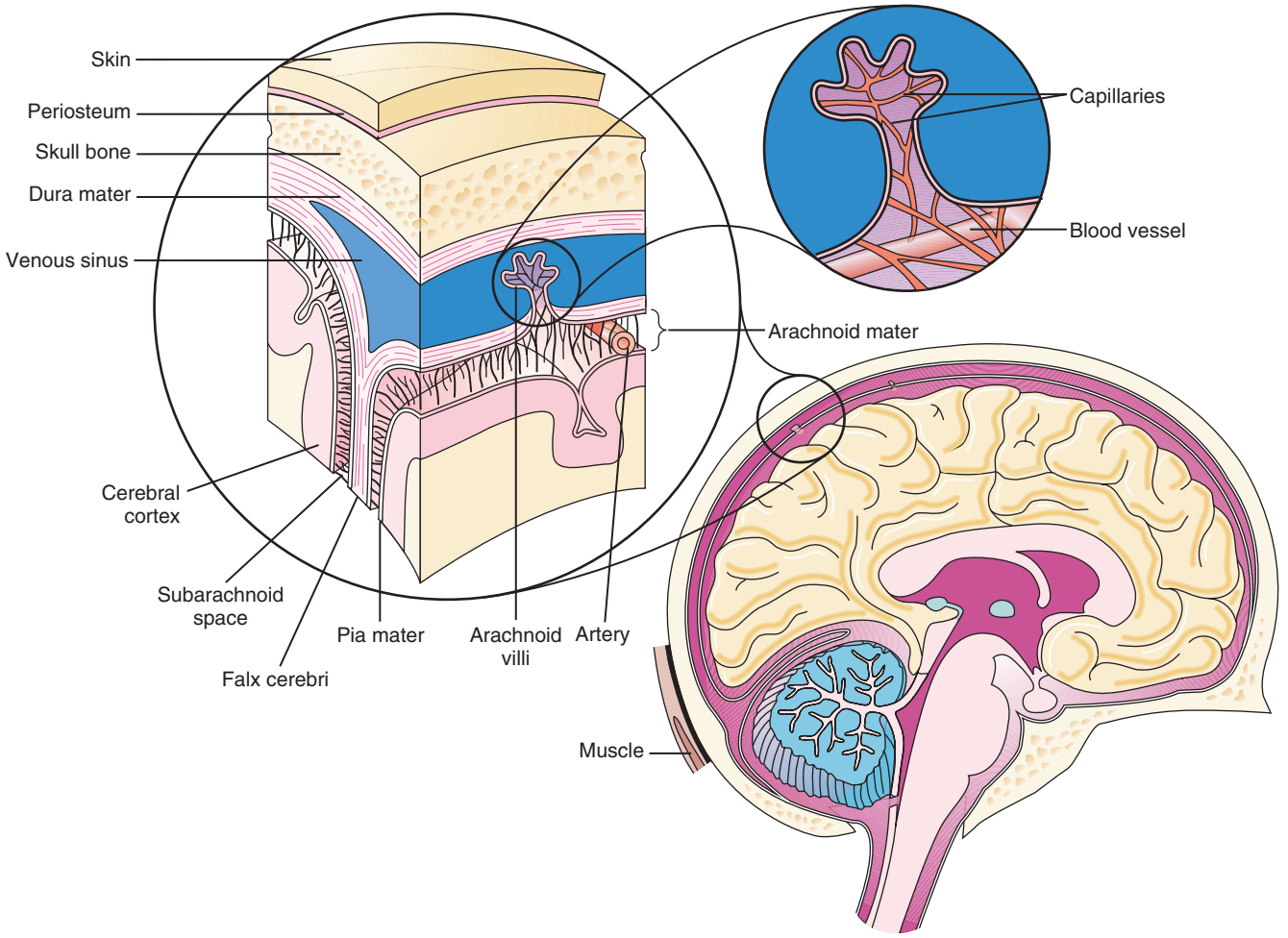


FIG 43.1 Principal membranes of the cranial meninges. Cerebrospinal fluid flows in the subarachnoid space and is reabsorbed by arachnoid villi within the dural sinuses.

TABLE 43.1 Composition of Cerebrospinal Fluid Compared With Plasma and Selected Cerebrospinal Fluid Abnormalities

Substance	Normal CSF	Abnormal CSF	Plasma
Na^+ (mEq/L)	148	—	136–145
K^+ (mEq/L)	2.9	—	3.5–5.0
Cl^- (mEq/L)	120–130	—	100–106
Glucose (mg/dL)	50–75	↓ Infection	70–100
Protein (mg/dL)	15–45	↑ Inflammation	6800
pH	7.3	—	7.4
Red blood cells (high-power field)	None	↑ Trauma, subarachnoid hemorrhage	—
White blood cells (high-power field)	<5	↑ Infection (e.g., meningitis)	—
Pressure (mm H_2O)	70–180	↑ Mass lesions	—

CSF, Cerebrospinal fluid.

ventricle through the median or lateral aperture and into the subarachnoid space. It flows around the spinal cord and up over the cerebral hemispheres to the arachnoid villi, where it is absorbed into the venous system.

CSF is absorbed by the arachnoid villi, which are small tufts of the arachnoid that invaginate into the dural sinus (see Fig. 43.1). These tufts bring CSF into close approximation with venous blood. CSF flows into the venous system through one-way valves because of pressure gradient differences. Although the CSF flows readily into the venous sinus, flow in the opposite direction cannot occur; that is, the fluid in the venous sinus cannot flow into the subarachnoid space. This mechanism is part of a system of barriers between the extracellular space in the nervous system and the rest of the body.

The rate of production of CSF is independent of blood pressure or intraventricular pressure. Thus CSF will continue to be produced even when its path of circulation or absorption is blocked. If this occurs, the amount of CSF increases, as does the size of the ventricles. This pathologic process is called *hydrocephalus.* Although hydrocephalus is

usually caused by blockage of CSF pathways, it can also be caused by overproduction and malabsorption of CSF (see Chapter 45).

Blood supply to the brain is provided by two pairs of arteries; the anterior circulation is supplied by the internal carotid arteries, and the posterior circulation is supplied by the vertebral arteries (Fig. 43.4). The internal carotid arteries have three principal branches: the anterior and middle cerebral arteries and the posterior communicating arteries. The vertebral arteries enter the skull at the foramen magnum and join at the level of the pons to form the basilar arteries. The ring of vessels that unites the anterior and posterior circulation at the base of the brain is known as the *circle of Willis* (see Fig. 43.4B). The cerebral veins drain into large vascular channels called *sinuses* that are formed by folds in the dura. From the sinuses, venous blood returns to the heart by way of the jugular veins (see Fig. 43.4C).

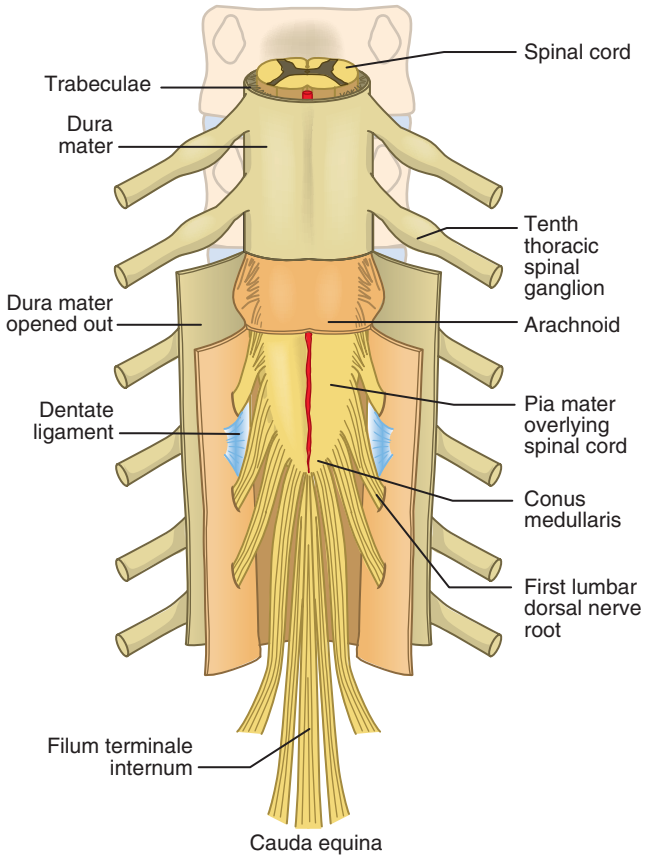


FIG 43.2 Spinal cord, spinal nerves, and meninges. Spinal meningeal layers are similar to cranial membranes. The spinal cord ends at L2, whereas the meninges continue on for several segments, creating a CSF-filled cistern where CSF fluid can be obtained by lumbar puncture.

The extracellular fluid that bathes the neurons is carefully shielded from elements in the CSF and blood by cellular barriers. Specialized tight junctions between the cells that line the CSF spaces and between the endothelial cells of brain capillaries prevent leakage of molecules through the spaces between the cells (Fig. 43.5). Therefore substances must move through the plasma membranes of these barrier cells to access the CNS. Lipid-soluble molecules move through more easily than do water-soluble ones. Thus the flow of ions, nutrients, drugs, proteins, and other charged or polar substances is highly restricted. The BBB is a crucial structure for protecting the brain, but it may also restrict access of beneficial molecules, such as antibiotics and cancer drugs, making treatment more difficult.

The integrity of the BBB is maintained in part by CNS cells called *astrocytes.* These specialized glial cells have foot processes that contact the brain capillaries and are thought to help regulate transport across the capillary endothelium (see Fig. 43.5). The BBB is less effective in infancy and can also be compromised by ischemia and chemical injury in adults.

A similar barrier exists between the circulating CSF and the interstitial fluid of the CNS—the CSF–brain barrier. The ependymal cells that line the ventricles are tightly joined and regulate the movement of water-soluble elements between the CSF and neurons. In addition, these cells serve the important function of removing unwanted substances from the CNS and secreting them into the CSF for eventual removal by the venous system.

Some areas of the brain need to sample the contents of the blood or CSF more directly to make regulatory adjustments in respiratory, autonomic, or endocrine functions, and these areas therefore have more permeable barriers (leaky BBB). These areas include the hypothalamus, pituitary, and other circumventricular organs (around the ventricles).

The Brain

Various schemes have been used to subdivide the structures of the brain using embryologic, evolutionary, and anatomic frameworks (Table 43.2). In this section, an anatomic framework is used that includes the cerebrum, diencephalon, cerebellum, and brainstem (Fig. 43.6).

Cerebrum

The cerebrum is divided into left and right hemispheres by the longitudinal fissure and is the largest part of the brain. The cerebral cortex is the outermost layer of the cerebrum and is composed of gray matter arranged in six histologically distinct layers (Fig. 43.7). Each layer makes connections with other parts of the brain. The cortex is characterized by its convoluted exterior with ridges (gyri), grooves (sulci), and deeper depressions (fissures). The sulci and fissures are used as landmarks to divide the cerebral cortex into lobes. The central sulcus separates the frontal and parietal lobes, the lateral sulcus separates the temporal lobe from the parietal and frontal lobes, and the parietooccipital line defines the occipital lobe (Fig. 43.8).

TABLE 43.2 Subdivisions of the Brain Using Embryologic, Evolutionary, and Anatomic Frameworks

	FRAMEWORK		
Structure	**Embryologic**	**Evolutionary**	**Anatomic**
Cerebral hemisphere	Telencephalon	Forebrain (includes diencephalon)	Cerebrum
Thalamus	Diencephalon	—	Diencephalon hypothalamus
Midbrain	Mesencephalon	Midbrain	—
Cerebellum	Metencephalon (includes pons)	Cerebellum	Cerebellum
Medulla	Myelencephalon	Hindbrain	Brainstem (includes midbrain, pons)

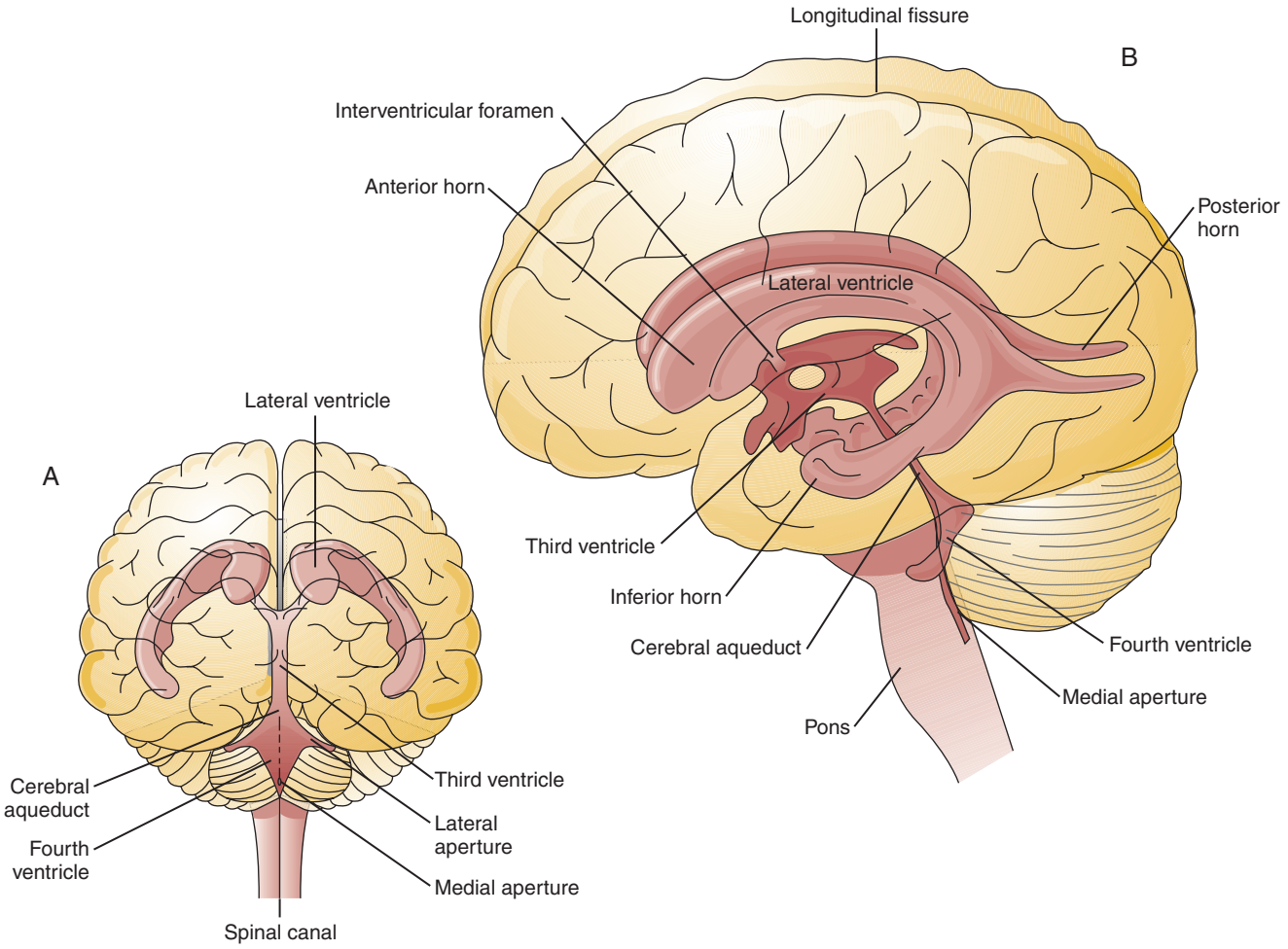


FIG 43.3 Ventricles within the brain from frontal **(A)** and lateral **(B)** views.

TABLE 43.3 Functional Areas of Brain Specialization

Area	Specialized Function
Occipital lobe	Visual cortex and association areas
Parietal lobe	Somatosensory cortex and association areas
Temporal lobe	Hearing and equilibrium, emotion, and memory
Frontal lobe	Motor cortex and association areas; prefrontal cortex involved in complex thought, ethical behavior, and morality
Limbic structures	Emotions, short-term memory, olfaction
Basal ganglia	Initiation and planning of learned motor activities
Broca and Wernicke areas	Interpretation and expression of language

Some anatomic locations are particularly associated with certain brain functions. The functional areas of specialization of brain loci are listed in Table 43.3. A partial map of Brodmann areas is shown in Fig. 43.9. Although the concept of functional anatomic areas is clinically useful, one should realize that even though an area may be critical for a particular function, it is not wholly responsible for that function, and many brain areas may be involved. A certain degree of reassignment of brain function from one area to another can occur, allowing the brain to adapt to loss of normal neural function (neural plasticity).

Functional areas of the cortex that can be mapped to specific sensory receptors or muscles are called *primary* areas. Primary areas are surrounded by secondary areas that provide greater character to sensations and greater complexity to movements. In addition to primary and secondary cortical areas, there are large areas of association cortex that add interpretive and learned responses. Organization of the primary and secondary cortex is best characterized for the somatosensory cortex and motor cortex (which are discussed later in the Somatosensory Cortex and the Central Control of Motor Function sections, respectively).

Cortical areas involved in visual perception are located in the occipital lobe. Interpretive association areas for vision are found in the adjacent temporal and parietal lobes. The primary auditory cortex is located on the superior temporal lobe, whereas vestibular information projects to the inferior temporal lobe.

Language expression and interpretation have been mapped to areas in the temporal lobe, particularly the Wernicke area. One hemisphere, usually the left, is dominant for language. Lesions in this area lead to difficulty recognizing written words (alexia) and spoken language

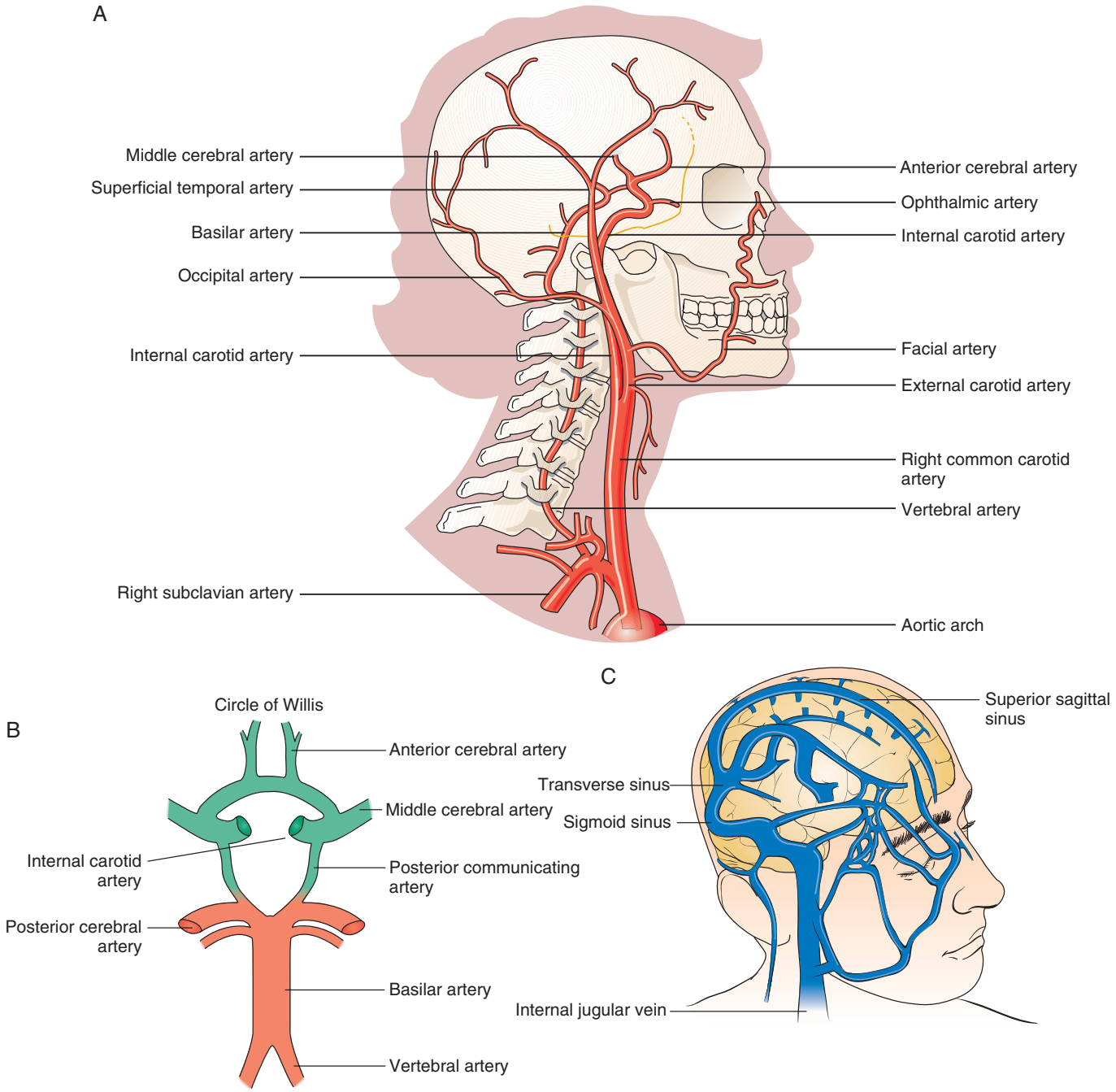


FIG 43.4 Blood supply to the brain. **A,** The internal carotid and vertebral arteries supply blood to the anterior and posterior aspects of the brain, respectively. **B,** At the base of the brain, the internal carotid and vertebral arteries join to form the circle of Willis. **C,** Major venous drainage from the brain.

(receptive aphasia). Another area closely associated with speech is the Broca area in the frontal lobe. Damage to this region interferes with the ability to use language (expressive aphasia).

The frontal lobe is usually credited with control over emotional responses, ethical behavior, and morality. It is also the site of initiative and motivation. Patients with lesions of the frontal lobe may fail to conform to societal behavioral norms.

The limbic lobe and limbic system are the parts of the cerebrum most closely associated with memory and emotion. The limbic lobe is a ring of cortex on the medial surface of each hemisphere containing the cingulate gyrus, isthmus, and parahippocampal gyrus (Fig. 43.10). Olfaction (the perception of smell) occurs within the limbic cortex. The limbic system is a group of structures that encircle the brainstem. In addition to the limbic lobe, the limbic system includes the amygdala, fornix, hippocampus, and portions of the thalamus (see Fig. 43.10). Lesions of the limbic system, particularly the hippocampus, cause impairment of short-term memory.

The basal ganglia are large masses of gray matter that lie deep within the cerebral hemispheres. They are intimately involved in the initiation, coordination, and execution of movement. The basal ganglia include the caudate nucleus, putamen, globus pallidus, subthalamus, and substantia nigra (Fig. 43.11). The caudate nucleus and putamen together are called the *striatum*. The five basal ganglia structures occur in pairs, with each cerebral hemisphere containing a set.

The basal ganglia are connected by complex neural circuits that incorporate sensory information about the current muscle conditions, cortical input about desired motor activities, and cerebellar signals about timing and coordination. Much of what is known about the function of basal ganglia has been learned from studying Parkinson disease. Parkinson disease is characterized by difficulty initiating voluntary movements (akinesia), stiff muscles (rigidity), and a tremor of the hands when idle (rest tremor). Improvement in symptoms occurs when the patient is given a precursor of dopamine (DA, levodopa), which can cross the BBB (see Chapter 45).

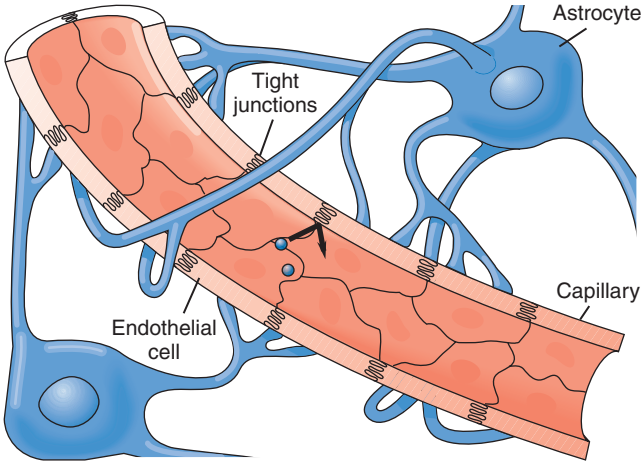


FIG 43.5 Tight junctions between brain capillary endothelial cells prevent polar and charged molecules from passing between cells. Astrocytes have foot processes on the capillary that help maintain integrity of the blood–brain barrier.

In addition to the gray matter of the cerebral cortex and the basal ganglia, the cerebrum contains thick layers of white matter that consist of myelinated axons. Some of these axons connect the two cerebral hemispheres (commissural fibers); some connect one area of cortex to another within the same hemisphere (association fibers); and others connect the cortex with lower brain centers, including the thalamus, basal ganglia, brainstem, and spinal cord (projection fibers). The corpus callosum and the anterior commissure connect the two hemispheres. The corpus callosum is a massive bundle of fibers crossing the brain just above the lateral ventricles and is the principal means of communication between the hemispheres.

In summary, the cerebrum is the largest brain structure, garnering about 70% of the neurons and supporting cells of the brain to accomplish its diverse and complex functions. Each of the 100 billion neurons in the brain may make hundreds of synaptic connections with other neurons, providing an incomprehensible number of potential interactions.

Diencephalon

The diencephalon lies deep in the brain, forming a connecting structure between the upper brainstem (midbrain) and the cerebral hemispheres. The principal structures of the diencephalon are the thalamus, hypothalamus, pineal gland, epithalamus, and ventral thalamus (Fig. 43.12). The third ventricle also traverses the diencephalon.

The thalamus is the principal receiving site and relay center for impulses traveling to the cerebral cortex from the spinal cord, cerebellum, and basal ganglia. In addition to processing and relaying sensory

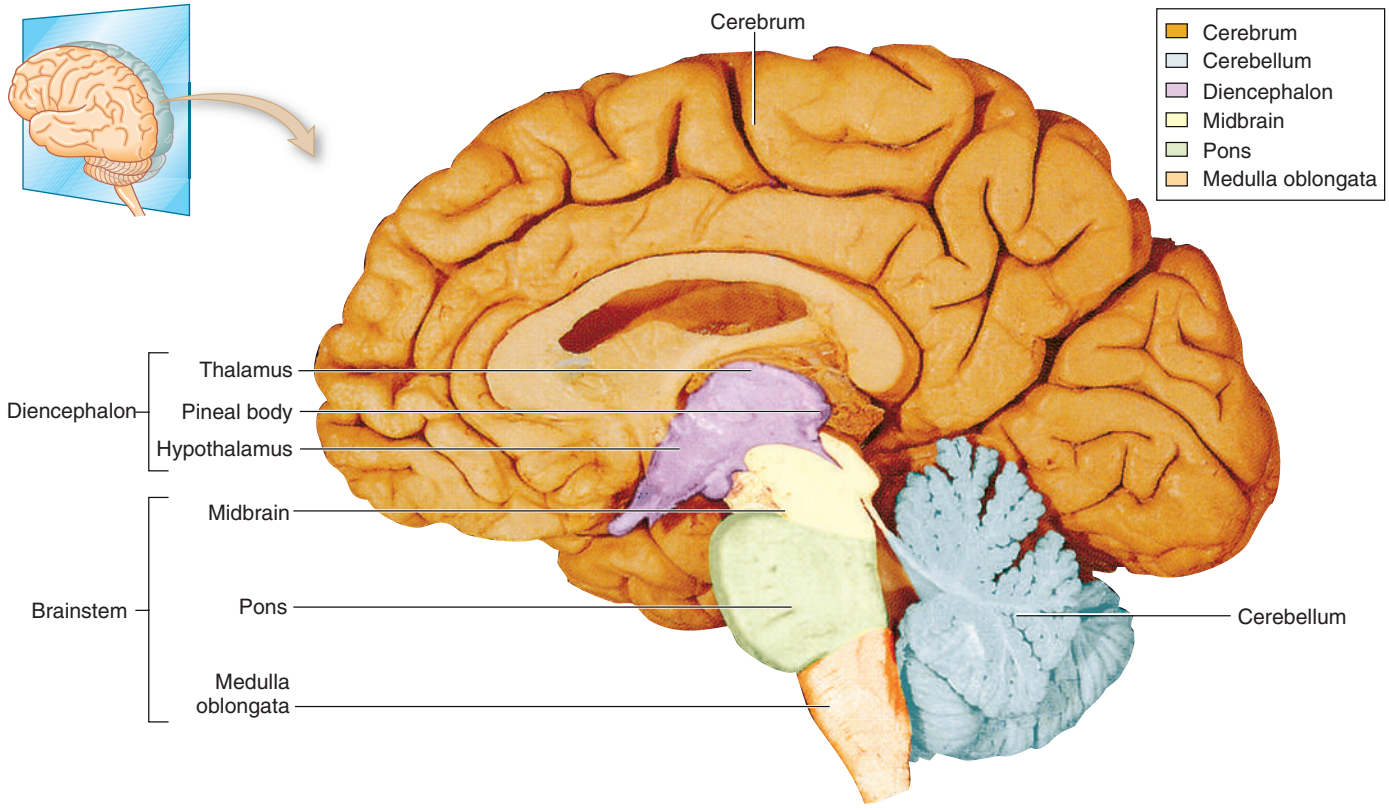


FIG 43.6 Four principal anatomic areas of the brain. (From Patton KT, Thibodeau GA: *Essentials of anatomy & physiology,* St Louis, 2012, Mosby.)

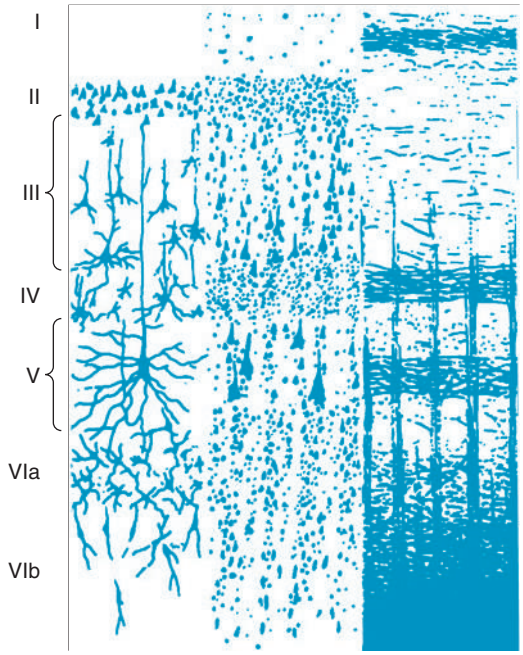


FIG 43.7 The cortex of the brain is histologically divided into six layers that differ in their connections to other parts of the nervous system. (From Ransom SW, Clark SL [after Brodmann]: *Anatomy of the nervous system,* Philadelphia, 1959, Saunders.)

information, the thalamus is integrally involved in executing motor activities. The thalamus also is involved in propagating the constant background electrical activity of the brain, which can be detected by electroencephalography. Connections between the brainstem reticular activating system and thalamus are necessary to maintain consciousness. Thalamic connections, including the limbic and association cortex, are integral to the expression of those qualities considered to be human: emotion, language, creativity, and complex thought.

The hypothalamus is located just beneath the thalamus on the floor of the diencephalon. The inferior aspect of the hypothalamus extends downward to form the pituitary gland (hypophysis). The posterior pituitary gland is an extension of the neuronal tissue of the hypothalamus, whereas the anterior pituitary gland is derived from glandular tissue (Fig. 43.13). Hormones secreted by the pituitary gland enter the systemic circulation and influence target cells at a distance. Neurons in the hypothalamus regulate the secretion of anterior pituitary hormones by releasing and inhibiting hormones (see Chapter 39 for a discussion of the endocrine system).

The hypothalamus is also an important regulatory center for the ANS and for basic functions, such as sleep, body temperature, appetite, and sex drive. Input from sensors of blood pressure, osmolarity, blood oxygen concentration, carbon dioxide level and pH, and temperature is received and integrated into appropriate regulatory responses. The hypothalamus is responsible for homeostasis of life-sustaining functions, including cardiovascular, respiratory, metabolic, fluid and electrolyte, and stress responses.

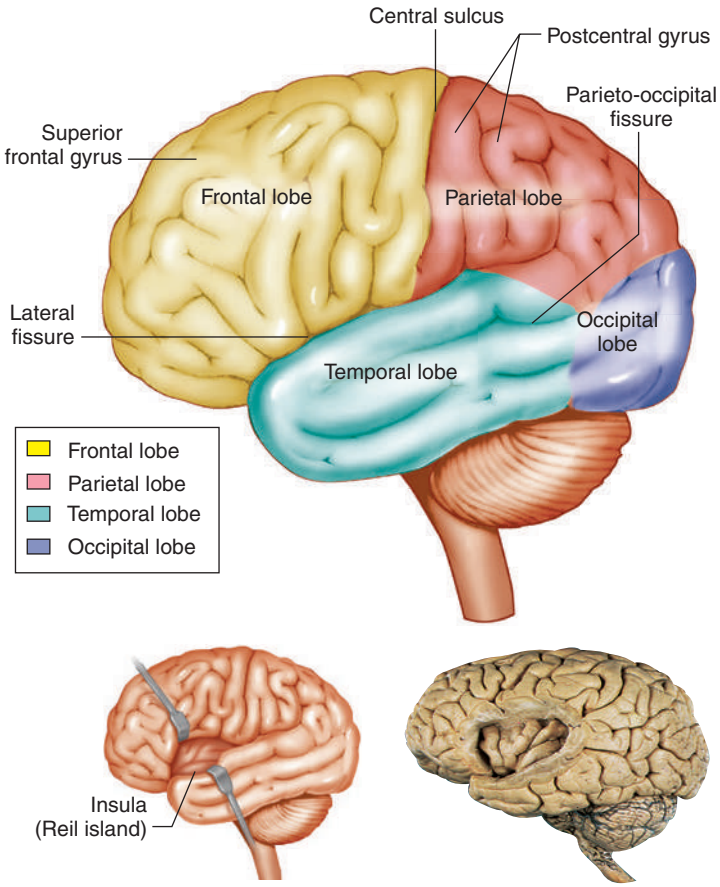


FIG 43.8 Four principal lobes of the cerebral cortex. (From Patton KT, Thibodeau GA: *Essentials of anatomy & physiology,* St Louis, 2012, Mosby.)

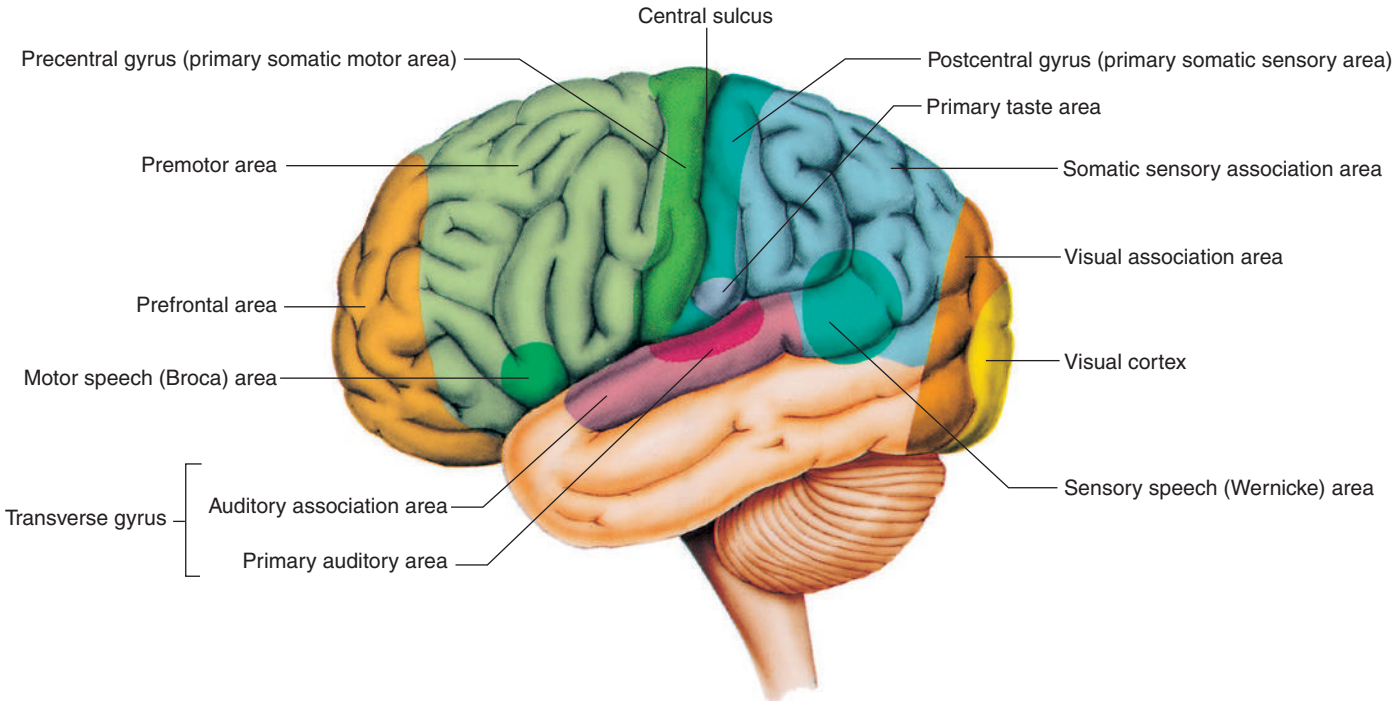


FIG 43.9 Partial Brodmann map of the cerebral cortex. Note the locations of Broca and Wernicke areas, which are important in the expression and understanding of language. (From Patton KT, Thibodeau GA: *Essentials of anatomy & physiology,* St Louis, 2012, Mosby.)

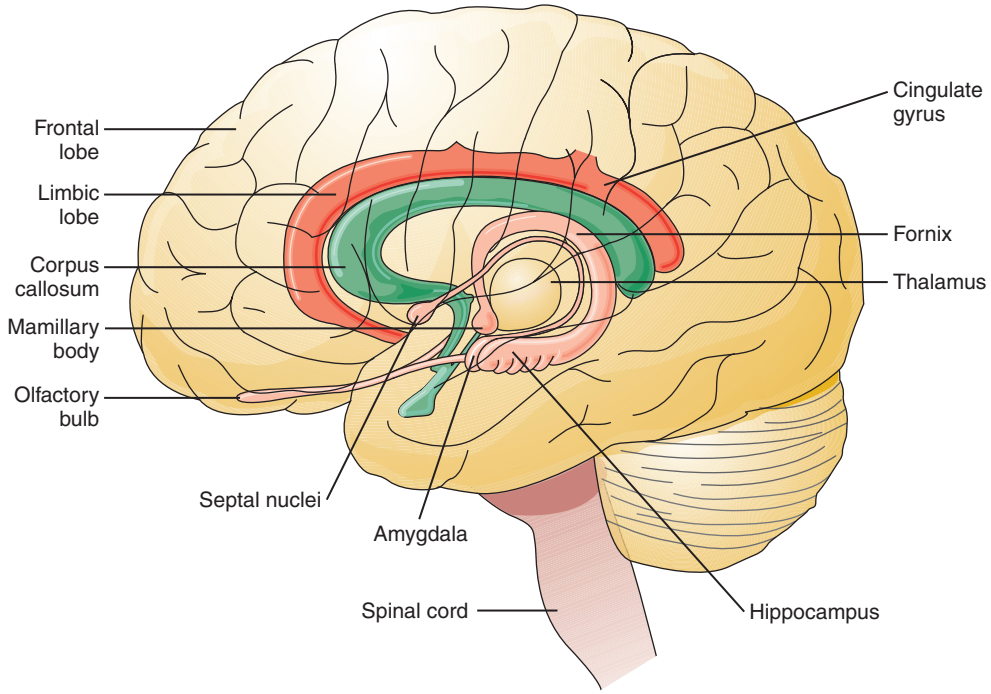


FIG 43.10 The limbic system is composed of a group of structures deep in the brain that are important in memory and emotion. These structures include the limbic lobe, amygdala, fornix, hippocampus, olfactory bulb, and portions of the thalamus.

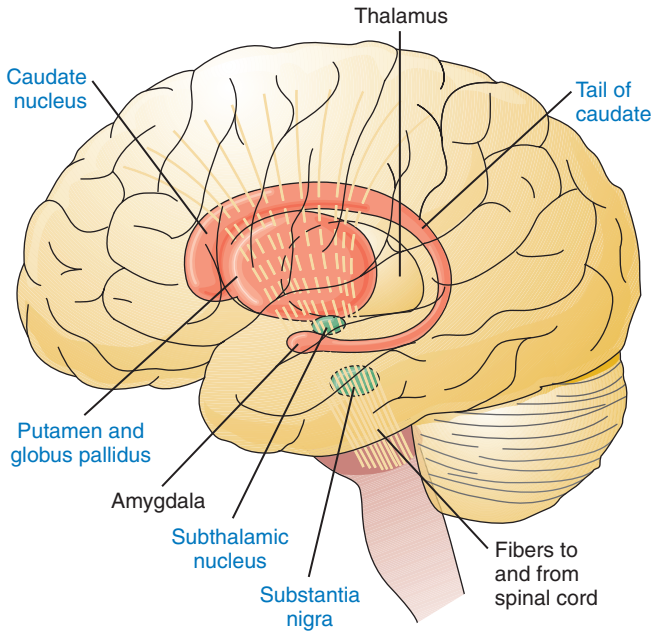


FIG 43.11 The basal ganglia include the caudate nucleus, putamen, globus pallidus, subthalamic nucleus, and substantia nigra (labeled in blue).

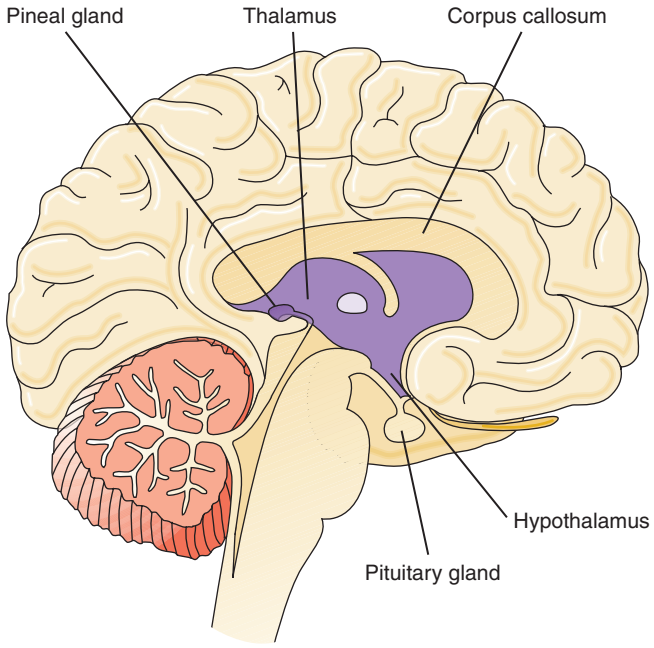


FIG 43.12 The diencephalon includes the thalamus, hypothalamus, pineal gland, and hypothalamic extension to the pituitary gland.

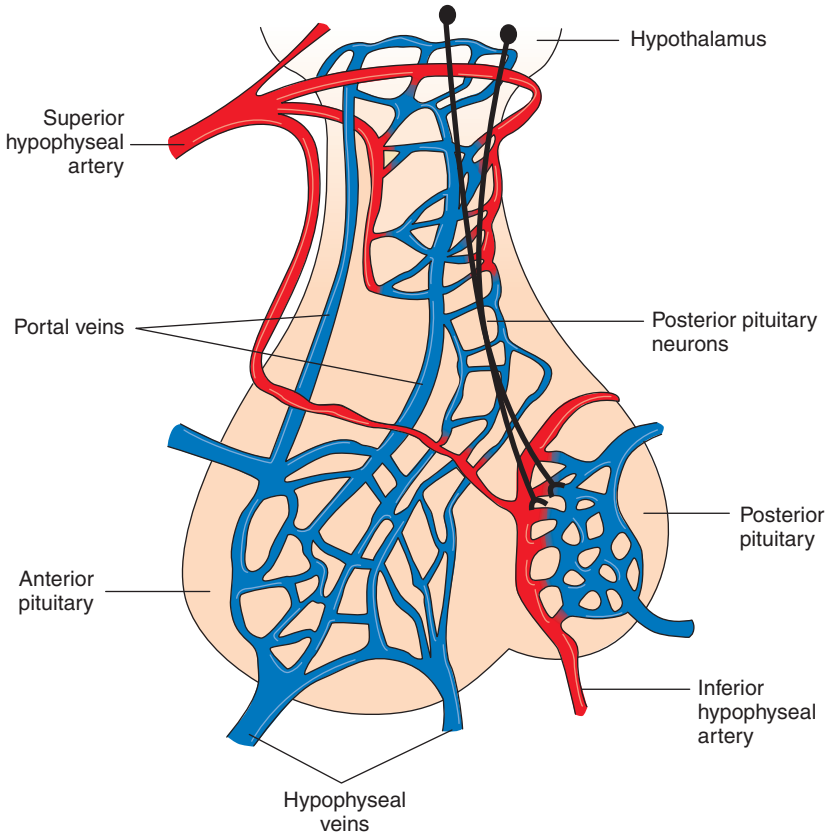


FIG 43.13 Anatomy of the hypothalamus and pituitary gland. Note that the posterior pituitary gland is connected to the hypothalamus by neuronal axons, whereas the anterior pituitary gland receives signals by way of a portal vein system. The portal veins drain blood from the capillaries of the hypothalamus and take it to the capillaries of the anterior pituitary gland.

The epithalamus contains the pineal gland, thought to be important in regulating circadian rhythms in response to light–dark cycles. The ventral thalamus contains the basal ganglia structure called the *subthalamic nucleus.*

Cerebellum

The cerebellum is located in the posterior fossa behind the pons. It is separated from the cerebrum by the tentorium cerebelli. The main roles of the cerebellum are to coordinate and smooth movements and to maintain posture and balance. The cerebellum compares the desired motor program with the moment-to-moment execution of the movement and makes instantaneous adjustments to improve the match. The cerebellum receives information from proprioceptors in muscles and joints and from the vestibular apparatus in the inner ear about the position of the head in space. Some of the fastest-conducting neurons in the nervous system are involved in relaying sensory information to the cerebellum.

The cerebellar cortex is folded much as the cerebral cortex is folded, in a way that significantly increases surface area. Its tightly folded shape gives it a banded appearance. The cortical ridges on the surface of the cerebellum are called *folia.* The white matter beneath is called the *medullary center* and is made up of fibers running to and from the cerebellar cortex.

The cerebellum is divided anatomically, first by the posterolateral fissure, which separates the flocculonodular lobe (the region immediately inferior to the middle cerebellar peduncles) from the main body (Fig. 43.14). The midline body is called the *vermis,* and it is straddled on either side by the cerebellar hemispheres.

The prominent tracts that attach the cerebellum to the brainstem are called the *inferior, middle,* and *superior* cerebellar peduncles. The inferior cerebellar peduncle is composed primarily of afferent fibers coming from the spinal cord and the brainstem. The middle peduncle contains afferent fibers from the contralateral pontine nuclei. The superior cerebellar peduncle is composed of major efferent pathways leaving the cerebellum.

Deep within the medullary center in each cerebellar hemisphere are the cerebellar nuclei. The deep cerebellar nuclei are the final pathway of cerebellar output. Input from several areas of the cerebral cortex is received by the cerebellar hemispheres and dentate nuclei and then sent back to the motor and premotor cortex. This circuit is believed to influence the planning and programming of voluntary movements, especially learned, skilled movements (those that become more rapid, precise, and automatic with practice).

The major input to the paravermal region, also called the *intermediate cortex,* consists of somatotopically arranged projections from the motor cortex and spinal cord. The intermediate cerebellum influences spinal cord and motor neurons through the corticospinal tract and the rubrospinal tract, where it is involved in interpreting and responding to the position and velocity of the moving body.

The vermis is most involved with regulation of posture and stereotyped movements that are programmed in the brainstem and spinal cord. The flocculonodular lobe helps maintain equilibrium and mediate the eye movements needed for visual tracking.

Lesions of the cerebellum result in ataxia (impaired balance), intention tremor, past pointing (failure of finger-to-nose test), and dysdiadochokinesia (failure of rapid movements).

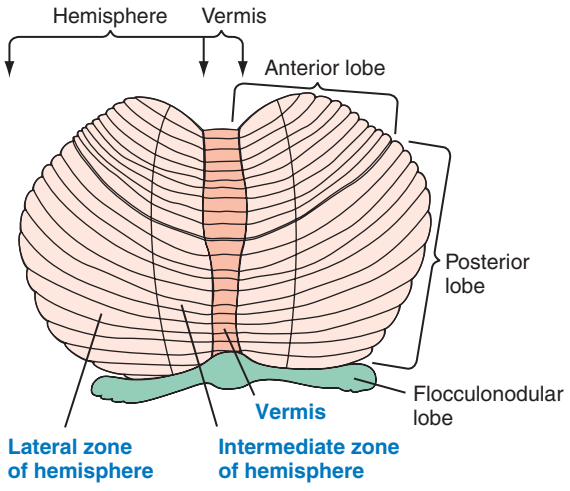


FIG 43.14 Lobes of the cerebellum from a posteroinferior view. See text for explanation. (Adapted from Hall JE, editor: *Guyton & Hall textbook of medical physiology,* ed 13, Philadelphia, 2015, Saunders, p 722.)

Brainstem

The brainstem is a stalk of neural tissue that lies between the upper spinal cord and the diencephalon. It has three parts: from top to bottom these are the *midbrain, pons,* and *medulla oblongata.* The brainstem is critical for transmission of impulses between the brain and spinal cord. Vital centers for regulating respiratory and cardiovascular function are located in the medulla and pons. In addition, the reticular activating neurons that maintain consciousness and alertness traverse the brainstem to reach the thalamus. Of the 12 pairs of cranial nerves, 10 originate from nuclei in the brainstem; only cranial nerves I (olfactory) and II (optic) originate elsewhere (diencephalon).

The midbrain, or mesencephalon, contains the cerebral peduncles, consisting of motor tracts to the spinal cord; the superior and inferior colliculi, which control head and eye movements; and the red nucleus, part of a major motor tract. Cranial nerve III (oculomotor) emerges from the midbrain and is prone to compression when pressure in one of the cerebral hemispheres is elevated. Increased intracranial pressure (e.g., from tumor, ischemia, edema, bleeding) is commonly manifested by dysfunction of cranial nerve III resulting in abnormal pupil size and poor reactivity to light (see Chapter 44). Cranial nerve IV (trochlear) also emerges at the level of the midbrain.

The pons (Latin for *bridge*) connects the midbrain above to the medulla below. The dorsal pons consists of reticular formation fibers, ascending sensory tracts, and descending motor tracts. Two respiratory centers (pneumotaxic and apneustic) located in the dorsal pons work in coordination with the principal respiratory centers in the medulla. A major pathway of voluntary motor control, the corticospinal tract, also passes through the ventral pons on its way from the motor cortex to the spinal cord.

The medulla oblongata makes up the lower third of the brainstem and is continuous with the spinal cord. Nuclei within the reticular formation of the medulla form the vital centers that regulate cardiac, vascular, and respiratory function. The medulla also contains centers that coordinate swallowing, vomiting, coughing, and sneezing. The medulla is the site of decussation (crossing over) of the major sensory (dorsal column) and motor (corticospinal) tracts such that innervation of one side of the body is connected to the opposite (contralateral) cerebral hemisphere. The corticospinal tract neurons decussate within ridges on the ventral surface of the medulla called *medullary pyramids.* Motor tracts that do not cross over within the pyramids (e.g., tectospinal, vestibulospinal) are sometimes referred to as *extrapyramidal tracts;* disorders associated with function of these tracts (balance, posture, gait) may be called *extrapyramidal disorders* (e.g., Parkinson disease). Although the anatomic correlation is not quite accurate, use of the term persists in a clinical context.

TABLE 43.4 Cranial Nerves

Cranial Nerve	Origin	Function
I (olfactory)	Nasal mucous membrane	Olfaction
II (optic)	Retina	Vision
III (oculomotor)	Midbrain	Movement of eyeball, eyelid, constriction of pupil
IV (trochlear)	Lower midbrain	Lateral eye movements
V (trigeminal)		
Ophthalmic	Forehead, eyes	Sensation from forehead, eye, scalp
Maxillary	Upper jaw, lip	Sensation from cheek, upper lip
Mandibular	Lower jaw area	Sensation from chin and lower jaw, motor chewing
VI (abducens)	Lower pons	Lateral eye movements
VII (facial)	Pons	Taste from anterior tongue, control of muscles of face
VIII (vestibulocochlear)	Cochlea	Hearing
	Inner ear	Equilibrium
IX (glossopharyngeal)	Medulla	Taste from posterior tongue, secretion of saliva, swallowing
X (vagus)	Medulla	Monitors oxygen, carbon dioxide, and pH levels in blood; senses blood pressure; inhibits cardiac action and has extensive gastrointestinal activities
XI (spinal accessory)	Medulla and cervical cord	Voice production, movement of head and shoulders
XII (hypoglossal)	Medulla	Movements of tongue during speech and swallowing

All of the remaining cranial nerves (VI, VII, VIII, IX, X, XI, and XII) originate in the medulla. The cranial nerves themselves are part of the PNS and are discussed in that section. The name, origin, and function of the 12 cranial nerves are included in Table 43.4.

The Spinal Cord

The spinal cord conveys nervous impulses between the brain and 31 pairs of spinal nerves that innervate sensory organs and muscle cells of the body. The spinal cord mediates spinal reflexes involved in maintenance of posture, protective responses to pain, urination, and muscle tone. A great deal of integration and processing occurs in the gray matter of the spinal cord, whereas the white matter contains bundles of myelinated axons forming tracts that run up and down the cord. Tracts in the spinal cord are somatotopically organized such that the innervation of a particular body region is connected to a specific region in the cerebral cortex.

The typical adult spinal cord is about 18 inches long, extending from the base of the skull (foramen magnum) to the first or second lumbar vertebra (L1 to L2). The vertebral column extends for several more inches, providing a reservoir for CSF and exit points for the lumbar and sacral spinal nerves. The vertebral column is formed by interlocking sections of bone separated and cushioned by intervertebral disks. At the lateral aspect of the intersection of two vertebrae is an opening (intervertebral foramen) that provides a passageway for spinal nerves to exit the cord. The spinal cord travels in a small lumen (1 cm) in the center of the vertebral column and is itself only slightly larger than the diameter of a pencil (Fig. 43.15).

On cross-section the spinal cord has a butterfly pattern of gray matter surrounded by white matter (Fig. 43.16). Three bumps on the butterfly wings are called *horns:* the ventral horn (motor neurons), the dorsal horn (sensory neurons), and the lateral horn (sympathetic neurons). The horns consist of neuron cell bodies, synapses, and small unmyelinated interneurons. The white matter is divided into columns that contain tracts of nerve fibers traveling to and from the brain. These are the posterior (dorsal) columns, anterior columns, and lateral columns. Some of the neurons in the columns convey signals from one level of the cord to another and are important in reflex and postural adjustments. The principal ascending sensory tracts include the dorsal column–lemniscal and the anterolateral (spinothalamic) tracts, which send afferent signals to the brain. The principal descending motor tracts include the corticospinal, rubrospinal, reticulospinal, and vestibulospinal tracts. These tracts are located in specific regions of the cord (Fig. 43.17). Sensory and motor pathways are discussed in later sections of this chapter.

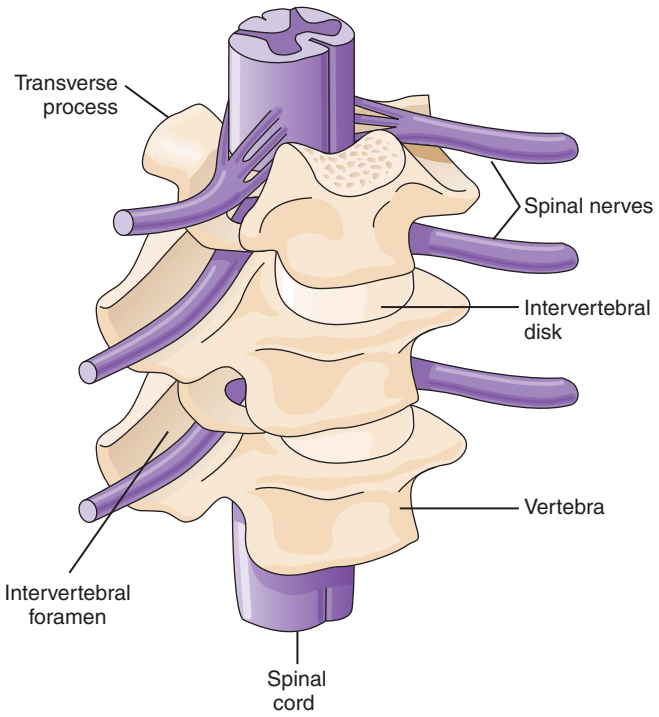


FIG 43.15 Spinal cord travels down the center of the vertebral column. A foramen at the intersection of two vertebrae forms an exit point for the spinal nerves.

Spinal nerves divide into two sections as they make contact with the spinal cord: the ventral and dorsal roots (Fig. 43.18). Ventral roots contain motor neurons that originate in the anterior horn and travel in the spinal nerve to skeletal muscles. Dorsal roots carry sensory information from somatic receptors to neurons in the posterior horn.

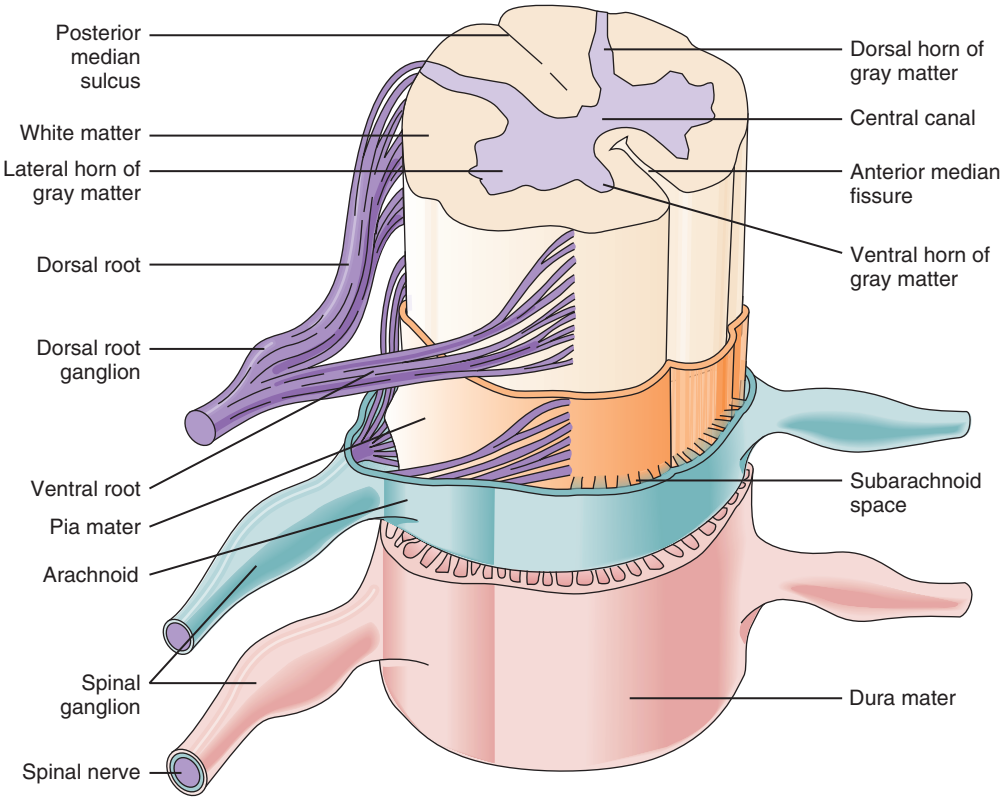


FIG 43.16 Spinal cord in cross-section, showing the butterfly pattern of white and gray matter.

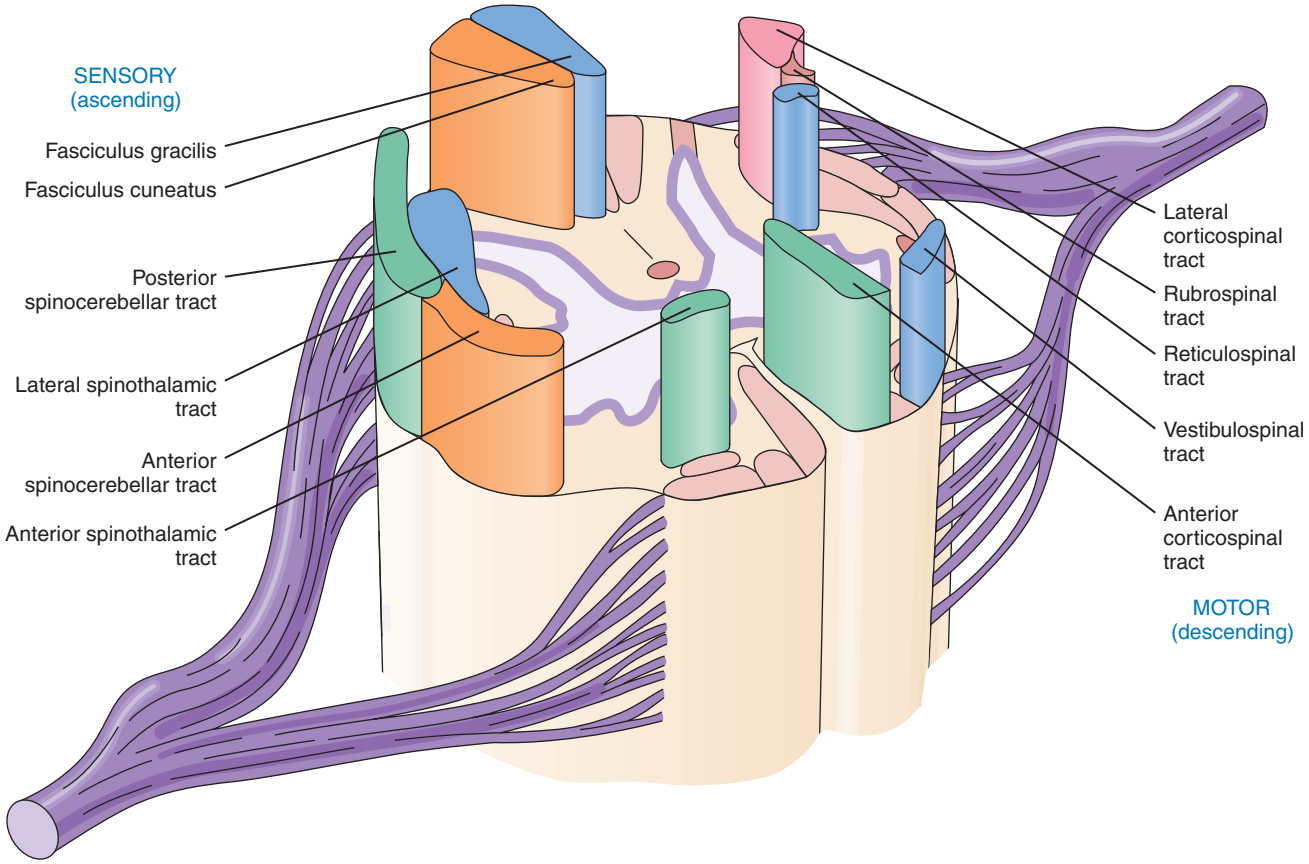


FIG 43.17 Main ascending *(left)* and descending *(right)* tracts of the spinal cord.

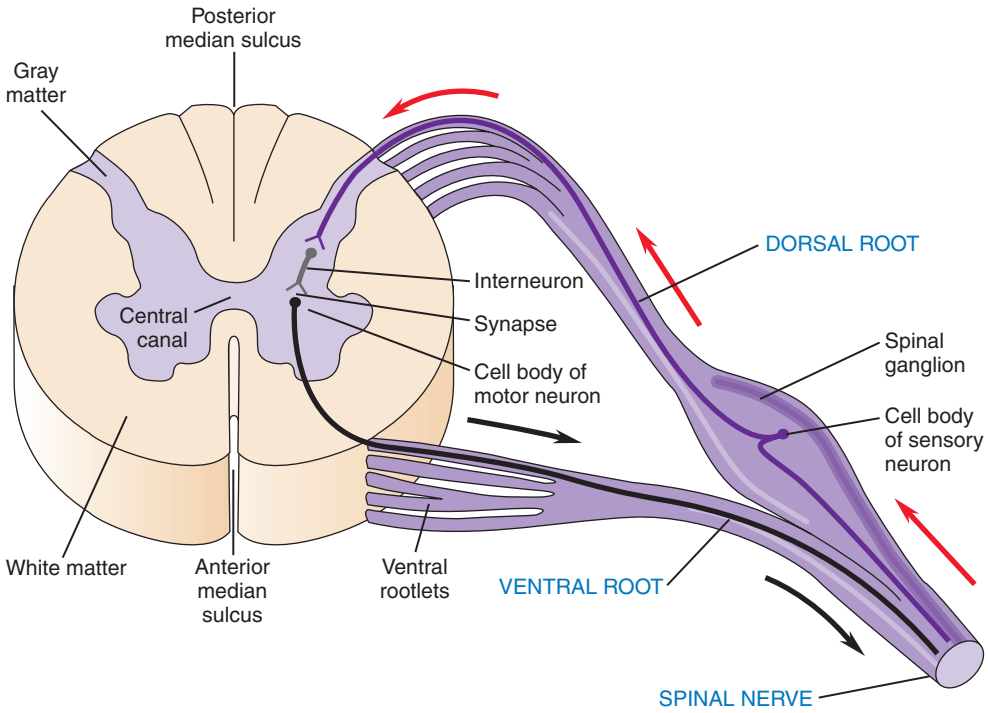


FIG 43.18 Spinal nerves split to form dorsal and ventral roots as they emerge from the spinal cord. Ventral roots carry motor efferent neurons, whereas dorsal roots carry sensory afferent neurons. See text for explanation.

The cell bodies of sensory afferents collect together in the dorsal root ganglion. Autonomic nerves also travel in the spinal cord and exit and enter the cord by way of the ventral and dorsal roots.
The points at which sensory neurons enter the cord and at which motor neurons exit represent the separation of the CNS and PNS.

PERIPHERAL NERVOUS SYSTEM

The PNS consists of the 31 pairs of spinal nerves and the 12 pairs of cranial nerves. These nerves are myelinated with Schwann cells, which differ somewhat from the oligodendrocytes that form the myelin sheaths of CNS neurons. By convention, groups of cell bodies are called *ganglia* in the PNS and *nuclei* in the CNS. A major exception to this naming rule is the basal ganglia of the CNS. The PNS is not protected by CSF, meninges, or bony coverings as is the CNS; however, a sheath of connective tissue covers the nerves and provides support.

The PNS serves both afferent sensory functions and efferent motor functions of the somatic and autonomic systems. Cranial nerves III, VII, IX, and X and spinal nerves S2 and S3 contain parasympathetic neurons, and spinal nerves T1 to L2 contain sympathetic neurons.

Cranial Nerves

As previously noted, all of the cranial nerves originate in the brainstem except cranial nerves I and II, which originate in the diencephalon (Fig. 43.19). Cranial nerve I is strictly sensory, transmitting olfactory signals from the 10 million to 20 million olfactory neurons in the nasal cavities to the olfactory bulbs. The olfactory bulb neurons then project to the olfactory cortex. Cranial nerve II is also sensory, conveying visual information from the retina to the brain. The optic nerve is unusual in that it is an extension of the CNS, myelinated by oligodendrocytes rather than Schwann cells. The neurons from the medial retina decussate in the optic chiasm, whereas the lateral retina neurons do not. Thus the right visual field projects to the left hemisphere and the left visual field projects to the right hemisphere. Damage to one hemisphere, as occurs in stroke, often interrupts visual signals from the corresponding sides of each retina—a condition known as *homonymous hemianopsia* (see Chapters 44 and 46).

Cranial nerves III, IV, and VI innervate motor structures in the eyes. Cranial nerve III mediates pupil constriction. The trigeminal nerve (cranial nerve V) is so named because it has three branches, which provide sensory innervation of the forehead and eyes (ophthalmic branch); upper lip, teeth, and palate (maxillary branch); and lower jaw (mandibular branch). The mandibular branch is both sensory and motor. Cranial nerve VII is also mixed sensory and motor, detecting taste in the anterior two-thirds of the tongue and innervating muscles of facial expression. Cranial nerve VII also contains autonomic fibers that innervate salivary and lacrimal (tear) glands.

Cranial nerve VIII has two important sensory functions: transmitting auditory information from the cochlea and vestibular information from inner ear structures. The vestibular neurons of cranial nerve VIII interact with the neurons of cranial nerves III and VI to reflexively control eye movements during head rotation such that a visual image can remain fixed on the retina. This reflex, the oculovestibular reflex, is commonly assessed in the unconscious patient to evaluate brainstem function (see Chapter 44).

Cranial nerve IX innervates tongue and pharyngeal muscles, conveying taste from the posterior tongue and controlling pharyngeal motion during swallowing. Cranial nerve IX also has autonomic functions and transmits sensory information from the carotid baroreceptors and carotid bodies to the brainstem. Cranial nerve X, the vagus nerve, contains parasympathetic afferent and efferent fibers that innervate many visceral structures, including the heart, lungs, and gastrointestinal (GI) tract from pharynx to anus. Sensory information from aortic baroreceptors and aortic bodies is conveyed to the brainstem by the vagus nerve.

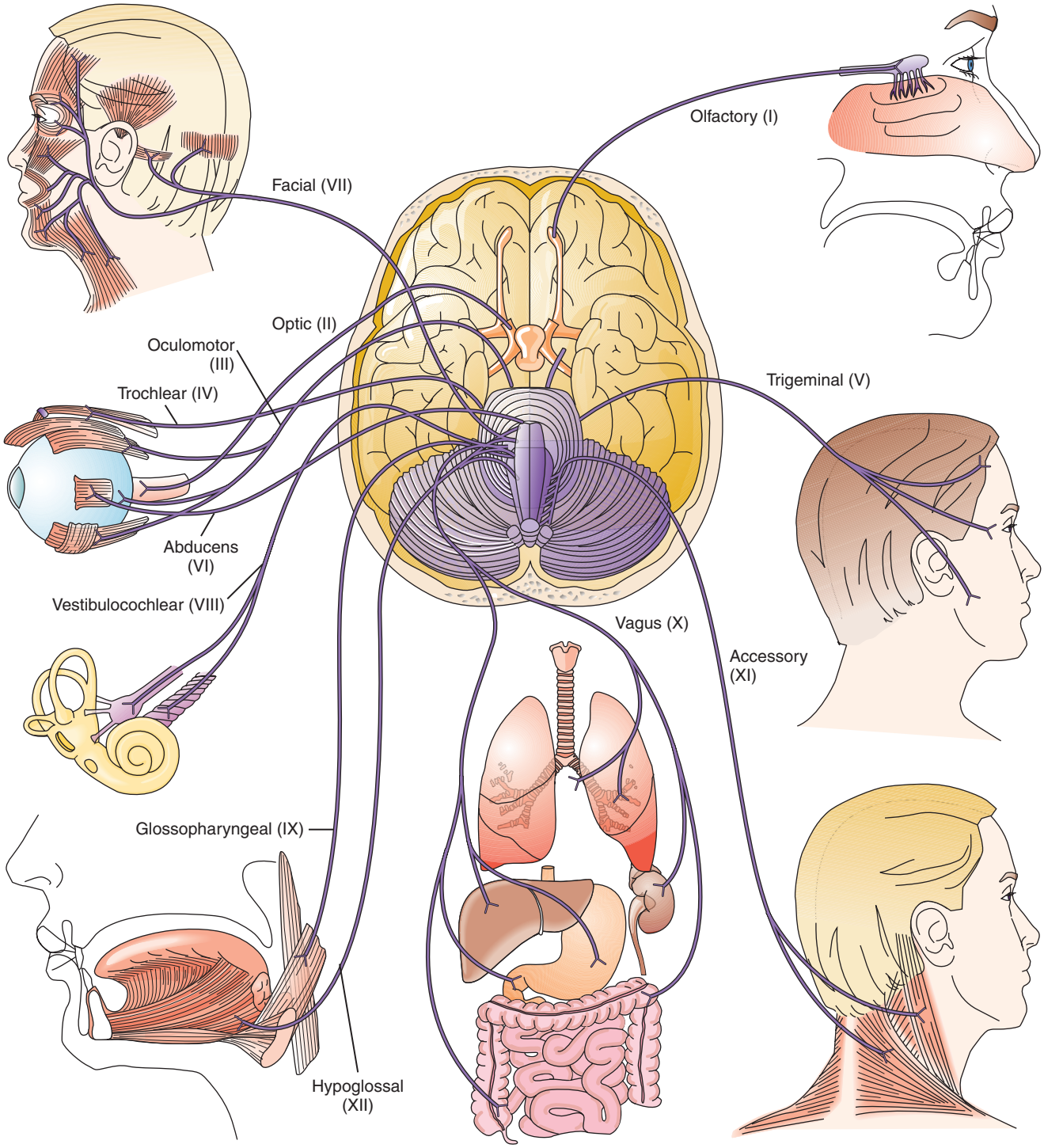


FIG 43.19 View of the inferior aspect of the brain showing the origin and distribution of the 12 cranial nerves. Only one of each pair is shown.

Cranial nerve XI innervates muscles of the larynx, neck, and shoulders and mediates voice production and neck and head movements. Cranial nerve XII innervates tongue muscles and controls their action during speech and swallowing.

Spinal Nerves

The 31 pairs of spinal nerves are named after the vertebral segments from which they emerge. There are 8 cervical, 12 thoracic, 5 lumbar, 5 sacral, and 1 coccygeal pair of spinal nerves (Fig. 43.20). The first cervical nerve exits above C1, whereas the others all exit below the vertebral segment; thus there is one more pair of cervical spinal nerves (8) than there are cervical vertebrae (7).

Except for spinal nerves T2 to T12, the spinal nerves travel a distance from the cord and then merge into a large group called a *plexus*. In the plexus, nerve fibers are recombined into different groups and emerge as peripheral nerves (Fig. 43.21). There are five plexuses:

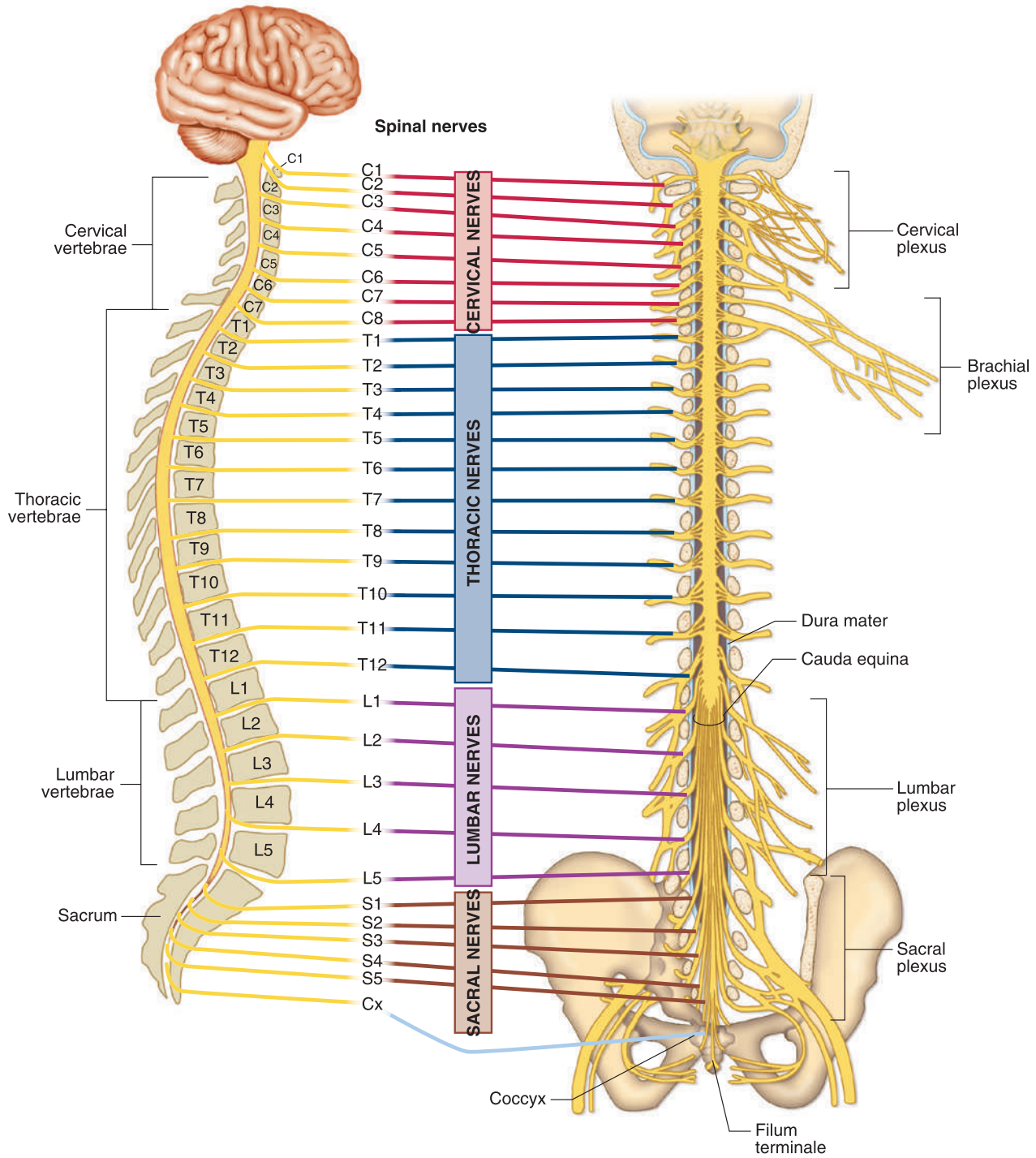


FIG 43.20 Spinal nerves. Each of the 31 pairs of spinal nerves exits the spinal cavity from the intervertebral foramina. Note that after leaving the spinal cavity, many of the spinal nerves interconnect to form complex networks called *plexuses* (shown on the right). (From Patton KT, Thibodeau GA: *Essentials of anatomy & physiology,* St Louis, 2012, Mosby.)

(1) the cervical plexus (C1 to C4), (2) the brachial plexus (C5 to C8, T1), (3) the lumbar plexus (L1 to L4), (4) the sacral plexus (L4 to L5, S1 to S3), and (5) the coccygeal plexus (S4 to S5, coccygeal) (Table 43.5). Because of this recombination of nerve fibers in the plexus, the spinal nerves and peripheral nerves have different somatic distributions. The segment of the body innervated by a spinal nerve is called a *dermatome,* whereas the peripheral nerve innervates a peripheral nerve field. Knowledge of dermatomes and peripheral nerve distribution can help the clinician differentiate between radiculopathy from spinal nerve compression (dermatomal sensory changes) and peripheral neuropathy. (Dermatomal and peripheral nerve maps are located in Chapter 47.)

TABLE 43.5 Spinal Nerve Plexuses

Plexus	Spinal Nerves	Peripheral Nerves	Distribution
Cervical	C1–C4	Phrenic	Diaphragm
		Cutaneous	Neck
		Ansa cervicalis	Hyoid bone
Brachial	C5–C8, T1	Axillary	Upper arm
		Ulnar	Forearm, wrist, fifth digit
		Median	Second through fourth digits
		Radial	Thumb
		Musculocutaneous	Upper arm
Lumbar	L1–L4	Femoral (saphenous)	Hip, knee (lower leg)
		Obturator	Inner thigh
Sacral	L4–L5, S1–S3	Superior gluteal	Gluteus medius, minimus
		Inferior gluteal	Gluteus maximus
		Sciatic (tibial, peroneal)	Thigh, leg, foot
		Pudendal	Perineum
Coccygeal	S4–S5, coccygeal	Coccygeal fibers	Skin on coccyx

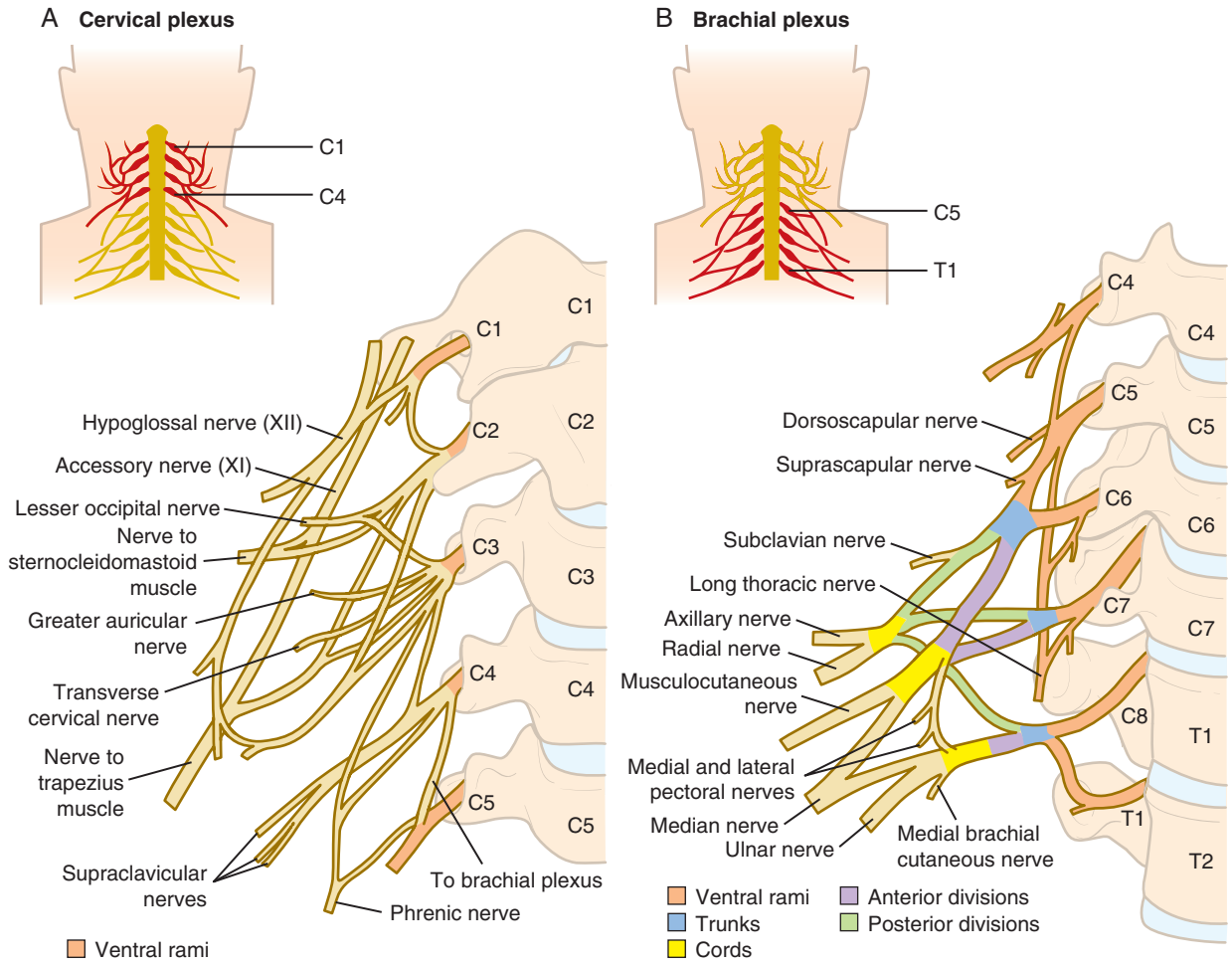


FIG 43.21 Examples of neuronal organization at the plexuses showing reassortment of spinal nerves into peripheral nerves. **A,** Cervical plexus. **B,** Brachial plexus.

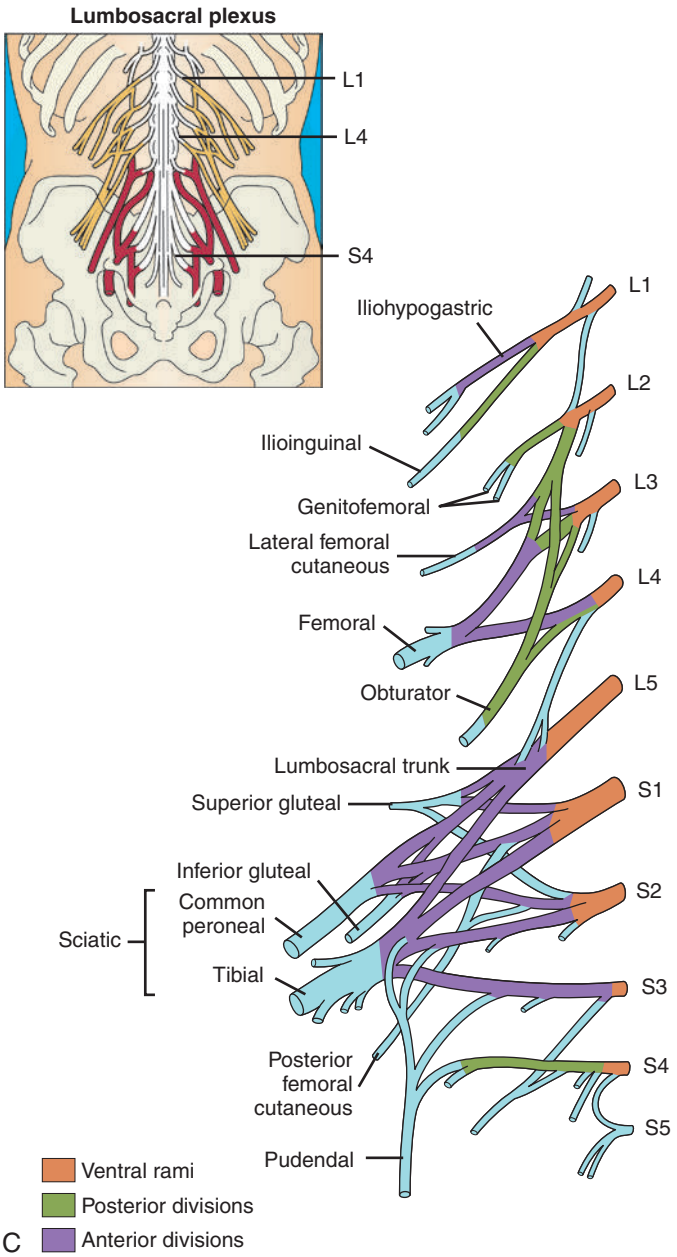


FIG 43.21, cont'd C, Lumbosacral plexus (because of their overlap, the lumbar and sacral plexuses are often considered together as in this example).

The intercostal nerves (T2 to T12) do not form plexuses; they travel in a course parallel to the ribs to innervate intercostal muscles and skin on the trunk and abdomen.

AUTONOMIC NERVOUS SYSTEM

The ANS is composed of neurons in the CNS and PNS that mediate automatic or involuntary functions. The ANS has both sensory afferents and motor efferents that primarily innervate visceral organs and blood vessels. As previously described, the hypothalamus and brainstem contain neurons responsible for integrating autonomic sensory information and creating appropriate homeostatic responses. This response is communicated to the effector organs by parasympathetic nervous system (PSNS) efferents located in cranial nerves III, VII, IX, and X; by spinal nerves S2 to S3; and by sympathetic nervous system (SNS) neurons in spinal nerves T1 to L2.

The distribution of parasympathetic nerves is shown in Fig. 43.22. Note the extensive role that cranial nerve X (vagus) has in cardiovascular, respiratory, and GI function. The distribution of sympathetic nerves is shown in Fig. 43.23. Note that after leaving the spinal cord, sympathetic neurons converge on a chain of ganglia that runs parallel to both sides of the spinal cord. Some sympathetic neurons synapse on secondary neurons in the ganglia, and others travel to other plexuses or ganglia before synapsing (see Fig. 43.23). The neurons that emerge from the spinal cord are called *preganglionic neurons,* whereas the neurons traveling to the target cell are called *postganglionic neurons.* This terminology is also used for the PSNS; however, the parasympathetic preganglionic neurons are long, traveling all the way to the target organ, and they do not terminate in ganglia. The postganglionic neurons are short and are located within the target organ (see Fig. 43.22).

The neurotransmitter secreted by preganglionic neurons is acetylcholine (ACh) for both the SNS and the PSNS. The postganglionic neurotransmitters differ (Fig. 43.24). The SNS secretes norepinephrine (NE) in most cases, although sweat glands and some skeletal muscle vessels are innervated by ACh-secreting SNS neurons. The PSNS secretes ACh as the postganglionic neurotransmitter. ACh is also the neurotransmitter of the motor neurons that innervate skeletal muscle; however, the target cell receptors differ. Skeletal muscle contains nicotinic ACh receptors, whereas autonomic organs contain muscarinic ACh receptors (see Fig. 43.24).

The effect of the SNS and PSNS on target organs is nearly always antagonistic. If one contracts smooth muscle, the other relaxes it; if one stimulates glandular secretion, the other inhibits it; if one speeds up a process, the other slows it down. The effects of SNS and PSNS stimulation on major target organs are shown in Table 43.6.

A specialized extension of the SNS is found in the adrenal gland. The adrenal medulla receives preganglionic neurons from SNS neurons emerging from the spinal cord, which stimulate the gland to secrete epinephrine (and smaller amounts of NE) into the bloodstream. These hormones have effects similar to those of direct SNS stimulation.

The manner in which target cells respond to SNS stimulation depends on the types of receptors they possess. Several subtypes of receptors bind and respond to NE and epinephrine; these include α_1, α_2, β_1, β_2, β_3, and several DA receptors (see Table 43.6). (The details of autonomic regulation of cardiac, genitourinary, and GI function can be found in Chapters 17, 29, and 35, respectively.)

The coordination of SNS and PSNS activity within a target organ is accomplished by centers in the brainstem and hypothalamus with input from sensory neurons and cortical neurons. Most of these systems work on a negative feedback principle to achieve homeostasis. Negative feedback requires accurate sensory input about the conditions being regulated. Much of this feedback is provided by the vagus nerves that obtain extensive sensory information from receptors in the GI tract and aorta. This sensory input is processed in lower brain centers and does not reach the level of perception.

KEY POINTS

- The nervous system can be divided into three principal systems: (1) the central nervous system (CNS), consisting of the brain and spinal cord; (2) the peripheral nervous system (PNS), consisting of 31 pairs of spinal nerves and 12 pairs of cranial nerves; and (3) the autonomic nervous system (ANS), consisting of the sympathetic and parasympathetic branches.
- Meninges affix the brain to the skull so that the brain is suspended and supported. Meninges have three layers: (1) The dura mater is the tough

outer layer attached to the periosteum of the skull. (2) The arachnoid is a delicate weblike membrane spanning the space between the dura mater and the pia mater. (3) The pia mater covers the contours of the brain surface. The spinal cord has a similar arrangement of meningeal coverings.

- Cerebrospinal fluid (CSF) is produced in the brain ventricles and circulates in the subarachnoid spaces, providing cushioning and nutritive functions.
- The brain is protected by specialized tight junctions between the cells of the capillary endothelium (blood–brain barrier) and between the ependymal cells that line the ventricles (CSF–brain barrier).
- The brain can be anatomically divided into four principal structures: (1) the cerebrum (cerebral cortex, basal ganglia, limbic cortex, and corpus callosum); (2) the diencephalon (thalamus and hypothalamus); (3) the cerebellum; and (4) the brainstem (midbrain, pons, and medulla).
- Certain cortical areas are closely associated with specific functions: the frontal lobe contains the motor cortex and is involved in complex thought, motivation, and morality; the temporal lobe contains the auditory and vestibular centers and parts of the language center; the occipital lobe contains the visual cortex; the parietal lobe contains the somatosensory cortex; the limbic area is involved in memory and emotion.
- Basal ganglia are located deep within the cerebral hemispheres and are important in the control of skeletal muscles.
- The thalamus is a centrally located structure that processes and relays most of the signals traveling to and from the cortex and lower centers. Connections between the thalamus and the brainstem and cortex are needed to maintain consciousness and allow higher brain functions.
- The hypothalamus and brainstem are important structures regulating the ANS. The sympathetic nerves originate in spinal cord segments T1 to L2. The parasympathetic nerves emerge from the sacral segments and also travel in cranial nerves III, VII, IX, and X.

NEURONAL STRUCTURE AND FUNCTION

The ways in which the nervous system achieves its rapid communication function can be understood by examining the structure and behavior of neurons and neuronal synapses.

NEURONS AND SUPPORTIVE CELLS

The nervous system is composed of two principal cell types: neurons, which generate and transmit nerve impulses, and glial cells, which provide supportive functions to neurons but do not transmit action potentials. There are approximately 10 glial cells per neuron and about 100 billion neurons in the CNS.

Neurons

A neuron has three basic components: (1) the cell body containing cellular organelles, (2) the dendrites that receive signals and conduct them to the cell body, and (3) the axon that generates and conducts action potentials. Neurons can be categorized according to their structure or by the neurotransmitters they secrete. Neurons are present in three basic configurations based on the location of the cell body and the relative length and number of dendrites and axons (Fig. 43.25).

Multipolar neurons have a large number of dendrites extending from the cell body and one axon. Most neurons are of this type. Bipolar neurons have only one dendrite and one axon extending from the cell body. These neurons are prevalent in the retina, cochlea, and olfactory structures but are rare elsewhere. Unipolar neurons have a single process protruding from the cell body, which splits to form a dendrite and axon. This arrangement makes the cell body appear to be off-center. Unipolar neurons are prevalent in the somatosensory nerves in which the cell bodies are grouped in the dorsal root ganglia. The dendrites extend to the sensory receptors, and the axons enter the spinal cord.

Neurons also can be grouped as excitatory or inhibitory based on the nature of the neurotransmitter they secrete. Each neuron secretes one principal neurotransmitter, which is excitatory if it depolarizes the target neuron or inhibitory if it results in hyperpolarization.

Glia

Glial cells in the nervous system (neuroglia) serve a number of supportive functions, but they are not capable of generating action potentials. Four major types of neuroglia are recognized: oligodendrocytes, astrocytes, microglia, and ependymal cells (Fig. 43.26). The Schwann cells of the PNS are similar to oligodendrocytes. Both of these cell types form the myelin sheath that wraps around nerve axons to insulate and speed the rate of action potential conduction (Fig. 43.27). Myelin gives the white matter its color.

Astrocytes serve many functions in the CNS. Some astrocytes have foot processes that contact the brain capillaries and help maintain the integrity of the BBB. Astrocytes regulate ionic balance of the interstitial fluid and may influence the transfer of nutrients from capillaries to neurons. Astrocytes also participate in nervous system signaling and have been shown to take up and release molecules, such as neurotransmitters and cotransmitters that modulate neurotransmission.

Microglia are derived from the monocyte–macrophage cell type and provide phagocytic functions within the CNS. Ependymal cells line the ventricles and central canal of the spinal cord, producing CSF and maintaining the CSF–brain barrier.

Terminally differentiated neurons are not capable of cell division and cannot replace themselves if they die. However, certain areas of the brain, particularly the hippocampus and ventricles, are populated by neural stem cells. These cells are capable of cell division to produce two daughter cells, one of which retains stem cell characteristics, whereas the other may differentiate into a neuron or glial cell (Fig. 43.28). Specific signals are thought to guide the new cell as it migrates to the brain tissue and begins differentiation. Approximately half of the newborn cells will not find a suitable place and undergo apoptosis (programmed cell death). The discovery that neural stem cells provide a reservoir for producing new neurons has numerous implications for treating neurodegenerative disorders such as Alzheimer disease and Parkinson disease, as well as brain damage after stroke and trauma. However, the methods for stimulating proliferation and coaching the neurons to migrate to the right places to make the correct synaptic connections have not been completely elucidated.

The term *neural plasticity* is used to describe the potential for the brain to change its structure and function. Traditionally, neural plasticity was thought to be a result of recruitment of formed neurons into new functional networks. With the discovery of neural stem cells, plasticity in some regions of the brain is likely to include the addition of new neurons, as well as reassignment of the participants in the neuronal circuits. Neural plasticity is used to train different brain areas to assume new functions. For example, when a person suffers a stroke causing destruction of neurons in the motor cortex, it is possible to train nearby cortical neurons to assume some of the lost motor functions. Persistent attempts to use the muscles in an affected area may recruit cortical neurons into the neuronal circuit and improve motor strength and coordination over time. Neural plasticity is a fundamental process that endows the brain with the potential for memory and learning. Within certain boundaries, it appears that greater exposure to a particular stimulus prompts the brain to dedicate more neurons to that stimulus (and a lack of stimulation allows the brain to reassign neurons to a different function)—and so the old adage "use it or lose it" appears to

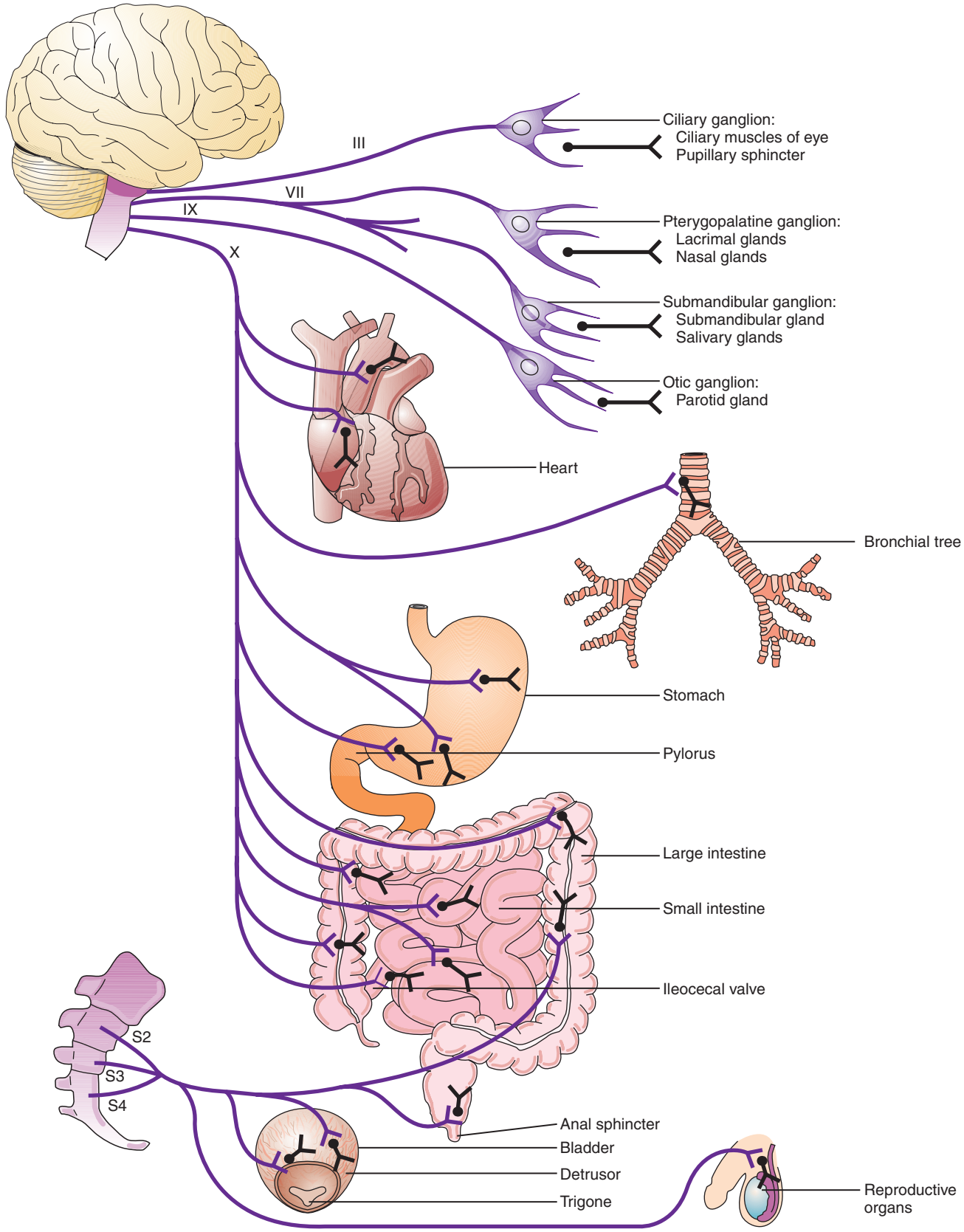


FIG 43.22 Distribution of parasympathetic nerves.

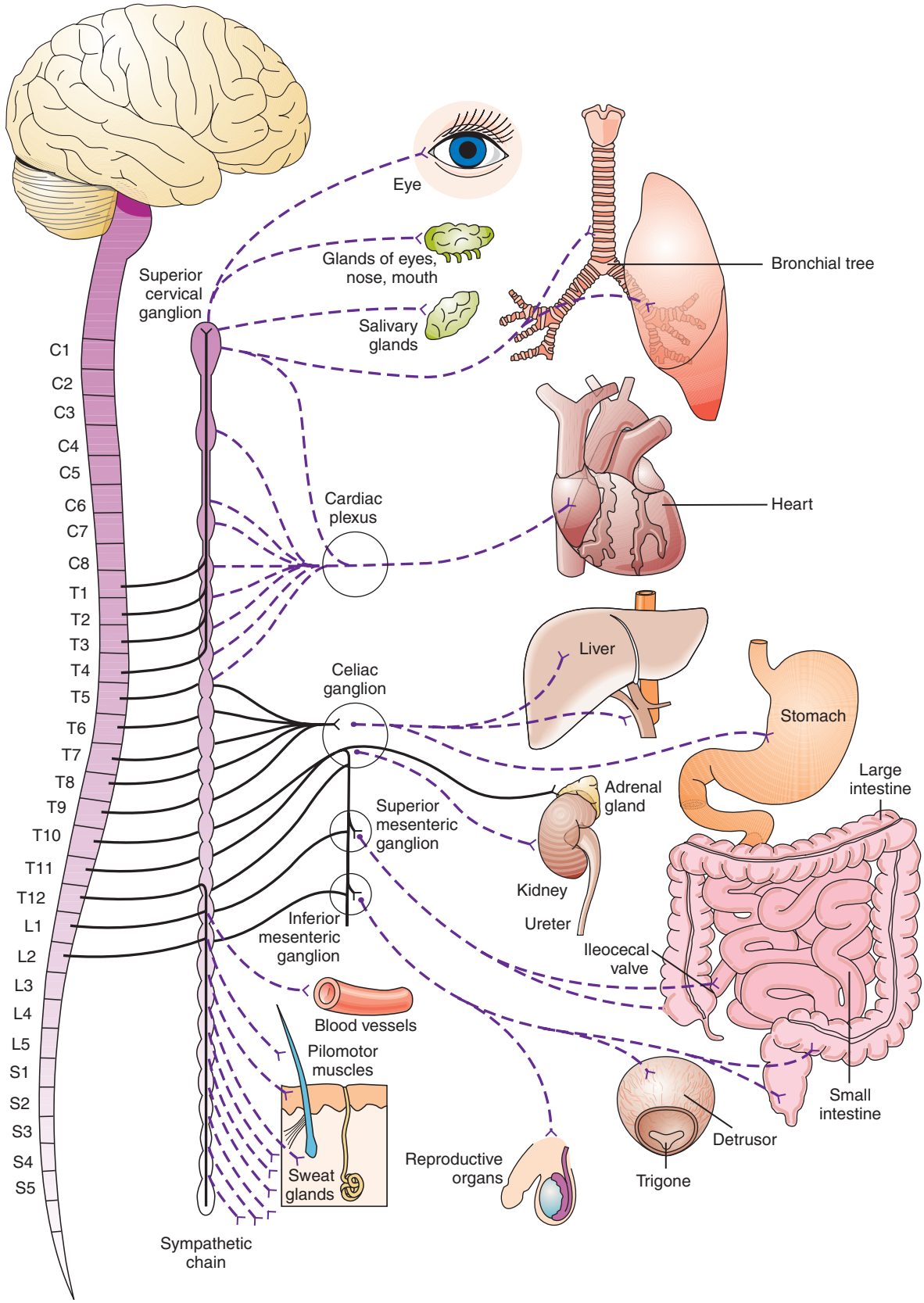


FIG 43.23 Distribution of sympathetic nerves.

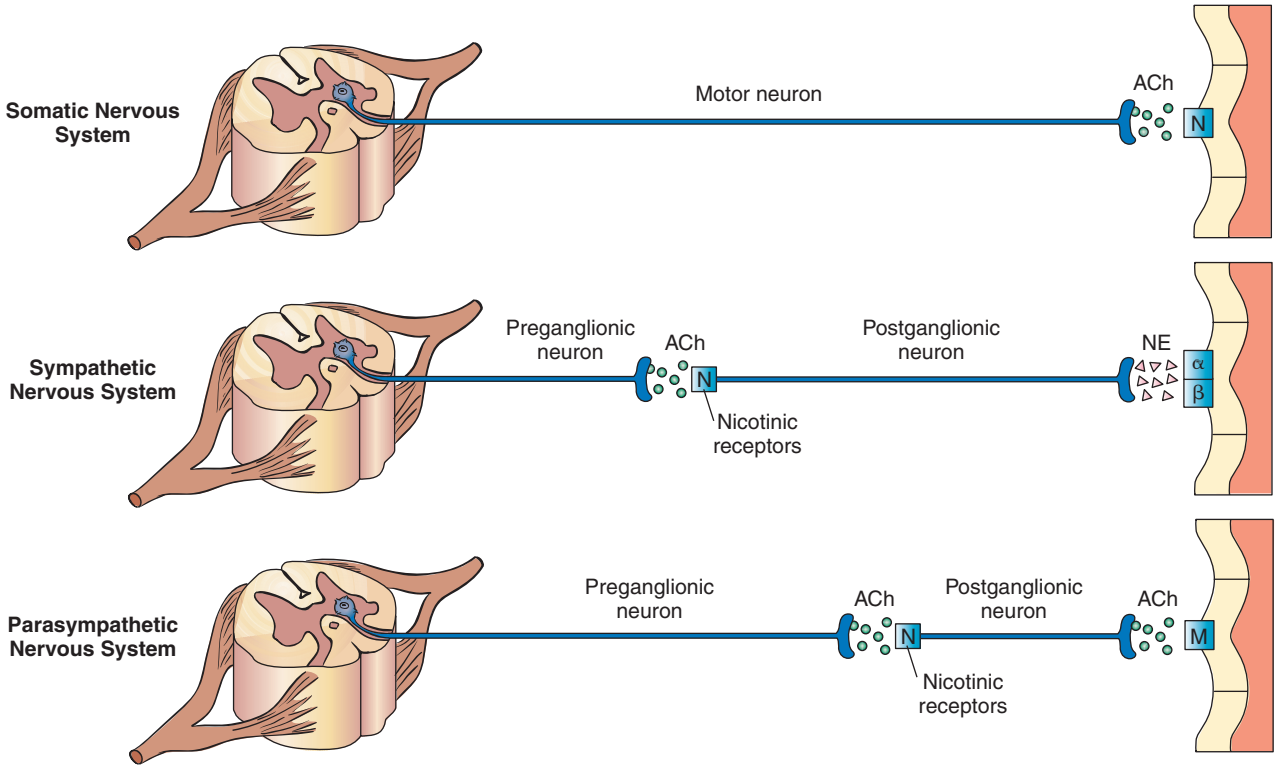


FIG 43.24 Comparison of preganglionic and postganglionic neurotransmitters in the sympathetic and parasympathetic systems. Acetylcholine (ACh) is the neurotransmitter of the motor neuron, but its receptor differs from that of the parasympathetic terminations. *N,* Nicotinic receptors; *NE,* norepinephrine; *M,* muscarinic receptors.

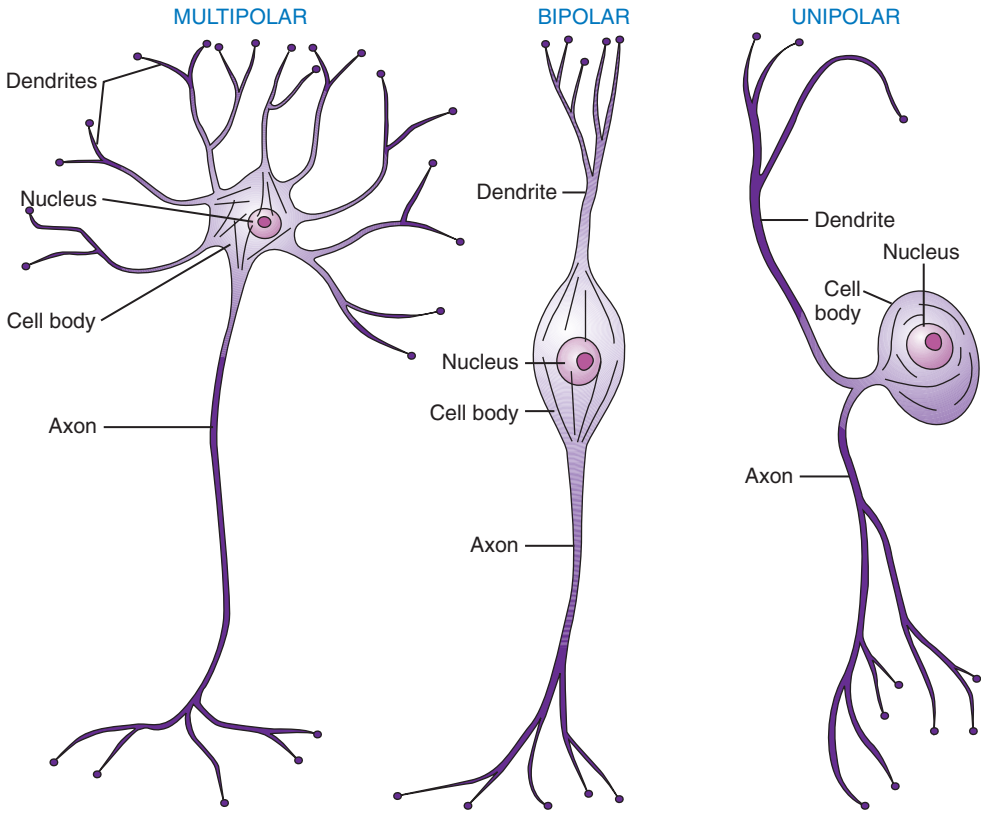


FIG 43.25 Three basic types of neurons: multipolar, bipolar, and unipolar.

TABLE 43.6 Effects of the Autonomic Nervous System on Organ System Function

	SYMPATHETIC		PARASYMPATHETIC	
Organ	**Action**	**Receptor**	**Action**	**Receptor**
Heart				
SA node, heart rate	↑	β_1	↓	M
AV nodal conduction	↑	β_1	↓	M
Contractility	↑	β_1	↓ (atria only)	M
Vascular Smooth Muscle				
Skin; splanchnic	Constricts	α_1		
Skeletal muscle	Dilates	β_2		
Skeletal muscle	Constricts	α_1		
Endothelium			Releases EDRF	M
Bronchioles	Dilates	β_2	Constricts	M
Gastrointestinal Tract				
Smooth muscle, walls	Relaxes	α_2, β_2	Contracts	M
Smooth muscle, sphincters	Contracts	α_1	Relaxes	M
Saliva secretion	↑	β_1	↑	M
Gastric acid secretion			↑	M
Pancreatic secretion			↑	M
Bladder				
Wall, detrusor muscle	Relaxes	β_2	Contracts	M
Sphincter	Contracts	α_1	Relaxes	M
Male Genitalia	Ejaculation	α_1	Erection	M
Eye				
Radial muscle, iris	Dilates pupil (mydriasis)	α_1		
Circular sphincter muscle, iris			Constricts pupil (miosis)	M
Ciliary muscle	Dilates (far vision)	β_2	Contracts (near vision)	M
Skin				
Sweat glands, thermoregulatory	↑	M*		
Sweat glands, stress	↑	α_1		
Pilomotor muscle (goose bumps)	Contracts	α_1		
Lacrimal Glands			Secretion	M
Liver	Gluconeogenesis; glycogenolysis	α_1, β_2		
Adipose Tissue	Lipolysis	β_1		
Kidney	Renin secretion	β_1		

*Sympathetic cholinergic neurons.
AV, Atrioventricular; *EDRF,* endothelial-derived relaxing factor; *M,* muscarinic receptor; *SA,* sinoatrial.

hold true for the brain. However, with significant effort, it is possible to reclaim at least some of what was lost.

NEURONAL COMMUNICATION

Neurons communicate primarily through the release of neurotransmitters into the synapses adjacent to target neurons. Postsynaptic neurons have receptors for these neurotransmitters and respond by changing the flow of ions through channels in the cell membrane. Some neuronal communication occurs through gap junctions that connect the cytoplasm of one neuron to the next, forming electrical synapses (see Chapter 3).

Sufficient depolarization of the neuronal membrane results in the generation of action potentials, which transmit signals quickly from one end of the neuron to the other. Action potentials reaching the axon terminal open voltage-gated Ca^{2+} channels and stimulate the release of neurotransmitter into the synapse. Not all neurotransmitters are excitatory; some are inhibitory and suppress the formation of action potentials in the postsynaptic neuron. Most neurons have many contacts, some inhibitory and some excitatory, such that the response of the postsynaptic neuron is a summation of all the input.

Membrane Potentials

A detailed discussion of membrane potentials can be found in Chapter 3, and the major points are reviewed here. All cells of the body contain slightly more negatively charged molecules than positively charged ones. These negative ions are trapped intracellularly because they cannot

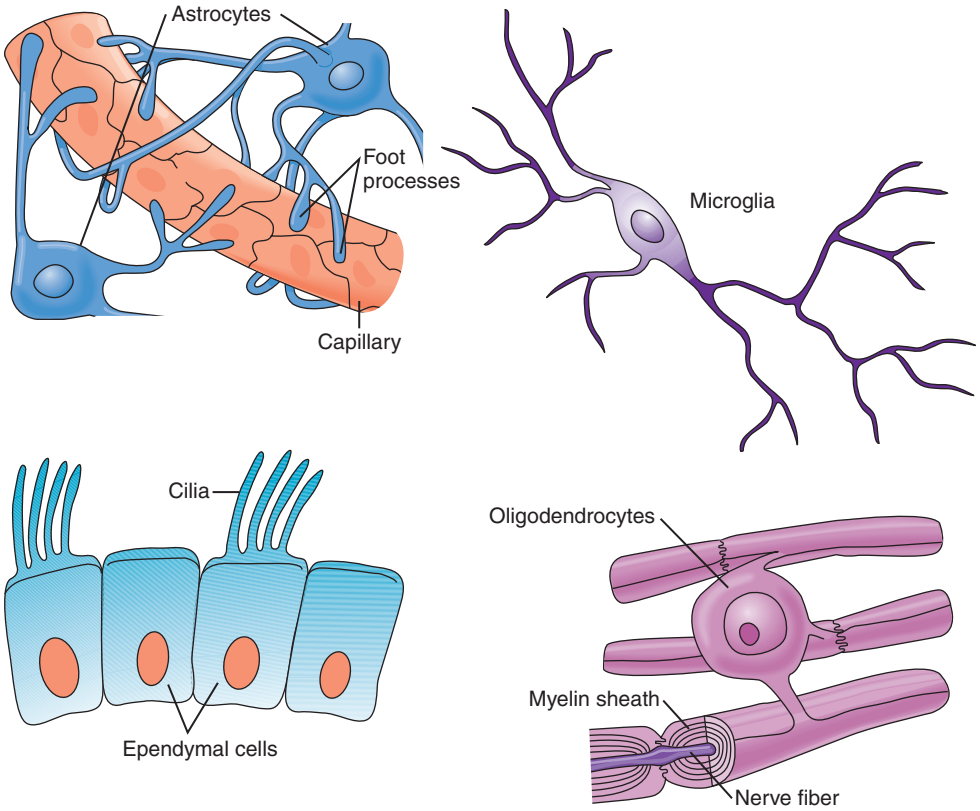


FIG 43.26 Four types of neuroglial cells: astrocytes, microglia, ependymal cells, and oligodendrocytes.

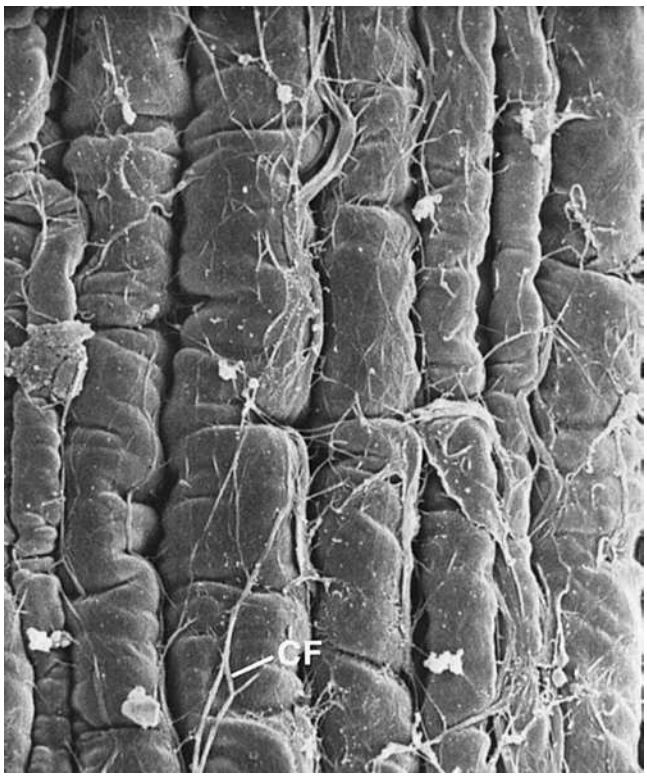


FIG 43.27 Oligodendrocytes wrap around nerve axons to form a myelin sheath. The nodes of Ranvier are the small spaces between the oligodendrocytes. *CF,* C-fiber (unmyelinated). (From Kessel RG, Kardon RH: *Tissues and organs: a text-atlas of scanning electron microscopy,* San Francisco, 1979, Freeman, p 80.)

pass through the plasma membrane. Positive ions are attracted to the cell membrane by the negatively charged intracellular ions. Because the cell membrane is permeable at rest to K^+ ions but not to Ca^{2+} or Na^+ ions, potassium accumulates in the cell to neutralize the intracellular anions. The unequal distribution of K^+ across the cell membrane creates a concentration gradient, pulling K^+ back out of the cell. At equilibrium, the electrical gradient pulling K^+ into the cell and the chemical gradient pulling it out are balanced. This equilibrium point leaves a few extra negatively charged ions inside the cell with no positive ion to neutralize them. The negative ions line up on the inside of the cell membrane to interact with positive ions on the other side (Na^+, K^+, Ca^{2+}). This separation of charge across the membrane at rest creates a membrane potential that can be measured and is about −65 to −90 mV. The membrane potential changes when the concentration of K^+ changes and when the permeability of the membrane to other ions changes.

Excitable cell types, like nerve and muscle, have ion channels in their cell membranes that open and close in response to fluctuations in membrane voltage. The most important voltage-gated ion channels in nerves are the fast Na^+ channels and the K^+ channels. Fast sodium channels allow Na^+ influx during the upstroke of the action potential, whereas potassium channels allow K^+ to leave the cell and help repolarize the membrane (see Chapter 3).

An action potential is initiated when neurotransmitters bind to receptors on the dendrite and cell body and allow cations, especially Na^+, to leak in. These channels are not voltage-gated channels; they are ligand-gated channels that open in response to a neurotransmitter binding to their receptor domain. If sufficient Na^+ leaks into the cell to raise the membrane potential to threshold, the fast voltage-gated Na^+ channels open and an action potential results. Opening of fast Na^+ channels in one section of the membrane allows Na^+ to flow in and bring the next section to threshold, thus opening the fast Na^+ channels

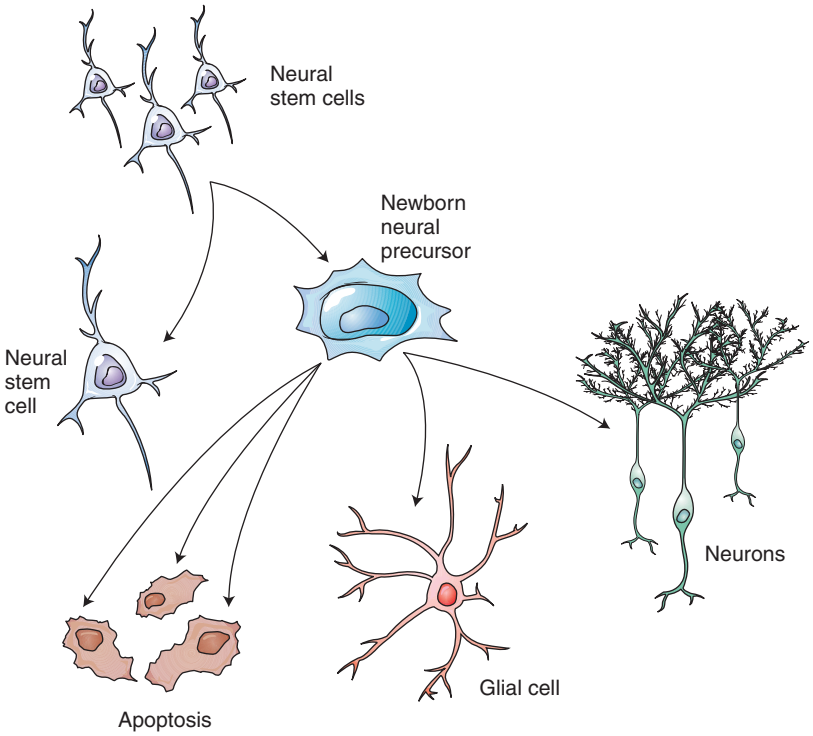


FIG 43.28 Schematic drawing of neural stem cell proliferation. Stem cells can differentiate into glial cells or neurons under the right conditions, but half fail to find a home and undergo apoptosis (programmed cell death).

in that section. This pattern repeats over and over again down the length of the axon. Threshold represents the amount of membrane depolarization required to cause fast Na^+ channels to flip into their open conformation (see Chapter 3). The axon hillock, the point at which the axon emerges from the cell body, is the usual site of action potential initiation because it has a high density of fast Na^+ channels and, therefore, a lower threshold.

Voltage-gated K^+ channels assist with repolarization because K^+ is allowed to flow out of the cell. During an action potential, the electrical gradient holding K^+ in the cell temporarily disappears as the membrane voltage moves toward zero. Potassium flows out of the cell passively down its concentration gradient. The Na^+–K^+ pumps work continuously to remove Na^+ from the cell interior and bring K^+ back in. The majority of a nerve cell's energy expenditure is used to power the Na^+–K^+ pumps.

The speed at which an action potential travels is determined by axonal diameter and myelination. Larger and myelinated neurons conduct impulses more quickly (Fig. 43.29). In myelinated neurons, action potentials are generated only at the nodes of Ranvier, allowing the impulse to hop quickly from node to node down the axon. This is called *saltatory conduction.*

Synaptic Transmission

The great majority of synapses responsible for signal transmission in the CNS function by using neurotransmitters. A neurotransmitter is released from the synaptic terminal of one neuron, proceeds across the synaptic cleft, and acts on the receptor proteins in the membrane of the second neuron to excite, inhibit, or modify its activity. The response at the postsynaptic membrane depends on the type of ion channel that is opened or closed when the neurotransmitter binds to the receptor. Once neurotransmitters are released into the synaptic cleft, their potential to activate the postsynaptic receptors is limited by deactivation processes. Neurotransmitters are either actively transported back into the axon terminals for reuse or destroyed by enzyme activity.

An *excitatory postsynaptic potential* (EPSP) results when a neurotransmitter has a depolarizing effect on the postsynaptic membrane. The EPSP may be too small to bring the axon hillock to threshold, and EPSPs from several presynaptic neurons may be required to generate an action potential. Thus postsynaptic potentials are not all-or-none phenomena, as are action potentials. Neurotransmitters that produce EPSPs do so by opening channels in the membrane that allow Na^+ influx. In some cases, the receptor itself is a channel *(ionotropic receptor);* in others, the receptor is linked to the channel through a second messenger cascade *(metabotropic receptor)* (Fig. 43.30). The ion channels regulated by metabotropic receptors may participate in action potential generation, but receptor activation also exerts more long-lasting effects on cell structure and behavior.

Some neurotransmitters inhibit depolarization and may produce hyperpolarization of the postsynaptic membrane by opening Cl^- or K^+ channels. Chloride ions leaking into the cell or potassium ions leaking out of the cell serve to short-circuit the effect of sodium ion influx, thus making it more difficult to reach threshold. This effect is called an *inhibitory postsynaptic potential* (IPSP). Neurotransmitters that result in IPSPs include γ-aminobutyric acid (GABA) and glycine.

Most synapses in the CNS have many presynaptic neurons, some producing EPSPs and some producing IPSPs. The membrane potential of the postsynaptic membrane is an algebraic sum of all the IPSPs and EPSPs occurring at any one moment in time. This is called *summation* and is the basis of neuronal processing and integration (Fig. 43.31). The term *spatial summation* is applied when multiple presynaptic neurons release their neurotransmitters onto one postsynaptic neuron at the same time. The IPSPs and EPSPs sum algebraically to produce the overall postsynaptic potential. *Temporal summation* occurs when one presynaptic neuron fires in rapid succession so that it releases more neurotransmitter onto the postsynaptic cell before the postsynaptic neuron has completely recovered from a previous dose.

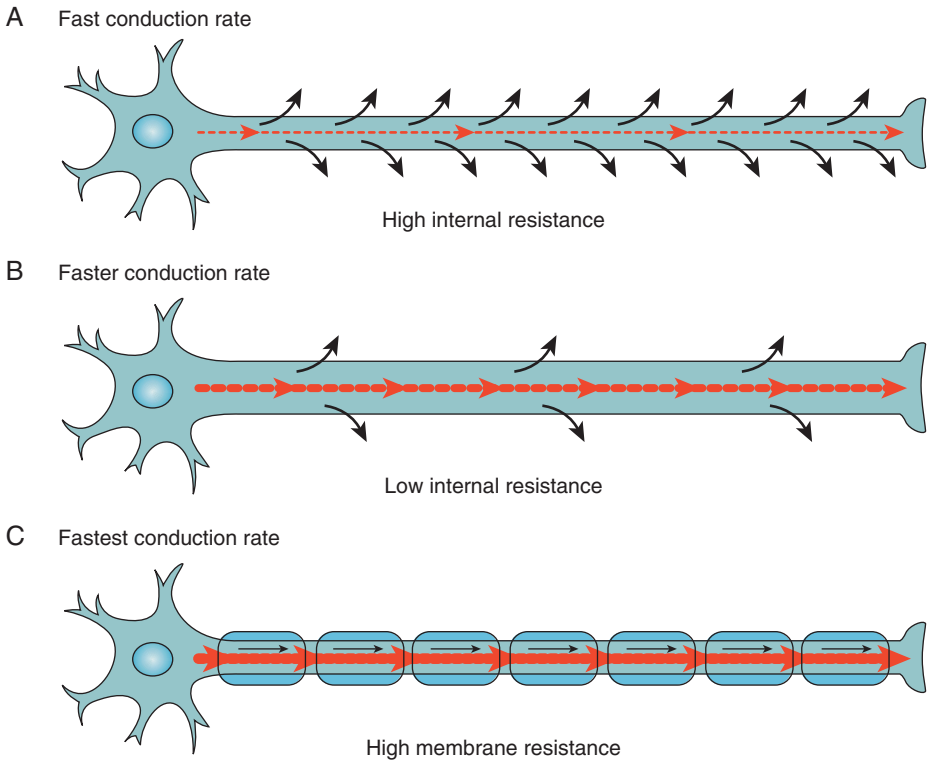


FIG 43.29 Rate of action potential conduction down an axon depends on the relative degree of internal resistance to current flow. When the diameter is small **(A),** there is higher internal resistance and slower conduction than in large-diameter neurons. A larger diameter **(B)** reduces internal resistance and accelerates the rate of conduction. Myelination **(C)** produces the fastest rate of conduction by increasing membrane resistance and decreasing internal resistance.

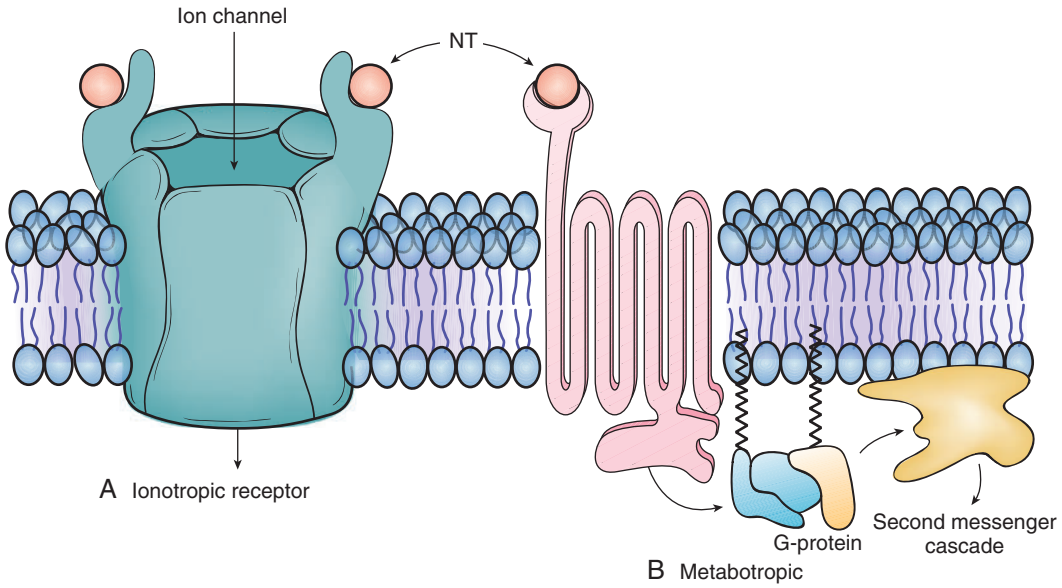


FIG 43.30 Neurotransmitter (NT) receptor classes. **A,** Ionotropic receptors are channel proteins that open when a neurotransmitter binds to them. **B,** Metabotropic receptors activate intracellular signaling cascades that generate second messengers in the cell when the neurotransmitter binds.

Neurotransmitters

Neurotransmitters are grouped according to their chemical structure into six principal categories (Box 43.1). *ACh* is the sole neurotransmitter in its class and is prevalent in numerous areas in the CNS. It is the neurotransmitter in autonomic ganglia, postganglionic parasympathetic synapses, and neuromuscular junctions. There are two major types of ACh receptors: the nicotinic receptors (N) are of the ionotropic variety, and the muscarinic receptors (M) are metabotropic (Table 43.7). When ACh is released into the synapse, it is quickly degraded by acetylcholinesterase to limit the duration of action (Fig. 43.32). Choline is actively taken back up into the presynaptic membrane for resynthesis. ACh

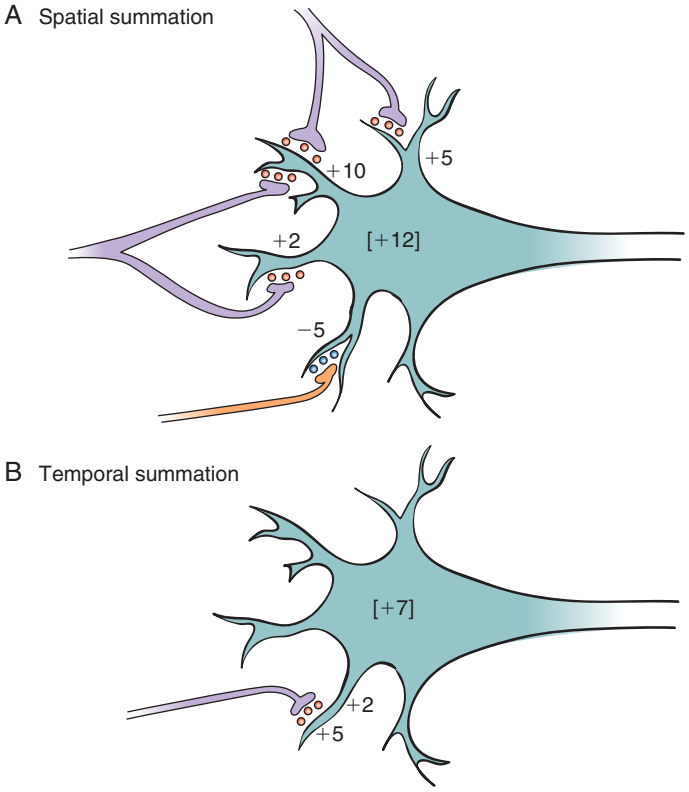


FIG 43.31 Summation. **A,** Spatial summation occurs when two or more presynaptic neurons release neurotransmitter onto one postsynaptic cell at the same time. The various excitatory postsynaptic potentials (EPSP) and inhibitory postsynaptic potentials (IPSP) add algebraically to determine the overall postsynaptic potential (PSP) reaching the axon hillock. **B,** Temporal summation occurs when one presynaptic neuron fires in rapid succession such that a previous PSP has not fully dissipated before the next PSP is added to it.

BOX 43.1 Six Major Classes of Neurotransmitters

Acetylcholine
Amines
Dopamine
Norepinephrine
Epinephrine
Serotonin
Histamine
Amino Acids
Excitatory
  Glutamate
  Aspartate
Inhibitory
  Glycine
  γ-Aminobutyric acid (GABA)
Polypeptides
Substance P, other tachykinins
Vasopressin
Oxytocin
Corticotropin-releasing hormone
Thyrotropin-releasing hormone
Growth hormone–releasing hormone
Somatostatin
Gonadotropin-releasing hormone
Endothelins
Enkephalins
β-Endorphin, other derivatives of proopiomelanocortin
Cholecystokinin
Vasoactive intestinal polypeptide
Neurotensin
Gastrin-releasing peptide
Gastrin
Glucagon
Motilin
Secretin
Calcitonin gene–related peptide α
Neuropeptide Y
Activins
Inhibins
Angiotensin II
Galanin
Atrial natriuretic peptide
Brain natriuretic peptide
Purines
Adenosine
Adenosine triphosphate
Gases
Nitric oxide
Carbon monoxide

TABLE 43.7 Mechanism of Action of Selected Nonpeptide Neurotransmitters

Transmitter	Receptor	Second Messenger or Ion Channel
Acetylcholine	Nicotinic	Cation channel
	M_1, M_3, M_5	↑ IP_3, DAG
	M_2, M_4	↓ cAMP
Dopamine	D_1, D_5	↑ cAMP
	D_2, D_3, D_4	↓ cAMP
Norepinephrine	α_{1A}, α_{1B}, α_{1D}	↑ IP_3, DAG
	α_{2A}, α_{2B}, α_{2C}	↓ cAMP
	β_1, β_2, β_3	↑ cAMP
5HT (serotonin)	$5HT_{1A}$, $5HT_{1B}$, $5HT_{1D}$, $5HT_5$	↓ cAMP
	$5HT_{2A}$, $5HT_{2C}$	↑ IP_3, DAG
	$5HT_3$	Na^+ channel
	$5HT_4$, $5HT_6$, $5HT_7$	↑ cAMP
Adenosine	A_1, A_3	↓ cAMP
	A_2	↑ cAMP
Glutamate	Metabotropic ($mGluR_1$ to $mGluR_8$)	Some ↑ cAMP, some ↓ cAMP, some ↑ IP_3, DAG
	AMPA, kainate	Na^+ channel
	NMDA	Ca^{2+} channel
GABA	$GABA_A$	Cl– channel
	$GABA_B$	↓ cAMP

AMPA, α-Amino-3-hydroxy-5-methyl-4-isoxazolepropionate; *cAMP,* cyclic adenosine monophosphate; *DAG,* diacylglycerol; *GABA,* γ-aminobutyric acid; *5HT,* serotonin; *IP_3,* inositol trisphosphate; *NMDA, N*-methyl-D-aspartate.

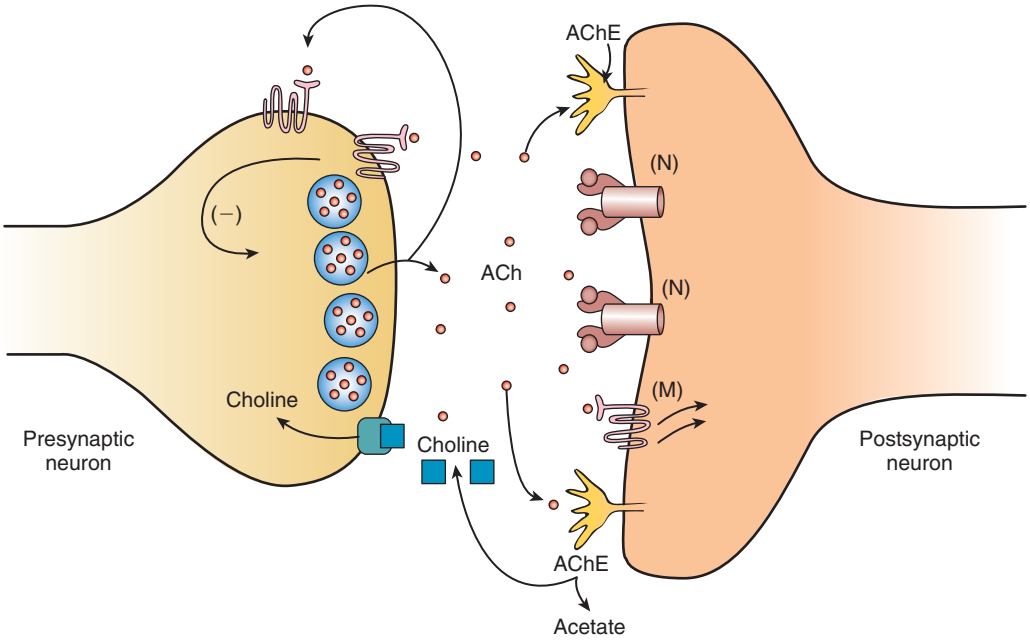


FIG 43.32 Acetylcholine synapse. Acetylcholine (ACh) released into the synapse binds to nicotinic (N) or muscarinic (M) receptors on the postsynaptic membrane. ACh also can bind to presynaptic receptors that are linked to a decrease in ACh release (negative feedback). Acetylcholinesterase (AChE) quickly degrades the ACh into acetate and choline. Choline is actively taken back into the presynaptic neuron for resynthesis.

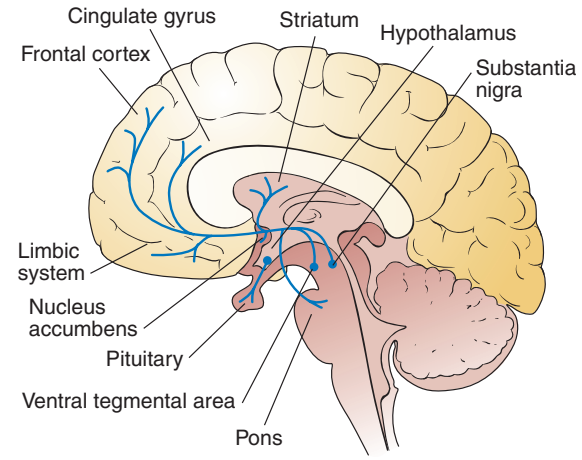
receptors located on the presynaptic membrane provide a negative feedback loop, whereby the presynaptic neuron monitors the amount of ACh in the synapse.

The *amines* include DA, NE, epinephrine, serotonin (5-hydroxytryptamine, 5HT), and histamine. Amines are particularly involved in the limbic system, hypothalamus, and basal ganglia. NE is the neurotransmitter released at SNS postganglionic nerve endings. DA, NE, and serotonin are important in regulating thought processes and mood. DA-secreting neurons project to the striatum (basal ganglia), pituitary gland, limbic system, and frontal cortex (Fig. 43.33A). DA can be degraded by enzymes in the extracellular fluid (catechol-*O*-methyltransferase, COMT) or by enzymes in the presynaptic nerve (monoamine oxidase, MAO). The primary means of clearing DA from the synapse is by active reuptake into the presynaptic membrane (see Fig. 43.33B). There are at least five DA receptor subtypes; all of these subtypes are metabotropic and linked to the production of second messengers (see Table 43.7). Abnormality of DA metabolism is apparent in various diseases, including Parkinson disease and schizophrenia.

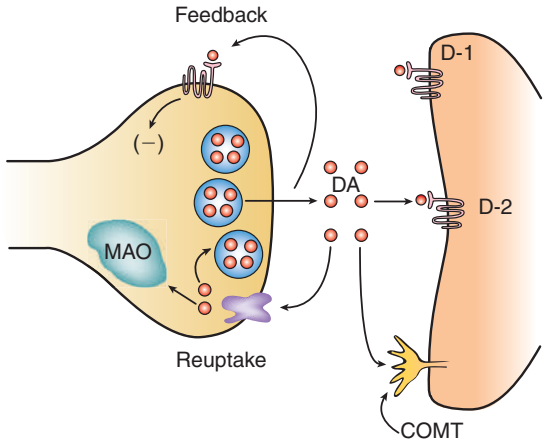
Some neurons have an enzyme for the hydroxylation of DA to form NE. The NE-secreting neurons originate in the brainstem (locus coeruleus) and project widely throughout the brain, including the cerebral cortex, cerebellum, limbic structures, brainstem, and spinal cord (see Fig. 43.33C). Most of the NE released into the synapse is cleared by active reuptake into the presynaptic neuron, where it can be repackaged for release or broken down by MAO. A number of receptor subtypes can bind NE in the synapse, and all are of the metabotropic variety and linked to second-messenger cascades (see Table 43.7). The receptor subtype α_2 is commonly located on the presynaptic membrane where it provides a negative feedback loop for the presynaptic cell to monitor the amount of NE in the synapse (see Fig. 43.33D). Stimulation of the presynaptic α_2 receptor by NE or by α_2-agonist drugs reduces the amount of NE released into the synapse by the neuron.

Serotonin is another amine that affects numerous areas of the brain in a pattern similar to that of NE (see Fig. 43.33E). Numerous serotonin receptor subtypes have been identified, including one ionotropic ($5HT_3$) and several metabotropic subtypes (see Table 43.7). Like NE and DA, serotonin is cleared from the synapse by an active reuptake carrier on the presynaptic membrane (see Fig. 43.33F). It is also subject to degradation by MAO in the presynaptic cell. Like other neurotransmitters, serotonin can bind to presynaptic receptors that regulate its release. Numerous drugs have been developed to manage disorders associated with serotonin pathways, including depression, anxiety, and migraine headache. The class of medications known as *selective serotonin reuptake inhibitors* blocks the reuptake carrier for serotonin. The tricyclic antidepressants also block reuptake of serotonin, but they are less specific than the selective serotonin reuptake inhibitors and also affect reuptake of other amines.

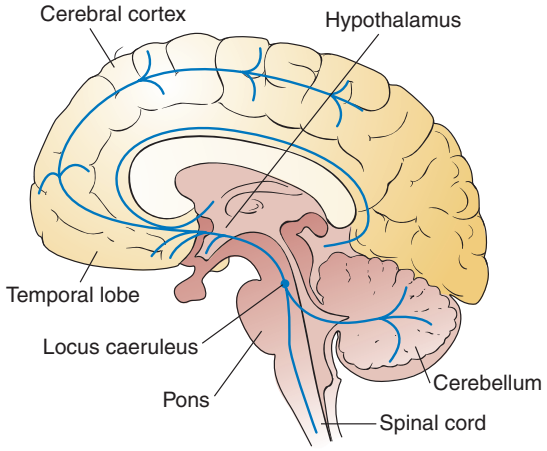
The category of *amino acids* can be subdivided into excitatory and inhibitory mechanisms of action. Glutamate and aspartate are the principal excitatory amino acids. Glutamate neurons are widely distributed throughout the brain, and glutamate is considered to be the primary excitatory neurotransmitter. Glutamate is involved in memory and has been implicated as a neurotoxin when released in excessive amounts (Chapter 44). Glutamate is removed from the synapse by active reuptake transporters on the presynaptic membrane. When energy stores are low because of interrupted blood supply or hypoxia, the transporters do not function effectively and glutamate remains in the synapse where it can behave as a neurotoxin. Most glutamate receptors are ionotropic; the metabotropic types are poorly understood. The α-amino-3-hydroxy-5-methyl-4-isoxazolepropionate (AMPA) receptors are classic ligand-gated sodium channels that depolarize the postsynaptic membrane. The *N*-methyl-D-aspartate (NMDA) receptors are interesting because they will not open unless the binding of glutamate is paired with a cotransmitter (such as glycine or D-serine) and concurrent depolarization of the membrane (Fig. 43.34). The NMDA receptor is a ligand-gated calcium ion channel, but it is blocked by a magnesium ion when the postsynaptic membrane is polarized. It only opens in response to glutamate binding if a depolarization is produced at the



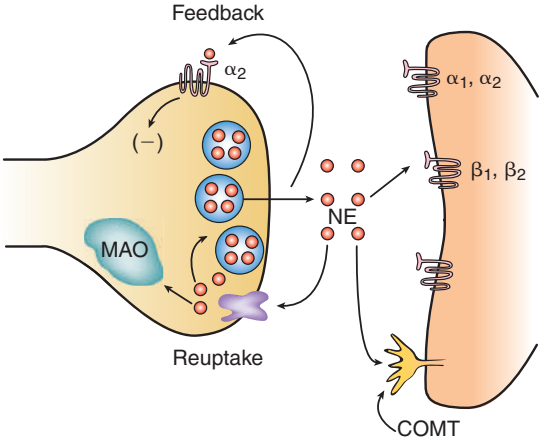
A Dopamine



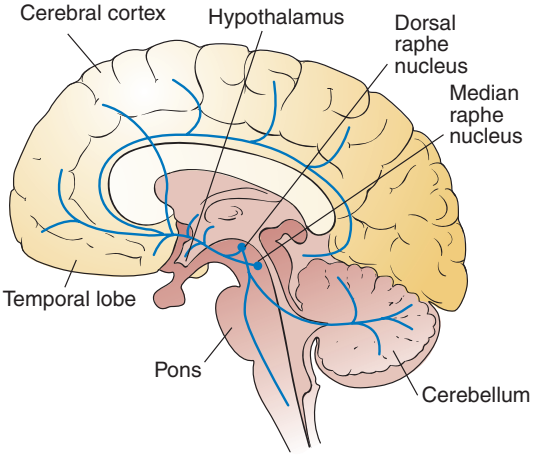
B Dopamine



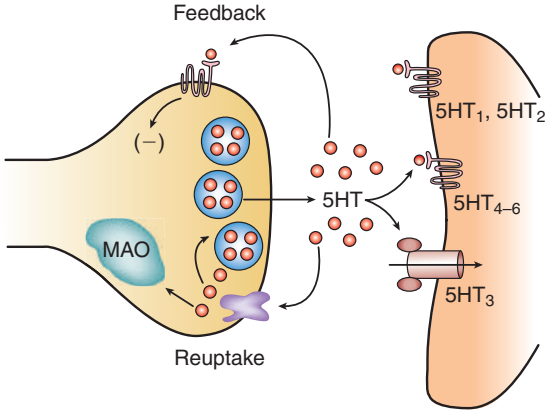
C Norepinephrine



D Norepinephrine



E Serotonin



F Serotonin

FIG 43.33 Amine synapses. **A,** Dopamine (DA) distribution in the brain. **B,** The DA synapse. **C,** Norepinephrine (NE) distribution in the brain. **D,** The NE synapse. **E,** Serotonin (5HT) distribution in the brain. **F,** The 5HT synapse. *COMT,* Catechol-*O*-methyltransferase; *MAO,* monoamine oxidase.

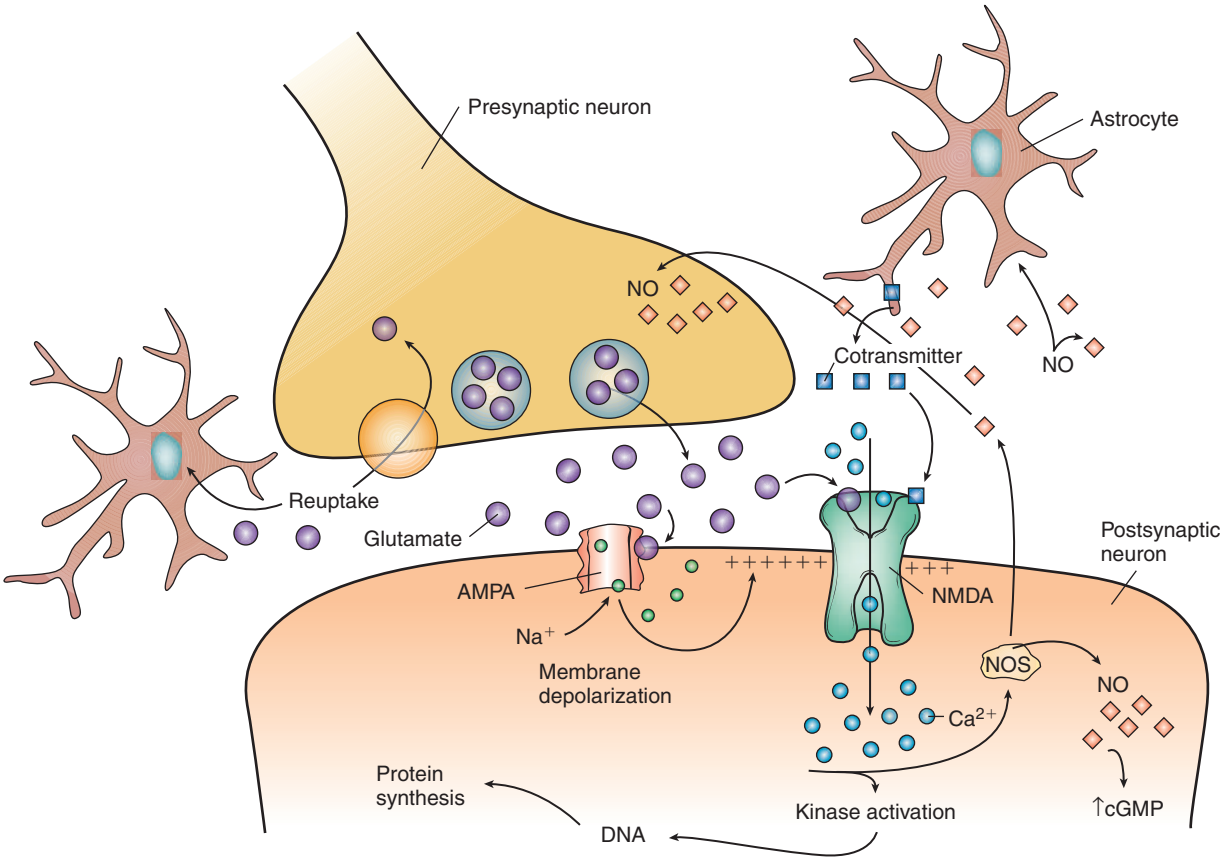


FIG 43.34 Glutamate synapse. Glutamate signaling is complex, having several receptor subtypes and costimulating molecules. The *N*-methyl-D-aspartate (NMDA) receptor is of special interest because it requires binding of glutamate and a cotransmitter (glycine or D-serine). In addition, the NMDA ion channel is blocked by Mg^{2+} and cannot open unless the postsynaptic membrane is already depolarized. Glutamate binding to its α-amino-3-hydroxy-5-methyl-4-isoxazolepropionate (AMPA) receptors can provide this depolarization by allowing sodium influx. When the NMDA channel opens, it allows calcium ions to flow in, triggering intracellular signaling cascades that produce nitric oxide (NO). NO is a gas that can diffuse throughout the synapse. NMDA receptor activation has been linked to long-term changes in synaptic efficiency. *NOS,* Nitric oxide synthase.

same time by another neurotransmitter–receptor interaction. The depolarization releases the blocking magnesium ion from the channel so that when glutamate binds, the channel opens to allow calcium influx. It is also unusual to use calcium ions to produce membrane depolarization because they can trigger signaling cascades within the cell. The NMDA receptor is thought to be responsible for long-term changes in the synapse that may relate to long-term memory. Drugs that interfere with NMDA receptors block memory; those that activate these receptors produce hallucinations and nightmares.

Glycine and GABA are inhibitory amino acids and are located throughout the spinal cord and brain. A large number of synapses (30%) are inhibitory in nature, and GABA is the principal neurotransmitter in these synapses. GABA is formed by decarboxylation (removal of CO_2) of glutamate, which transforms it from an excitatory amino acid to an inhibitory one. The GABA receptors are of two types: the $GABA_A$ is a classic ligand-gated chloride channel that produces an IPSP when activated. The $GABA_B$ receptor is a metabotropic receptor that also produces an IPSP and is linked to a reduction of cyclic adenosine monophosphate (cAMP) concentration in the cell. Barbiturates and benzodiazepines are thought to exert their depressive effects by increasing GABA activity.

A long list of neurotransmitters is found in the *neuropeptide* category (see Table 43.7). Neuropeptides may function as the primary neurotransmitter in the synapse, but more often they are released together with another neurotransmitter. Amines and neuropeptides are commonly released together into synapses. The neuropeptides have long-lasting effects on the postsynaptic cell, mediating changes in receptor number or structure and altering the responses of intracellular signaling pathways. Well-known neuropeptides include substance P, endorphins, and enkephalins, which are involved in the transmission and perception of pain. Neuropeptides are synthesized in the neuronal cell body and not in the nerve terminal like other neurotransmitters. The amount produced depends on the degree of gene activity that produces messenger RNA to direct the synthesis of the neuropeptide. Once synthesized and packaged into vesicles, the neuropeptides must be actively transported along the axon to the nerve terminal. All neuropeptide receptors are linked to second-messenger cascades. Neuropeptides are released in very small quantities in comparison to other neurotransmitters, and reuptake mechanisms are not required to turn off their activity. The neuropeptide with its bound receptor may be internalized into the postsynaptic cell, where the receptor is degraded or recycled to the synaptic membrane.

Purines, including adenosine triphosphate (ATP) and adenosine, function as neurotransmitters in various brain regions. Adenosine is thought to be continuously released by most neurons and modulates neurotransmission by blocking neurotransmitter release. It may be important in preventing seizure activity. The role of ATP as a

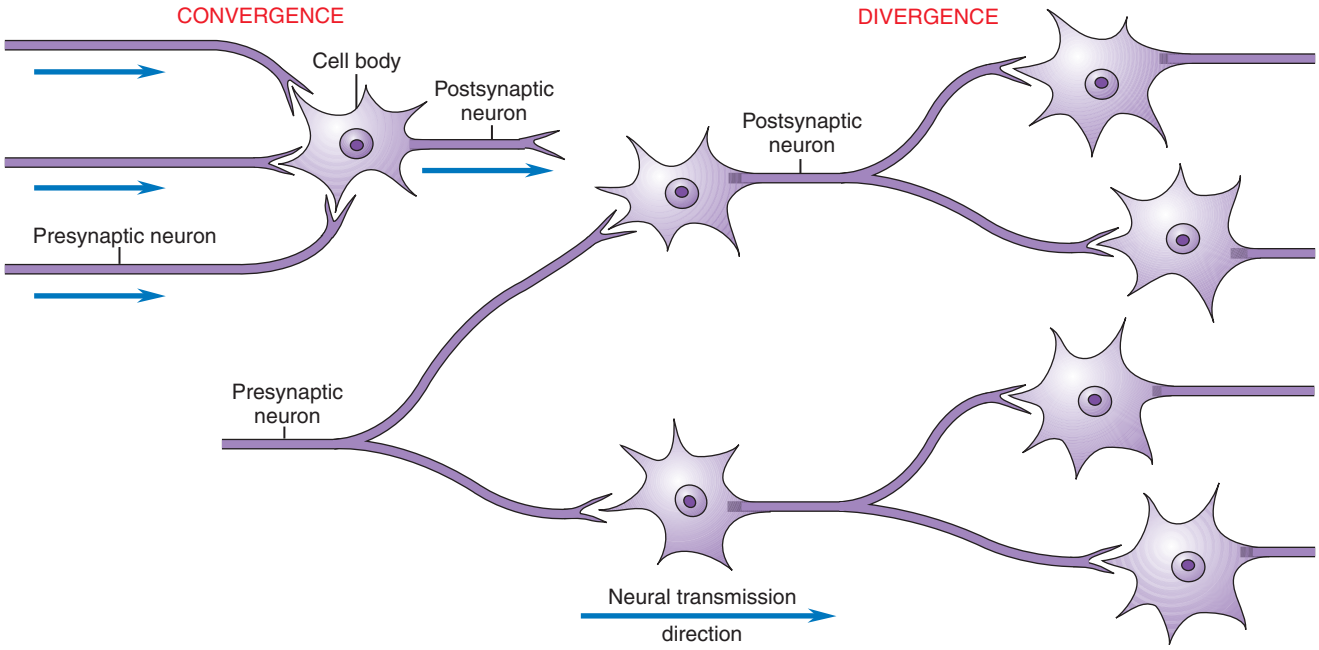


FIG 43.35 Convergence of several presynaptic neurons on one postsynaptic neuron is compared with divergence of one presynaptic neuron to several postsynaptic neurons.

neurotransmitter continues to be elucidated. There are at least three ATP receptor subtypes, two that decrease cAMP production and one that increases cAMP production.

Nitric oxide (NO) is a gas that can diffuse through cell membranes and therefore does not require a synaptic receptor for its activity. NO has several different potential targets within cells. For example, NO can bind and stimulate guanylyl cyclase, an enzyme that produces cyclic guanosine monophosphate (cGMP), a second messenger in the cell, or it can alter the activity of ion pumps, metabolic enzymes, and DNA transcription factors. Unlike other neurotransmitters that are produced and released by presynaptic neurons, NO can be synthesized in the postsynaptic neuron and diffuse locally to affect presynaptic neurons and nearby glial cells. The functions of NO are not completely known, but it is thought to be important in memory and pain perception. One trigger known to stimulate NO is activation of the previously described NMDA receptor. The calcium ions that flow in through the open NMDA receptor cause activation of an enzyme called *nitric oxide synthase* (NOS), which produces NO. NO may be the messenger that alerts the presynaptic membrane that the paired stimuli required to open the NMDA receptor were received.

Neuronal Circuits

Patterns of neuronal synaptic connections are called *neuronal circuits.* Activity in particular groups of neurons in one or more circuits is the basis of nervous system function: thoughts, memories, sensations, movements, and learning.

Divergence is a term used to describe neuronal circuits in which one presynaptic neuron makes contact with more than one postsynaptic neuron (Fig. 43.35). Divergence is a strategy used to send sensory input to a large number of receiving neurons. *Convergence* occurs when many presynaptic neurons synapse with one postsynaptic neuron (see Fig. 43.35). This arrangement is typical in the motor pathways, in which sensory, reflex, and voluntary inputs must be integrated into a response by the motor neurons that innervate skeletal muscle. Convergence is a mechanism of processing and integration of input.

The efficiency of circuits can be altered by changes in synaptic function. Synaptic transmission can be facilitated or inhibited in various ways. Alteration in the ease of synaptic transmission is the basis of memory and is discussed in more detail in the Consciousness, Memory, and Sleep section at the end of this chapter.

NEURAL DEVELOPMENT, AGING, AND INJURY

Development

The nervous system starts to take shape during the third week of embryonic development. At this time, three primary tissues of the embryo are distinguishable: the ectoderm, endoderm, and mesoderm. A thickened plate of ectoderm, running longitudinally on the dorsal surface of the embryo (neural plate), gives rise to the CNS and PNS. By the end of the third week, the neural plate folds to form a neural tube. Openings at either end of the neural tube are called *neuropores.* The neural tube is the precursor of the future brain and spinal cord.

Fusion of cells and formation of the neural tube start in the cervical region of the future spinal cord and then progress rapidly in a rostral direction toward the future brain. Failure of the neural tube to close properly is a cause of congenital malformation of the nervous system. Anencephaly (absent brain) results from failure of the rostral portion to close, whereas failure of the caudal portion to close results in myelomeningocele. In this defect, the spinal cord and meninges are displaced into a sac on the back (see Chapter 45). A reduction in neural tube defects has been achieved through prenatal maternal supplementation with folic acid.

Neurons grow and divide at an incredible rate during embryologic development and make primitive synaptic connections according to a basic architecture that is genetically programmed. The number of neurons and synapses ultimately dedicated to particular functions is determined in large part by their use. For example, if no visual sensory input is relayed to the primary visual cortex (as occurs with congenital cataracts), the cortical neurons will be reassigned to other functions. Similarly, a person born without arms will lack representation of these structures in the primary somatosensory cortex. In contrast, greater stimulation

appears to increase the number of neurons dedicated to a particular function. Critical periods in the early neonatal period have been identified when neuronal assignment to specific functions is determined. In recent years it has been recognized that a significant degree of neural plasticity exists throughout life; however, it is much less than that during early childhood. The brain continues to increase in size until puberty and remains stable until middle age.

Aging

A gradual loss of neurons in later adulthood does not result in significant alteration in brain function during the usual life span. However, the older one becomes, the greater the neurologic impairment. Excessive neuronal degeneration in adulthood is called *Alzheimer disease* or *senile dementia* and is distinguished from the normal changes of aging. The effects of aging on nervous system function are summarized in Geriatric Considerations: Changes in the Nervous System.

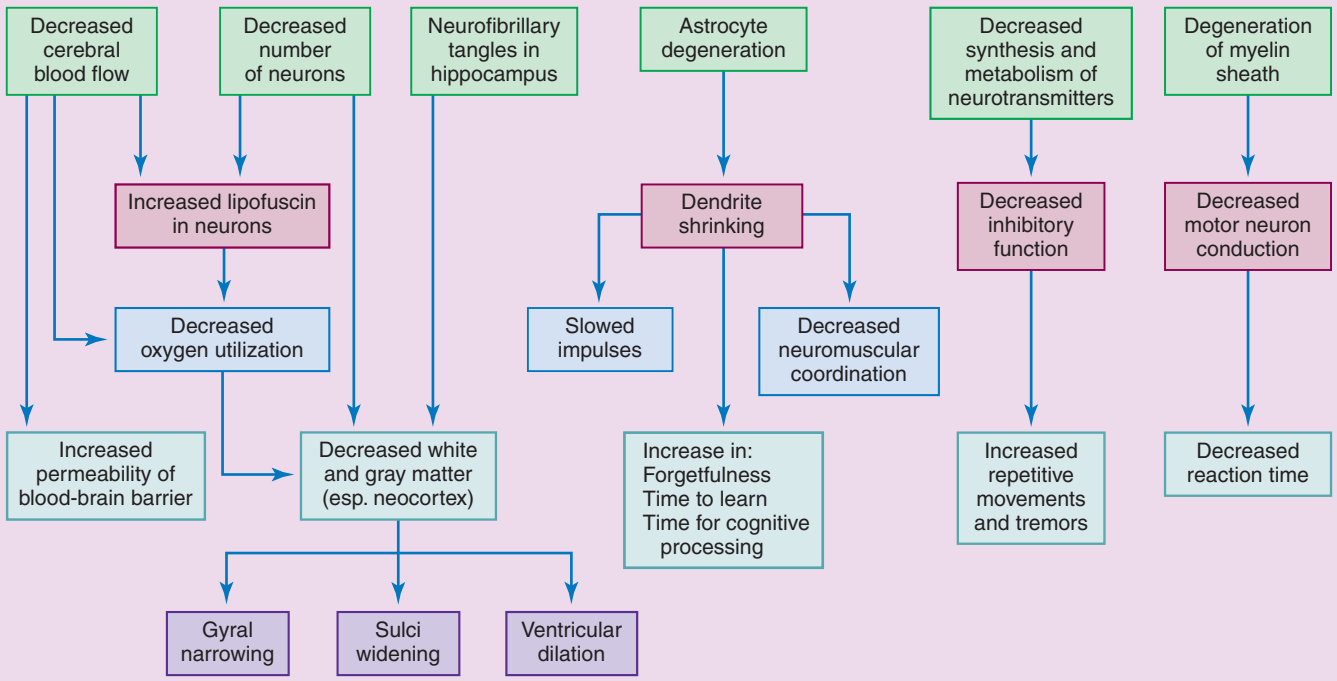
Injury

Injury to neurons usually results in neuronal cell death and loss of function; however, some regrowth is possible in peripheral nerves if the injury is not severe. Stem cells in the brain are capable of producing new, immature neurons; however, the likelihood that they will find and repair a site of injury and make appropriate synaptic connections is uncertain.

Injury to axons of neurons in the PNS results in degeneration of the distal segment (Wallerian degeneration). Within a few days to a week, the axons break into irregular fragments, and after 3 weeks they disappear completely. In the axonal segment proximal to the injury, a small section also degenerates. If the nerve is myelinated, degeneration usually extends back to the next one or two nodes of Ranvier. As axons and myelin sheaths degrade, Schwann cells of the injured nerve swell and divide. Over the next 2 to 3 weeks, continuous columns of short

GERIATRIC CONSIDERATIONS

Changes in the Nervous System



With aging, brain atrophy and a decrease in brain weight occur. This is evidenced by a decrease in the amount of white matter and gray matter up to 0.5% per year, with gyral narrowing, sulci widening, and ventricular dilation. There is a gradual atrophy and loss of neurons in the brain and spinal cord over time; but neuron loss is not uniform within the brain. Most of the neuron loss is in the neocortex (20%), Purkinje cells of the cerebellum, substantia nigra, and locus coeruleus. Some parts of the brain, such as the vestibular nucleus, have no neuron loss. Blood supply to the brain is decreased because of the decreased metabolic demands and brain atrophy. There is also increased permeability of the blood–brain barrier.

Intracellularly there is an increase in the amount of lipofuscin, which hampers cellular oxygen use, crowds intracellular organelles, and reduces the number of mitochondria. There are also neurofibrillary tangles in the hippocampus and neuritic plaques that are found only in the elderly.

Nerve fibers in the brain decrease in number and show signs of splitting or fragmentation. The cortex, subcortex, and cerebellar astrocytes degenerate. Nerve axons develop swellings near their ends called *neuroaxonal dystrophy*. The relevance of these swellings is unknown. Dendrites shrink, decreasing the number of messages received from other cells and synaptic linkages. This causes slowing of impulses and decreases neuromuscular coordination. These changes result in decreased short-term memory, reduced speed of learning, prolonged new information processing, increased reaction time, diminished abstract reasoning, and impaired perception.

Changes in the secretion and metabolism of neurotransmitters also affect the aging brain. There is a decrease in norepinephrine and dopamine secretion with an increase in monoamine oxidase activity. The reduction of dopamine levels leads to decreased inhibitory motor functions.

In the spinal cord, posterior root fibers and sympathetic nerve fibers of the autonomic nervous system decline in number. Peripherally, there is degeneration of the motor nerve fibers and myelin sheath. Motor neuron axons remain intact. Decreasing motor neuron conduction velocity and prolonged muscle action potentials lead to decreased reaction times. Reflexes may be decreased or absent. There is an increased risk of falls.

Schwann cells mark the course of the lost axons. If there is little separation between the ends of a divided nerve, the proliferating Schwann cells bridge the gap. If the divided ends are farther apart, the proliferating Schwann cells form bulbous swellings at the end of the nerves and the surviving axons form fine fibrils, which extend into the surrounding Schwann cells at random. Those fibrils that find a column of Schwann cells in the distal part of the nerve grow down the column. There is often good functional return to the nerve secondary to this process. If continuity is not restored, the distal end gradually becomes replaced by collagenous scar tissue.

Damaged axons in the CNS show a pattern of degeneration that is similar to that of peripheral neurons. Damaged axons become irregular and beaded, break up, and disappear, but the process is significantly slower. Methods using stains for β-amyloid precursor protein to identify axonal injury in the CNS have revealed that axonal injury is a common event even in mild concussion. Axonal damage impairs axonal streaming and causes a buildup of β-amyloid precursor protein proximal to the injury. This buildup is taken as evidence of axonal injury. Contrary to previously held notions, a significant degree of axonal repair appears to occur in CNS neurons when the injury is not too severe. If neurons that were the principal source of stimulation to some other group of neurons are damaged and die, that other group of neurons may also degenerate because of the loss of trophic (growth and survival) signals. Researchers continue to discover new nerve growth and survival factors. Eventually they may find ways to minimize neuronal degeneration after injury and encourage repair or replacement.

KEY POINTS

- The fundamental unit of the nervous system is the neuron. Neurons have three basic parts: the cell body, dendrites, and axons. The dendrites receive signals and transmit them to the cell body. The axon generates and conducts action potentials. Conduction of action potentials is faster in large and myelinated axons.
- Neuronal communication through chemical synapses can be summarized as follows: Stimulation from other neurons occurs primarily at the dendrite and cell body. Action potentials are initiated at the axon hillock and conducted down the axon to the axon terminal, where neurotransmitter is stored. Depolarization of the terminal opens voltage-gated calcium channels. Calcium influx mediates exocytosis of neurotransmitter into the interneuronal synapse. Neurotransmitter binds and activates specific receptors on the postsynaptic cell, changing its ion conductance. With sufficient depolarization of the postsynaptic cell, an action potential is generated.
- There are six categories of neurotransmitters based on their chemical structure: acetylcholine, amines, amino acids, neuropeptides, purines, and gases.
- Excitatory neurotransmitters create excitatory postsynaptic potentials (EPSPs) in the postsynaptic neuron attributable to opening channels that allow Na^+ or Ca^{2+} influx. Inhibitory neurotransmitters create inhibitory postsynaptic potentials (IPSPs) in the postsynaptic neuron attributable to opening channels that allow Cl^- influx or K^+ efflux. The summation of EPSPs and IPSPs at the axon hillock determines whether an action potential will be initiated.
- Neuroglial cells are supportive cells in the CNS. Oligodendroglial cells form insulating myelin sheaths, astroglial cells moderate extracellular fluid composition and synaptic conditions, microglial cells are derived from circulating monocytes and destroy foreign materials, and ependymal cells form CSF.
- Development of the nervous system follows a basic architecture that is genetically programmed. However, the brain is quite plastic, especially during infancy, and alters the assignment of neurons to certain functions based on the degree of stimulation.
- Neurons in the CNS that are severely injured generally do not regenerate. Peripheral neurons may regenerate if the Schwann cells provide a pathway for growth. Neural stem cells in the ventricles and hippocampus of the brain can proliferate to produce either glial or neuronal cells depending on specific cues, most of which have yet to be elucidated.

SENSORY FUNCTION

The discussion of sensory function in this chapter is restricted to the somatosensory system. The special senses of hearing, vision, taste, and olfaction are discussed in Chapter 46. Neural pathways related to pain transmission are discussed in detail in Chapter 47.

Transmission of sensory signals begins with activation of specialized dendritic processes, called *sensory receptors,* at the ends of sensory afferents that project to the spinal cord (or brainstem in the case of some cranial nerves). Secondary neurons in the cord are activated and carry the signals up the cord to the brain. The thalamus is the principal receiving site for somatosensory signals, which are then relayed to various brain areas, including the somatosensory cortex in the parietal lobe.

An important principle of sensory transmission is that somatotopic organization is maintained from receptor to somatosensory cortex. This property allows for precise localization of the origin of sensory signals. The somatosensory system conveys a number of different sensory modalities, including fine touch, vibration sense, pressure, temperature, itch, crude touch, and pain. In general, different modalities are sensed by different types of sensory receptors.

SENSORY RECEPTORS

Sensory receptors are specialized terminations of the dendrites of primary sensory neurons. The receptor may be a free nerve ending or may have various connective tissue elaborations that affect its responsiveness (Fig. 43.36). All types of receptors respond to stimuli by altering their

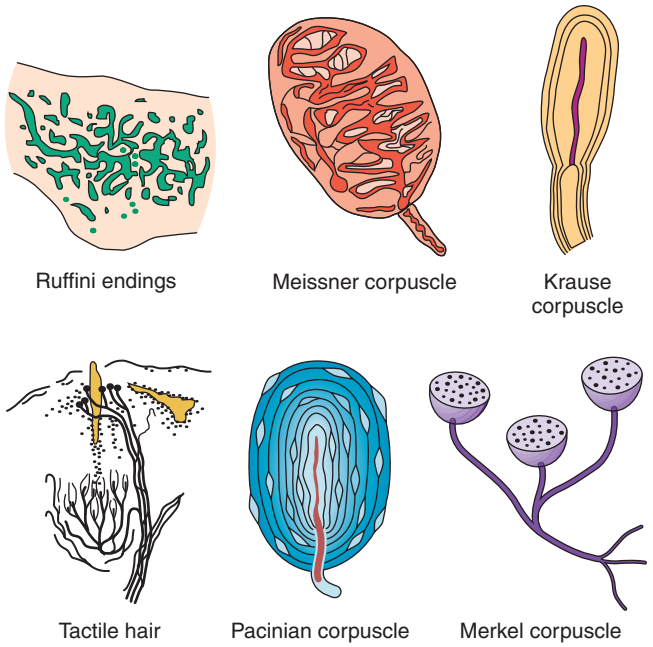


FIG 43.36 Common types of somatosensory receptors. Sensory receptors are specialized to respond to particular types of stimuli.

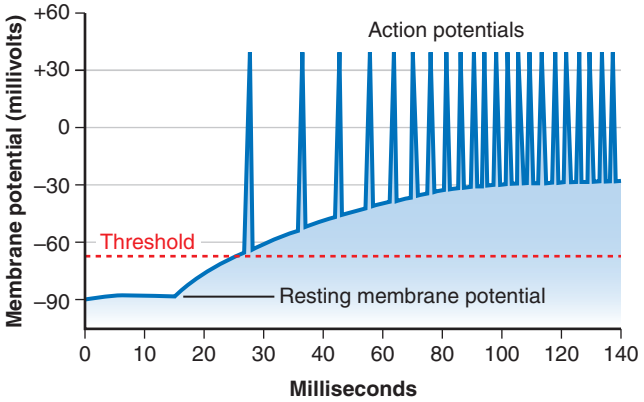


FIG 43.37 Relationship between stimulus intensity and action potential generation. As the stimulus increases, the receptor potential is greater, and action potentials are generated at a faster rate.

membrane permeability to ions, thus creating a change in the membrane voltage. These receptor potentials are similar to the EPSPs generated in postsynaptic neurons, except that the stimulus is not a neurotransmitter. Different receptor types respond to different kinds of stimuli: mechanical stretch, changes in temperature, or the binding of chemicals. When a stimulus depolarizes the receptor sufficiently, voltage-gated fast Na^+ channels in the membrane open and an action potential is generated. The rate of action potential generation by the receptor depends on the intensity of the stimulus (Fig. 43.37).

Some receptors are rapidly adapting and generate action potentials only when a change in the stimulus intensity occurs. The Pacinian corpuscle is a classic example of a rapidly adapting receptor that is well suited to transmitting the sense of vibration. Tonic receptors adapt slowly and are good for conveying information about stimulus intensity over time. Free nerve endings, such as pain receptors, are usually tonic receptors.

Impulses generated by receptors are transmitted to the dorsal root of the spinal nerve. Depending on the sensory modality, the nerve impulses may travel up the ipsilateral side of the cord or may cross the spinal cord to travel up the contralateral side. In general, the well-localized sensations of touch, pressure, and vibration travel up the ipsilateral side of the cord, whereas the sensations of pain, itch, and temperature usually cross over and travel to the brain on the contralateral side. Regardless of pathway taken, somatosensory signals eventually converge on the somatosensory cortex in the hemisphere opposite from the location of the primary receptor.

SENSORY PATHWAYS

Two major tracts, the dorsal column–medial lemniscal tracts and the anterolateral tracts, carry information from the spinal segments to the brain (Fig. 43.38). The dorsal column–medial lemniscal tract carries fine touch, vibration sense, and proprioception and remains ipsilateral until the level of the medulla. As the fibers progress upward, they gradually move toward the midline so that those corresponding to the lower extremities occupy the medial white column and those representing the arm are more lateral. From the nuclei in the medulla, neurons of the dorsal column pathway cross to the opposite side and travel up the brainstem, where they form the medial lemniscus, and then on to the thalamus. In the thalamus, fibers synapse with tertiary neurons, which in turn pass upward in a great band of fibers known as the *internal capsule,* and then travel on to the primary sensory cortex.

The anterolateral tract (previously called the *spinothalamic tract*) carries impulses for sensations of pain, itch, and temperature from small myelinated (Aδ) and unmyelinated (C-fiber) neurons. These neurons ascend one or more spinal segments before entering the posterior horn and synapsing with their secondary neuron. A few secondary fibers ascend ipsilaterally in the Lissauer fasciculus all the way to the thalamus. However, the majority of fibers cross the midline and ascend as the anterolateral tract. Fibers from the lower extremities and trunk are pushed laterally as they ascend in the spinal cord by the addition of fibers from the upper extremities and upper body, which enter medially.

On their way to the thalamus, these fibers give off collaterals to the reticular formation of the brainstem and the periaqueductal gray matter in the midbrain, where it is believed that one of their functions is pain inhibition (see Chapter 47). Secondary fibers of the anterolateral tract terminate in several thalamic regions. Tertiary neurons project from the thalamus to the somatosensory cortex. Thus although the sensations of fine touch and pain are separated in the cord, they reunite in the somatosensory cortex that lies in the cerebral hemisphere opposite the site of sensory receptor origin.

SOMATOSENSORY CORTEX

The primary somatosensory cortex is organized in columns of gray matter that correspond to specific body locations. All modalities of sensation are grouped together in adjacent sections. The somatotopic representation of the body along a strip of cortex creates a distorted picture of the human body, called a *homunculus map* (Fig. 43.39). Body areas with a greater density of receptors garner a larger part of the homunculus map. The homunculus map was created by stimulating discrete areas of the cortex in awake subjects and recording the sensations that they reported. It is now known that several different cortical areas contain somatotopically organized maps in addition to the well-known one in the primary somatosensory cortex. The *perception* of sensation from the body occurs at the level of the cortex.

KEY POINTS

- The body is somatotopically represented in the spinal cord and cerebral cortex. Stimulation of points in the primary somatosensory cortex results in discrete sensations in the contralateral side of the body.
- Projections to the somatosensory cortex begin in sensory receptors throughout the body. Receptors send axons to the spinal cord through the dorsal root. Stimulation of receptors by mechanical deformation, temperature, or chemicals alters membrane permeability, resulting in receptor potentials. The intensity of the stimulus is reflected in the rate of action potentials generated.
- Sensations of touch and proprioception (dorsal column–medial lemniscal tract) project up to the medulla on the ipsilateral side, then cross over and project to the thalamus.
- Sensations of pain, temperature, and itch (anterolateral tract) usually cross the cord near the level of entry and travel to the brain on the contralateral side.
- Sensory information from both tracts is transmitted from the thalamus to the same areas of the somatosensory cortex by way of the internal capsule.

MOTOR FUNCTION

The execution of voluntary movement requires interaction among basal ganglia, the cerebellum, and several regions of the cortex. The final program of voluntary muscle activity is transmitted from the brain down the spinal cord by way of the lateral corticospinal tracts. As previously noted, the corticospinal tract decussates in the medullary pyramids

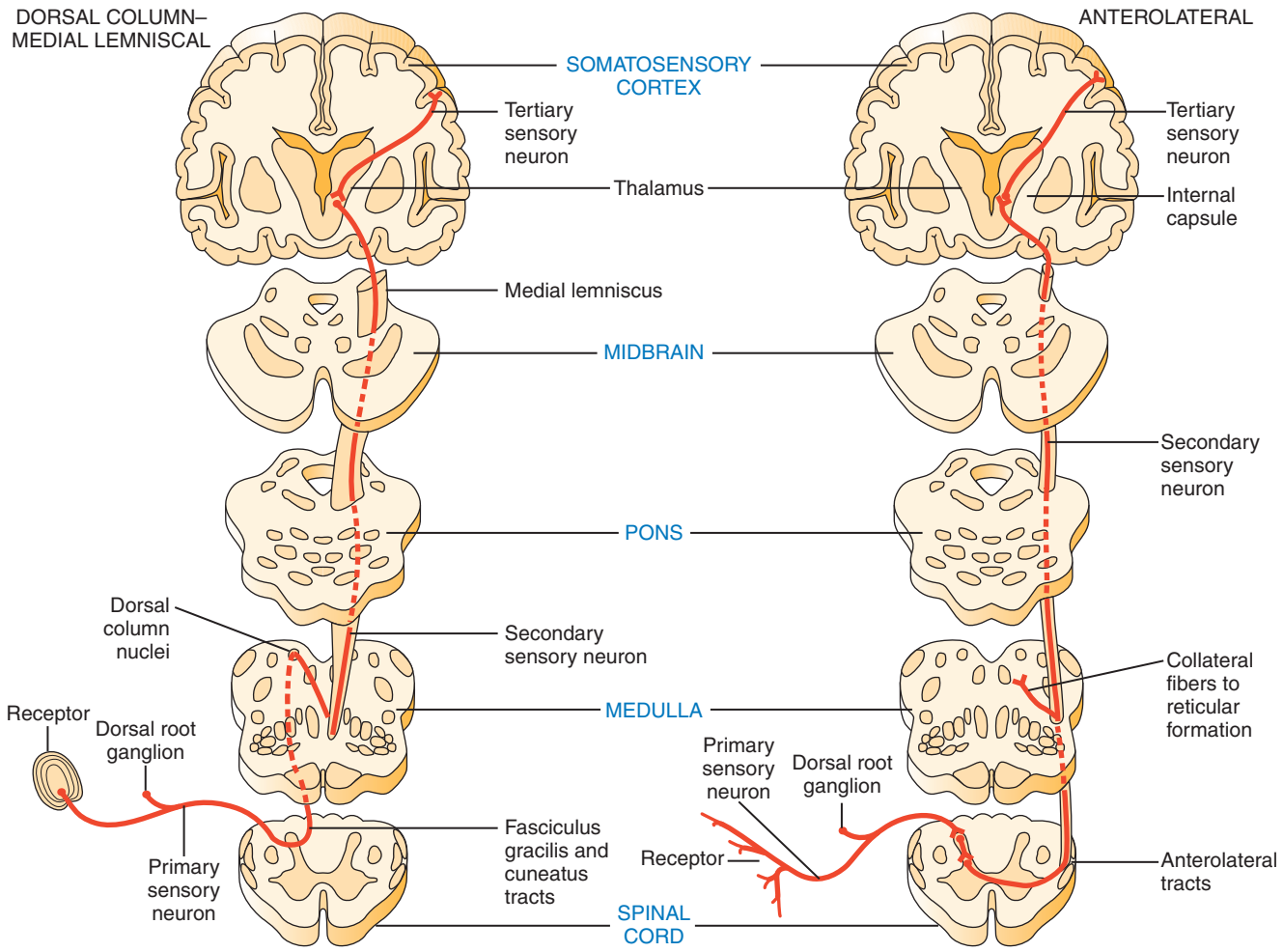


FIG 43.38 Comparison of the two major ascending somatosensory tracts. *Left,* Dorsal column–medial lemniscal tract. *Right,* Anterolateral tract. Note that the dorsal column tracts do not cross the midline until the level of the medulla, whereas the anterolateral tracts cross at the spinal cord level.

and travels down the spinal cord to control muscles on the contralateral side of the body (Fig. 43.40). The corticospinal tract primarily controls distal muscles of the arms, wrists, fingers, lower legs, feet, and toes. These are the muscles capable of fine-motor control. Another group of motor tracts innervate large proximal muscle groups and axial muscles that control posture and balance. These tracts include the vestibulospinal, reticulospinal, and tectospinal tracts. Motor tracts descending from the brain synapse on the cell bodies of motor neurons that lie in the anterior horn of the spinal cord and project to skeletal muscles.

MOTOR NEURONS

Motor neurons travel from the anterior horn of the spinal cord through the ventral root and within the spinal and peripheral nerves to finally innervate target muscles. The α motor neurons release ACh into neuromuscular junctions, depolarizing skeletal muscle cells and contracting all the fibers in the *motor unit.* A motor unit consists of a single motor neuron and all of the muscle fibers under its control. Some motor units are large, containing hundreds of muscle cells, whereas others may contain only one muscle cell per motor neuron. Smaller motor units produce finer gradations of muscle control.

A single action potential in the α motor neuron is sufficient to release enough neurotransmitter to contract the motor unit. Therefore the point of control of muscle contraction is at the cell body of the α motor neuron that lies within the anterior horn. A typical motor neuron receives hundreds of synaptic inputs, which summate to control the generation of action potentials. Some presynaptic inputs are from corticospinal neurons; others are sensory inputs from primary sensory afferents and spinal cord interneurons.

The γ motor neurons are small fibers that innervate structures within the body of the muscles, called *muscle spindles.* Muscle spindles are specialized sensory receptors that sense the length or stretch within the muscle and relay the information to the spinal cord. The γ motor neurons contract muscle fibers within the muscle spindle and regulate spindle sensitivity. Excessive γ motor neuron activity occurs in some types of brain coma, making the spindles hypersensitive and the muscles stiff and resistant to stretch (see discussions of decerebrate and decorticate rigidity in Chapter 44).

SPINAL REFLEXES

A great deal of motor coordination is exerted in the spinal cord through complex reflex pathways. These pathways allow upper motor neurons from the brain to initiate preprogrammed movements, rather than having to excite and inhibit each and every lower motor neuron individually. For example, experiments in lower animals have demonstrated

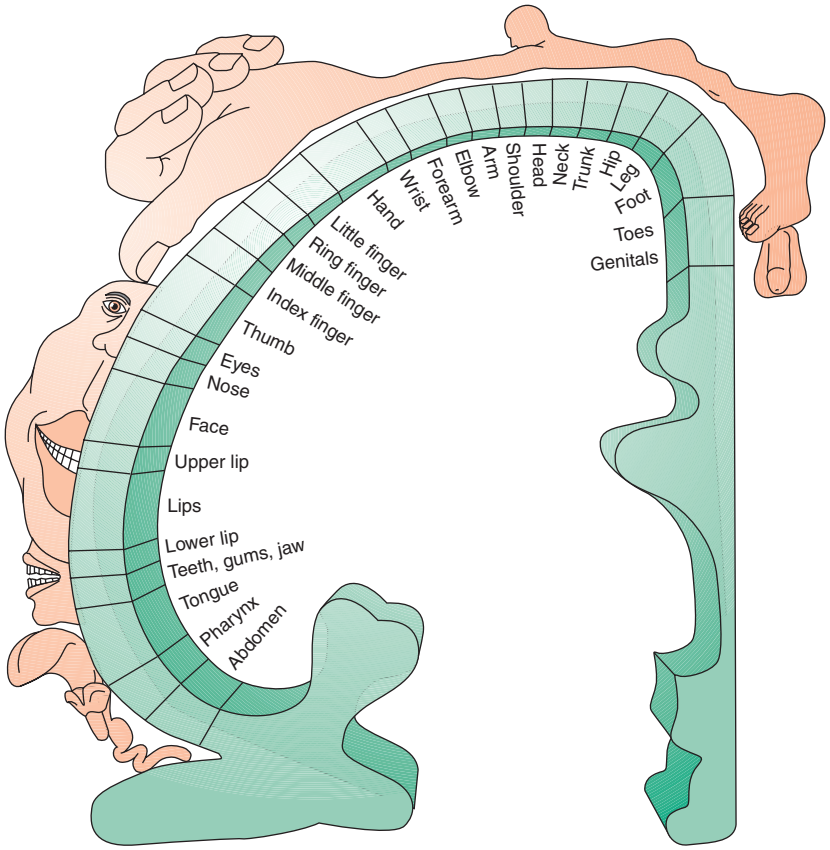


FIG 43.39 Topographic organization of the body on the somatosensory cortex, forming a homunculus map.

that pressure on the pads of the feet initiates complex walking movements even when the brain is no longer functional.

Spinal reflexes allow sensory information from pain receptors, proprioceptors, and muscle spindles to alter muscle contraction very quickly, even before the information reaches the brain. The stretch reflex and withdrawal reflex are illustrative examples.

The stretch reflex can be demonstrated by tapping the patellar tendon (below the knee) with a rubber hammer, which results in contraction of the quadriceps muscle and elevation of the lower leg. Evaluation of deep tendon reflexes is helpful in localizing a motor abnormality to the PNS or CNS. The deep tendon reflex tests the reflex arc between the sensory muscle spindles and the α motor neurons. Tapping the tendon produces a quick stretch in the muscle fibers of the quadriceps muscle and stimulates the muscle spindles. The muscle spindle sends action potentials along neurons (group Ia) that enter the dorsal horn of the spinal cord and then synapse directly on α motor neurons in the anterior horn that activate the muscle in which the spindles lie (Fig. 43.41). Other branches of the group Ia neurons make connections with antagonistic muscle groups and inhibit their contraction. The stretch reflex makes only one synapse (monosynaptic) and produces a very quick response in the muscle. The physiologic function of the muscle spindle stretch reflex is to provide feedback to α and γ motor neurons to adjust the strength of muscle contraction to match the load on the muscle.

The withdrawal reflex is an important protective mechanism that allows reflexive withdrawal of a body part from a physical threat while simultaneously making postural adjustments to avoid loss of balance. The withdrawal reflex is polysynaptic, making connections with interneurons in the cord to affect muscles on both sides of the body. A simplified model of the withdrawal reflex is shown in Fig. 43.42. Note that a painful stimulus to one extremity results in activation of flexor muscles and inhibition of extensor muscles on the ipsilateral side. This allows quick withdrawal of the extremity from the source of injury. Connections with muscle groups on the opposite side of the cord stimulate extensors and inhibit flexors to stabilize posture.

CENTRAL CONTROL OF MOTOR FUNCTION

Corticospinal tract neurons originate in the primary motor cortex, which is arranged in a similar manner to the somatosensory cortex. The motor homunculus map (Fig. 43.43) shows that muscles of the face, lips, tongue, and hands occupy most of the cortical neurons. These areas have small motor units and produce fine-motor control. Corticospinal neurons from the primary motor cortex, in association with neuronal output from the premotor area and the supplemental motor cortex, activate α motor neurons to execute voluntary motor commands.

First, a motivation to move is needed to spur an individual to action. Little is known about motivation; however, signals from the limbic system are thought to be important. Circuits between the basal ganglia and association areas of the cerebral cortex plan the intended movement. Somewhat different circuits are involved in new situations, such as first learning to type, than are involved in the execution of learned, but subconscious, patterns of movement (e.g., typing 100 words per minute). An important part of the planning process involves the somatosensory cortex, which provides information about the spatial coordinates of body parts in relation to the physical surroundings (e.g., placement of the fingers on the keyboard).

The cerebellum serves to adjust the timing and intensity of movements to improve the similarity between the intended and actual movements.

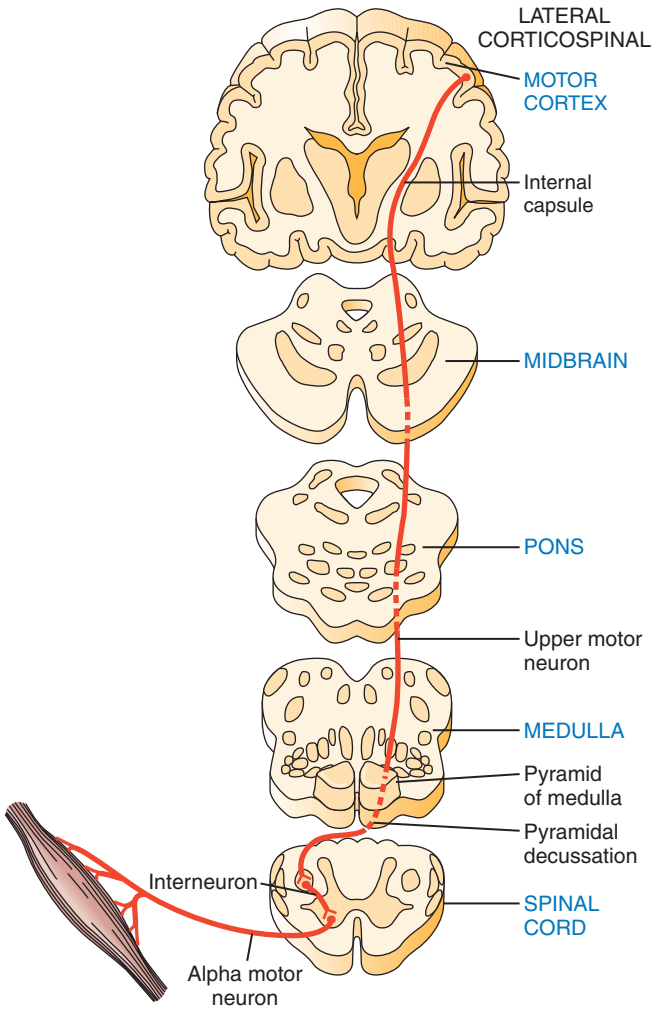


FIG 43.40 Corticospinal tract showing decussation at the level of the medulla.

Neurons in the cerebellum learn with practice. Visual, proprioceptive, and vestibular information is used by the cerebellum to make adjustments in the execution phase of the movement. In addition, the cerebellum functions with the cerebral cortex to make muscle adjustments in advance of the movement. For example, if a person is asked to lift what appears to be a pile of bricks, but the bricks are made of lightweight Styrofoam, the unsuspecting subject will throw the pile up into the air. Visual cues and past learned experiences are used to gauge the intensity of muscle contraction. The brain learns quickly, however, and if the subject is asked to repeat the maneuver, the intensity of muscle contraction will exactly match that needed to lift the load smoothly. Once a motor activity is well learned and can be performed "automatically," the cerebellum participates less and the basal ganglia and cortical neurons are most active.

KEY POINTS

- The body is somatotopically represented in the motor cortex. Stimulation of points in the primary motor cortex results in discrete movements in the contralateral side of the body.
- Projections from the motor cortex (corticospinal tract) travel by way of the internal capsule, cross over at the medulla, and travel down the contralateral spinal cord to synapse on α motor neurons in the anterior horn. The α motor neurons innervate skeletal muscle.
- Extrapyramidal tracts (tectospinal, vestibulospinal, reticulospinal) from subcortical nuclei innervate antigravity muscles and are primarily involved in balance, posture, and movement of large proximal muscle groups.
- A great deal of motor activity is preprogrammed into neuronal connections in the spinal cord. These connections produce reflexive alterations in muscle contraction in response to sensory information about the tension on the muscle or the need to move a body part away from a painful stimulus.
- The planning and execution of movements is accomplished through neuronal circuits between the basal ganglia and the premotor and association areas of the motor cortex. The cerebellum contributes primarily to the learning phase of a motor skill by making instantaneous adjustments in muscle force and timing to improve the match between the intended and the actual movement.

CONSCIOUSNESS, MEMORY, AND SLEEP

That the cerebral cortex is integral to the elaboration of complex thought, learning, memory, and so-called *higher brain functions* is well supported by lesion studies, yet little is known about the ways in which these higher functions are accomplished. Each thought involves neuronal circuits in portions of the cerebral cortex, thalamus, limbic system, and reticular formation of the brainstem. A thought or memory is not stored in any one place; rather, it is the outcome of a pattern or circuit of neuronal activation. Neurons in the limbic system, especially the amygdala, are thought to determine the general emotional value of the thought as being pleasant, painful, or neutral, whereas neurons in the cerebral cortex add the specific details, including remembered sensations and visual and auditory images. Consciousness and memory are prerequisites to thinking.

CONSCIOUSNESS AND MEMORY

Consciousness can be defined as awareness of the surroundings and of one's own thoughts and sensations. The neural correlates of consciousness are not known; however, continuous activation of neuronal circuits between the thalamus and cortex and between the thalamus and brainstem are believed to be necessary. Consciousness is assessed by the expression of motor behaviors, such as speech, response to questions, and body movements. However, behavioral responses are not necessarily an outcome of consciousness. A person completely paralyzed with neuromuscular blocking agents is still conscious even though all outward behaviors are suppressed. It is difficult to know the level of brain activity in an individual who cannot move. Brain waves are frequently measured in an attempt to assess brain activity, and a correlation with consciousness is assumed; α waves are thought to reflect search and retrieval functions, and θ waves are associated with memory encoding tasks. Groups of neurons firing synchronously produce the regular oscillations of brain waves that seem to underlie consciousness. Much remains to be discovered before the phenomenon of consciousness and its counterpart, unconsciousness, are understood.

To think and learn, one must be able to remember past events and link them to current circumstances. Memory is a synaptic phenomenon in which neurons in the memory trace or circuit alter the efficiency of synaptic transmission. Greater stimulation of neurons in the memory circuit results in longer-lasting effects. Once the memory trace is established, it can be reactivated by the thinking mind to reproduce the memory. Reactivation by the mind is called *retrieval* and may be enhanced by rehearsal or by strategies to link the memory trace with

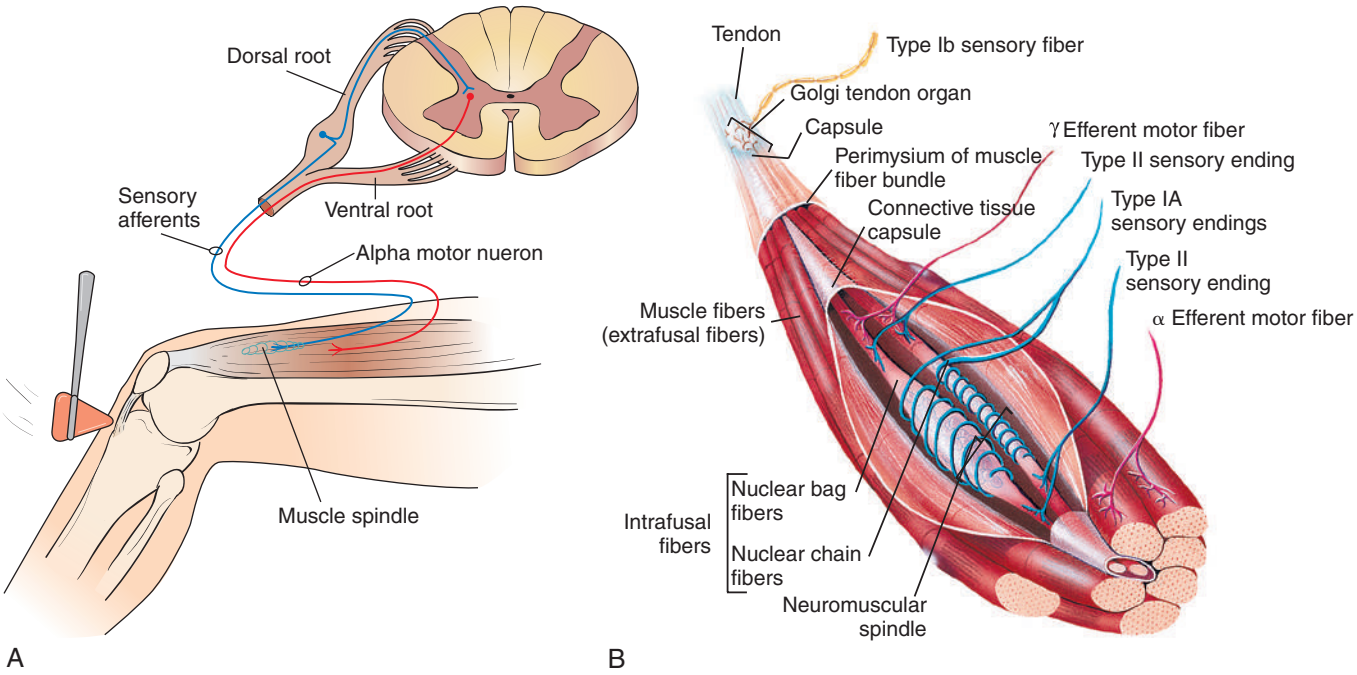


FIG 43.41 Diagram of the stretch reflex in which activation of the muscle spindle stimulates contraction of the stretched muscle. **A,** Stretch of the muscle sends action potentials to the cord, which make a monosynaptic connection to motor neurons from the same muscle fibers. **B,** Detailed view of the muscle spindle apparatus showing type I and type II sensory fibers that are large, myelinated, and rapidly conducting neurons that detect stretch. The γ motor neurons contract the spindle to keep it taut and sensitive to further stretch, whereas the α motor neurons contract the muscle fibers. Also shown in this figure is another stretch receptor located at the juncture between the muscle and the tendon called the *Golgi tendon organ.* Muscle shortening during contraction stimulates the Golgi tendon organ, causing it to send signals to the cord that inhibit muscle contraction. This is thought to protect the muscle from excessive contraction that could tear it from its tendon insertion points on the bone.

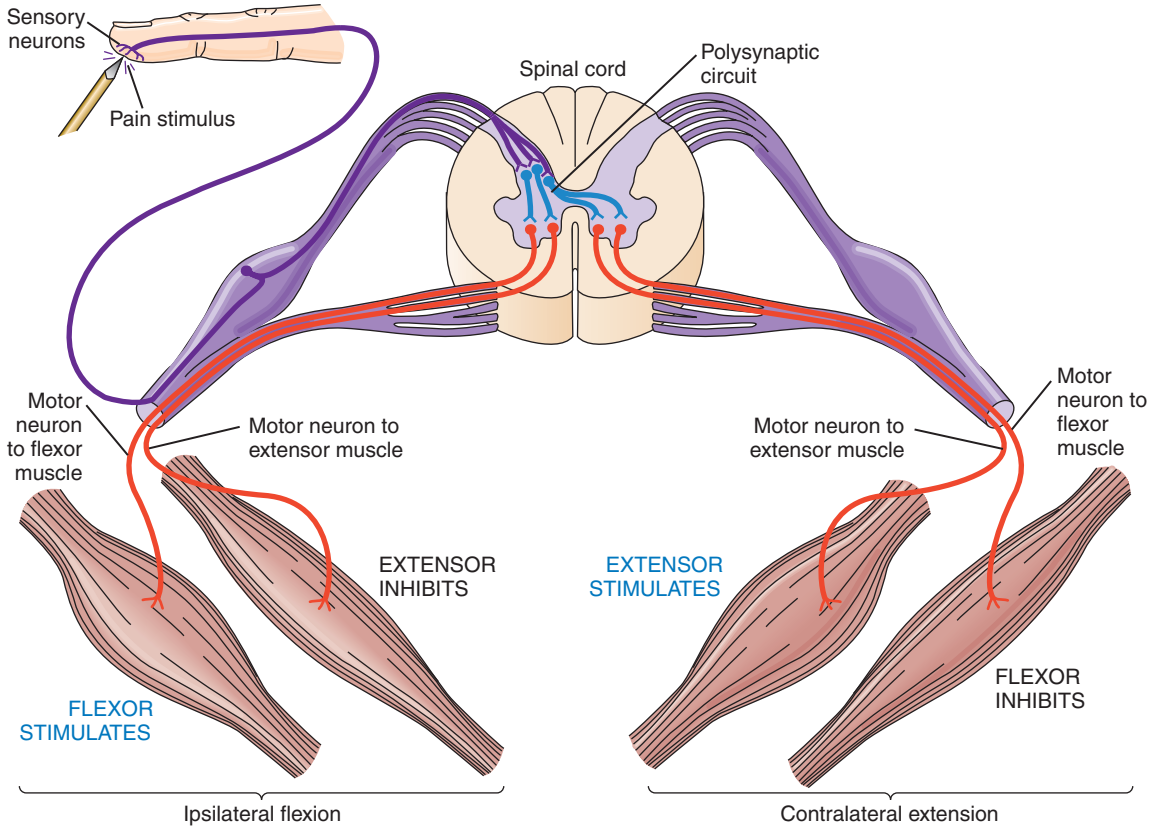


FIG 43.42 Neural connections mediating the flexor–withdrawal reflex.

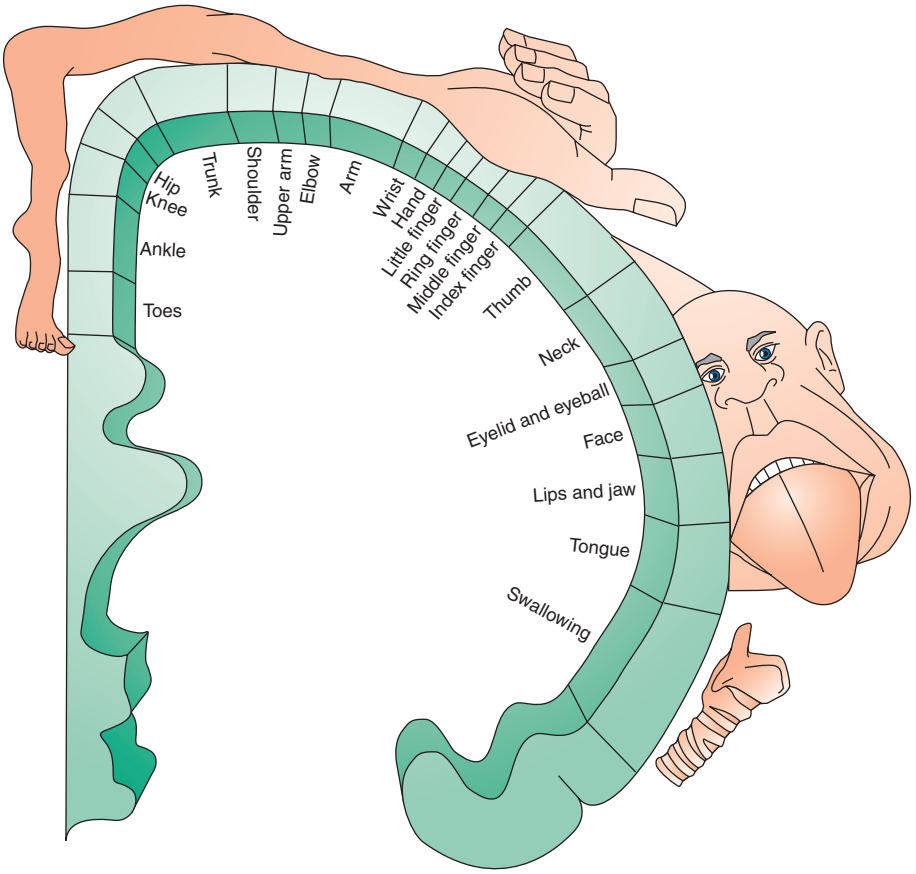


FIG 43.43 Cortical representation of the muscles of the body. Note the large area devoted to control of the hands and face.

associated circuits that are already established. Experiences that have important consequences, such as pain or pleasure, are usually enhanced and stored as memory traces, whereas experiences of little consequence may be suppressed. The value of a memory is determined primarily by the limbic system, which helps the brain learn to ignore information of little consequence (which can be construed as a form of negative memory, or remembering to ignore). This is an important function, because it prevents preoccupation and overload of brain circuits with useless stimuli.

Some memories last for a short time, and others persist for a lifetime. The mechanisms for different types of memory are mostly unknown; however, it is believed that short-term alteration of *presynaptic* neurons in the memory trace is responsible for short-term memory, whereas long-term memory requires more permanent changes in the *postsynaptic* neurons. A certain time period is required for memories to be consolidated into long-term memory. An interruption of the consolidation phase, by head trauma, for example, results in loss of memory for events occurring just before the injury.

Some examples of presynaptic modulation are shown in Fig. 43.44. The interaction between the presynaptic and postsynaptic neurons is modified by the activity of the modulating neuron. For example, the modulating neuron could send a signal from the limbic system to indicate that the incoming signal was important. The modulating neuron would then enhance neurotransmitter release from the presynaptic neuron to facilitate transmission to the postsynaptic cell. Presynaptic inhibition could suppress synaptic transmission by preventing or reducing the amount of neurotransmitter released from the presynaptic neuron.

Longer-term memory is thought to occur because of long-term changes in synaptic efficiency. In some cases memory may incorporate new neurons into the memory circuit. The hippocampus is an important limbic structure that must be intact for memory to be consolidated. It is also a site of neural stem cell proliferation, leading some to conclude that memory involves the growth of new neurons. Changes in synaptic efficiency could include alterations in receptor number or structure and changes in the components of second-messenger cascades. Protein synthesis in the involved neurons is necessary to consolidate long-term changes in synaptic efficiency. Memories are thought to be stored in the brain in association with memories of similar qualities that share some of the same neuronal circuits. Information is added to the memory circuit that is already in place. This makes it easier to learn information that is connected to previously learned material. Learning something completely new, such as a foreign language, requires a great deal of rehearsal.

SLEEP

Sleep is a state of decreased arousal from which a person is easily awakened. Different levels of brain activity from wakefulness to deep sleep are characterized by different electroencephalographic waveforms. Brain waves become progressively slower and more synchronized with deeper levels of sleep. These waveforms are called α, β, θ, and δ waves (Fig. 43.45). The α and β waves are found in awake individuals: α waves predominate during a relaxed state with the eyes closed; β waves occur during visual stimulation and with active problem solving. β waves are

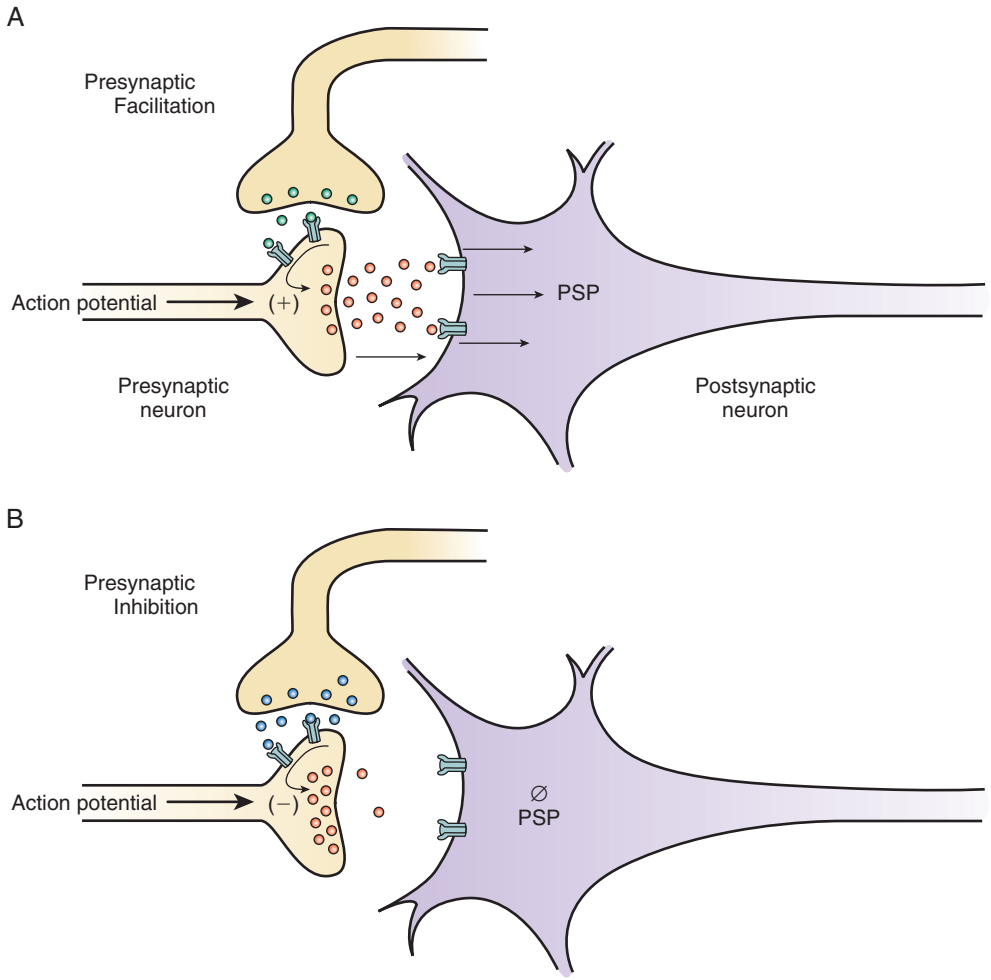


FIG 43.44 Example of a modulating neuron that terminates on the presynaptic cell and alters its response. **A,** Facilitation: The presynaptic neuron releases more neurotransmitter into the synapse when an action potential arrives. **B,** Inhibition: The presynaptic neuron releases less neurotransmitter into the synapse with each action potential. *PSP,* Postsynaptic potential.

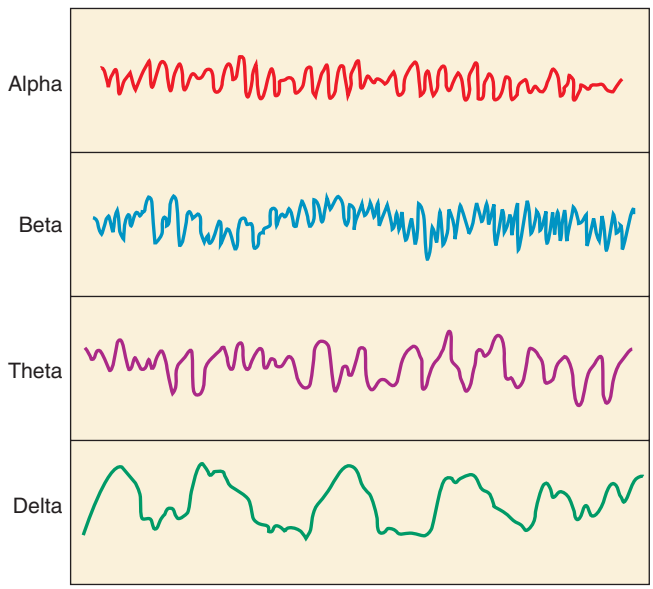


FIG 43.45 Brain waves are categorized according to frequency and synchrony as α, β, θ, and δ.

also apparent during a stage of sleep called *rapid eye movement* (REM) sleep. Both θ and δ waves occur during deep sleep.

Most sleep is of the restful, slow-wave type of deep sleep. Interspersed at about 90-minute intervals are episodes of REM sleep, which last 5 to 30 minutes. REM sleep is characterized by irregular breathing and heartbeat, rapid eye movements, depressed muscle tone, and active dreaming. The number and length of REM episodes usually increase over the course of the night. Individuals who have been awake for a prolonged period spend more time in deep sleep at the beginning of sleep and begin to have more REM sleep as the brain becomes more rested. Dreams occur in both types of sleep, but are more likely to be remembered when they occur during REM sleep.

The mechanisms of the sleep–wake cycle and the reason the brain needs to sleep are not well understood. Experiments in animals have demonstrated that sleep is an active process initiated by sleep-inducing substances within the brain. CSF removed from sleep-deprived animals promptly produces sleep when injected into another animal. Production of sleep-inducing substances within the brain occurs during wakefulness; they gradually accumulate, producing a desire to sleep. A period of sleep is thought to inhibit production of sleep-inducing chemicals, and they are cleared from the CNS.

The physiologic significance of sleep may be to rebalance synaptic transmission strength and to avoid the behavioral consequences of sleep deprivation. Prolonged sleep deprivation produces hallucinations, disordered thought processes, and personality changes. It has been noted that the smaller the animal and the higher the metabolic rate, the greater the sleep requirement. Non-REM sleep is associated with lowering of metabolism and body temperature and may provide an opportunity to avoid or repair metabolism-induced brain damage. REM sleep is associated with an active brain; however, certain types of neurons—those that secrete amines—are turned off during REM sleep. The younger and more immature brain spends more time in REM sleep, and some have speculated that the brain is laying down circuits for genetically programmed or instinctive neural pathways.

KEY POINTS

- Thoughts and memories are not stored in a particular location in the brain; rather, they are the outcome of activation of neurons in a neuronal circuit.
- Memories are stored by altering the synaptic efficiency of neurons in the memory trace. Short-term memory is thought to result from presynaptic mechanisms, whereas long-term memory is consolidated by more permanent changes in the postsynaptic neurons.
- Sleep is characterized as REM sleep (β waves) and deep sleep (θ and δ waves). Most sleep is of the deep-sleep variety, interspersed with periods of REM activity at about 90-minute intervals. Sleep is an active process produced by sleep-inducing chemicals in the brain. The reason that the brain requires sleep remains unknown.

SUMMARY

The nervous system is a complex network of neurons and supporting cells that provides the body with a rapid means of communication. The anatomy of the nervous system has been extensively studied, but the functional mechanisms of thought, memory, emotion, and sleep are poorly understood. The nervous system can be partially understood by examining the function of individual neurons and their synaptic connections. Neural communication occurs primarily through the secretion of neurotransmitters into synapses between neurons. Neurotransmitters bind to specific receptors to exert their effects on the membrane potential of the target cell.

Activation of groups of neurons in a particular circuit is the basis for neuronal processing of information, thoughts, and memories and for learning. Although certain brain locations have been associated with particular functions, most brain activities require participation by neurons in widespread locations. The ability of the nervous system to learn and adapt is remarkable in the early childhood period, and a significant degree of neural plasticity is maintained throughout life. However, damage to large numbers of neurons in a particular location usually results in significant disability because mature neurons cannot regenerate and neural stem cells may not survive or establish appropriate connections.

RESOURCES

Amtul Z, Atta-Ur-Rahman: Neural plasticity and memory: molecular mechanism. *Rev Neurosci* 26(3):253–268, 2015.

Cheslow L, Alvarez JI: Glial-endothelial crosstalk regulates blood-brain barrier function. *Curr Opin Pharmacol* 26:39–46, 2016.

Ezenwanne EB: Current concepts of neurophysiological factors in central regulatory mechanism of rapid eye movement sleep: a review. *West Indian Med J* 2015. doi:10.7727/wimj.2014.215. [Epub ahead of print].

Hall JE, editor: *Guyton & Hall textbook of medical physiology*, ed 13, Philadelphia, 2015, Saunders.

Hensch TK: Critical period plasticity in local cortical circuits. *Nat Rev Neurosci* 6(11):877–888, 2005.

Kumar V, et al: *Robbins and Cotran pathologic basis of disease*, ed 9, Philadelphia, 2015, Elsevier.

Meredith RM: Sensitive and critical periods during neurotypical and aberrant neurodevelopment: a framework for neurodevelopmental disorders. *Neurosci Biobehav Rev* 50:180–188, 2015.

Nolte J: *The human brain: an introduction to its functional anatomy*, ed 7, St Louis, 2016, Mosby.

Oberdick J, Sillitoe RV: Cerebellar zones: history, development, and function. *Cerebellum* 10(3):301–306, 2011.

Patton KT, Thibodeau GA, editors: *Anatomy & physiology*, ed 8, St Louis, 2016, Mosby.

Penhune VB, Steele CJ: Parallel contributions of cerebellar, striatal and M1 mechanisms to motor sequence learning. *Behav Brain Res* 226(2):579–591, 2012.

Rothwell JC: The motor functions of the basal ganglia. *J Integr Neurosci* 10(3):303–315, 2011.

Rubinson K, Lang E: The somatosensory system. In Koeppen BM, Stanton BA, editors: *Berne & Levy physiology*, ed 6, Philadelphia, 2010, Mosby, pp 105–122.

Saijo K, Glass CK: Microglial cell origin and phenotypes in health and disease. *Nat Rev Immunol* 11(11):775–787, 2011.

Smith Y, et al: The thalamostriatal systems: anatomical and functional organization in normal and parkinsonian states. *Brain Res Bull* 78(2-3):60–68, 2009.

van Niekerk EA, Tuszynski MH, Lu P, Dulin JN: Molecular and cellular mechanisms of axonal regeneration after spinal cord injury. *Mol Cell Proteomics* 2015 Epub.

Vincent SR: Nitric oxide neurons and neurotransmission. *Prog Neurobiol* 90(2):246–255, 2010.

Ward LM: The thalamic dynamic core theory of conscious experience. *Conscious Cogn* 20(2):464–486, 2011.

Yu JH, Seo JH, Lee JY, et al: Induction of neurorestoration from endogenous stem cells. *Cell Transplant* 2016. **DOI:** http://dx.doi.org/10.3727/096368916X690511.

44

Acute Disorders of Brain Function

Joni D. Marsh and Jacquelyn L. Banasik

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- What are the proposed mechanisms and potential consequences of primary and secondary brain injury?
- Which brain components determine intracranial pressure, and under what conditions might each contribute to elevated intracranial pressure?
- How are level of consciousness and cranial nerve reflexes used to assess changes in neurologic status in the brain-injured patient?
- What are the common manifestations of types of traumatic brain injury (focal, polar, diffuse) and hemorrhage (epidural, subdural, subarachnoid)?
- How do the three most common causes of stroke (thrombi, emboli, and hemorrhage) differ with regard to risk factors, prevention strategies, and acute management?
- How do the clinical manifestations of ischemic stroke vary depending on the location of cerebral artery obstruction?
- What are the common long-term sequelae of stroke, and how are they managed?
- What are the causes and usual presentations of cerebral aneurysm and arteriovenous malformation?
- How do brain abscess, meningitis, and encephalitis differ with regard to usual infective organisms, cerebrospinal fluid analysis findings, clinical manifestations, and treatment?

CHAPTER OUTLINE

Mechanisms of Brain Injury, 892
Ischemia and Hypoxia, 892
Cellular Energy Failure, 892
Excitatory Amino Acids, 893
Reperfusion Injury, 894
Abnormal Autoregulation, 895
Increased Intracranial Pressure, 895
Etiology, 895
Manifestations, 897
Brain Compression and Herniation, 897
Management, 899
Manifestations of Brain Injury, 900
Level of Consciousness, 900
Glasgow Coma Scale, 900
Cranial Nerve Reflexes, 901
Pupil Reflex, 901
Oculovestibular Reflex, 902
Corneal Reflex, 902
TRAUMATIC BRAIN INJURY, 902
Epidemiology, 902
Types of Traumatic Brain Injury, 903
Primary Injury, 903
Intracranial Hematomas, 903
Epidural Hematoma, 903
Subdural Hematoma, 904
Subarachnoid Hemorrhage, 905
Secondary Injury, 905
Treatment, 905
CEREBROVASCULAR DISEASE AND STROKE, 905
Epidemiology, 906
Ischemic Stroke, 906
Hemorrhagic Stroke, 906
Treatment, 906
Stroke Sequelae, 908
Motor and Sensory Deficits, 908
Language Deficits, 908
Cognitive Deficits, 908
CEREBRAL ANEURYSM AND ARTERIOVENOUS MALFORMATION, 909
Cerebral Aneurysm, 909
Etiology, 909
Pathogenesis and Manifestations, 910
Treatment, 910
Arteriovenous Malformation, 911
Etiology, 911
Pathogenesis and Manifestations, 911
Treatment, 911

CENTRAL NERVOUS SYSTEM INFECTIONS, 911
Meningitis, 911
Etiology, 911
Pathogenesis and Clinical Manifestations, 911
Treatment, 911
Encephalitis, 912
Etiology, 912
Pathogenesis and Manifestations, 912
Treatment, 912
Brain Abscess, 912
Etiology, 912
Pathogenesis and Manifestations, 912
Treatment, 913

Disorders of brain function can result from a wide variety of pathophysiologic processes. The focus of this chapter is on primary causes of acute brain injury, including brain trauma, cerebrovascular disease, brain hemorrhage, and central nervous system (CNS) infections. These conditions are acute because they generally have a sudden onset and progress rapidly. Thus early detection and prompt management are necessary to prevent death and minimize morbidity. However, the majority of patients who survive acute injury to brain structures will be left with some degree of permanent neurologic damage and chronic dysfunction. The designation of acute and chronic disorders is, therefore, somewhat arbitrary. The chronic aspects of neurologic diseases, including seizure disorders and dementias, are discussed in Chapter 45. Acute neurologic dysfunction often is a complication of diseases primarily affecting other systems. Hypoglycemia, renal failure, liver failure, human immunodeficiency virus infection, drug overdoses, fluid imbalances, and many other abnormalities may cause acute brain dysfunction. Accurate determination of the source of acute alterations in brain function is an important step in developing an appropriate treatment plan. However, there are many common elements in the pathogenesis of acute brain injuries, regardless of etiologic factors. These cellular aspects are presented as a foundation for understanding the specific disorders that follow.

MECHANISMS OF BRAIN INJURY

The mechanisms of brain injury are varied, complex, and incompletely understood. Mechanical trauma, ischemia, cellular energy failure, reperfusion injury, excitotoxins, edema, vascular failure, and injury-induced apoptosis (programmed cell death) are all factors thought to be operative in most kinds of acute brain injury. These mechanisms are often separated into two categories: *primary injury* and *secondary injury.*

The primary injury is that which occurs immediately at the onset of brain injury. This definition implies that there is little that can be done to reverse the process once it has occurred. For example, in the case of head trauma, some tissues will be irreversibly damaged at the time of impact owing to mechanical forces such as shearing, tearing, and stretching of neurons, axons, glia, and blood vessels. This damage represents the primary injury. Similarly, with the sudden cessation of blood flow to an area of brain tissue, as occurs in stroke, an area of irreversible ischemia in cells may develop quickly, and this constitutes the primary injury. Brain tissue necrosis occurs rapidly as cells lose membrane integrity, rupture, and release their intracellular contents into the extracellular space. Cytotoxic edema quickly follows, which can cause deleterious effects to surrounding brain tissue.

Secondary injury refers to the development of further neurologic damage and changes subsequent to the primary injury; these may progress over days and weeks and continue to cause changes to the brain for years. Delayed cell death may involve necrosis from further acute injury or may be a delayed consequence of the primary injury. Cells that die slowly after injury are said to undergo apoptosis, or programmed cell death. Apoptosis requires energy and protein synthesis, and the cells shrink and die in a tidy fashion without releasing their internal contents. Necrosis follows severe ischemic injury, whereas apoptosis is associated with moderate injury. A critical factor in determining the neuronal cell fate after injury is the degree of adenosine triphosphate (ATP) depletion. If ATP levels fall profoundly, increased membrane permeability and necrosis ensue. If the ATP level is partially maintained for a period of time after the injury, apoptosis is the likely consequence. Mild reductions in the amount of ATP are associated with reversible injury and cellular recovery. ATP level reduction is a consequence of ischemia and hypoxia, which accompany many types of acute brain injury, including trauma and stroke.

Mechanisms of secondary injury are the subject of much interest because of the potential to effectively intervene to limit brain damage. Unfortunately, effective means of preventing secondary damage have remained elusive, leading to high rates of mortality and morbidity. The effort to elucidate mechanisms of secondary injury and develop effective treatments is worthwhile. Secondary injury often leads to progressive neurodegeneration and delayed cell death, surrounding the primary insult and at times causing damage to areas some distance away. Thus, the high rates of mortality and morbidity may be attributed in large part to mechanisms of secondary injury.

Ischemia and Hypoxia

Ischemia occurs when the delivery of oxygenated blood is below the level needed to meet metabolic demands of the brain tissue. Ischemia is a contributing factor in most forms of acute brain injury, either as the primary insult (e.g., stroke) or as part of the secondary response to injury (e.g., vasospasm, vascular compression, or abnormal autoregulation). Hypoxia is a deficiency of oxygen at the cellular level, which may occur as a result of decreased blood flow (ischemia) or decreased blood oxygenation (hypoxemia). In practice, ischemia and hypoxia usually occur together and are considered together in this discussion. Ischemia results in immediate neurologic dysfunction because of the inability of neurons to generate the ATP needed for energy-requiring processes. In addition, ischemia sets the stage for secondary injury by oxygen free radicals, excitatory amino acids, and inflammatory cells.

Cellular Energy Failure

Neuronal tissue is highly sensitive to oxygen deprivation because it has great ATP requirements and limited capacity for anaerobic metabolism during ischemia. The normal brain receives about 15% of the total cardiac output and garners 20% of the body's oxygen consumption (750 mL/min), despite contributing only 2% of the body weight. Neurons are dependent on glucose for production of ATP; however, they store little in the form of glycogen. Thus, when oxygen supply is decreased, not only is oxidative phosphorylation impaired, but also the low supply of stored glucose restricts anaerobic production of ATP. Brain cells tolerate loss of ATP for several minutes; about 5 to 10 minutes of complete occlusion is necessary for irreversible brain damage in humans. Complete occlusion of blood flow is rare, but even a partial occlusion, if allowed to continue for a sufficient amount of time, may produce irreversible brain damage. Once blood flow to cerebral neurons diminishes, two

mechanisms can independently lead to brain cell death: anaerobic metabolism and deterioration of ion gradients.

A general sequence of events after acute brain ischemia has been proposed (Fig. 44.1). The critical event is mitochondrial dysfunction owing to lack of cellular oxygen. Recall that oxygen is required to accept electrons from the mitochondrial electron transport chain. In the absence of oxygen, the transport proteins and cytochromes remain reduced and unable to accept any more electrons from the Krebs cycle (tricarboxylic acid cycle). Anaerobic glycolytic pathways continue to function in the affected region to compensate for the loss of oxygen and to provide a source of energy. Glycolysis may continue for a short time, producing pyruvate, which is converted to lactate. However, this conversion releases H^+ and contributes to cellular acidosis, a damaging by-product of glycolysis. Toxicity of hydrogen ions leads to loss of neuronal integrity.

Inadequate energy supply leads to deterioration of ion gradients. Most of the ATP used by neurons is for maintenance of ion gradients across the plasma membrane. The sodium–potassium (Na^+–K^+) pump consumes three fourths of the ATP in a typical neuron. Anoxic depolarization causes potassium to leave the cell and sodium, chloride, and calcium ions to enter. Energy also is required to maintain calcium balance and regulate neurotransmitter synthesis and reuptake. Not surprisingly, energy failure results in neuronal dysfunction, injury, and, if severe or prolonged, necrotic cell death. Ischemia also is the probable inciting factor for apoptosis (see Chapter 4).

The mitochondria are also important regulators of calcium ion concentration in the cell. The mitochondrial membrane contains calcium transporters that sequester calcium ions within the mitochondria when cytoplasmic calcium levels are elevated. Mitochondrial energy failure impairs the ability of mitochondria to perform this sequestering function. Thus the mitochondria become severely overloaded with calcium, which activates enzymes (phospholipases) that damage mitochondrial membrane structures. Ischemic cells are prone to calcium overload because pumps that move calcium out of the cell are energy dependent. Calcium ions have a large electrochemical gradient for diffusion into the neuron and tend to accumulate intracellularly.

Unfortunately, the activity of many intracellular enzymes is regulated by intracellular calcium. Calcium overload is thought to be a critical factor leading to activation of enzyme cascades, which disrupt function and cause irreversible damage to cell membranes (lipid peroxidation).

Excitatory Amino Acids

Glutamate is a principle neurotransmitter in the adult CNS. It is important for rapid neuron-to-neuron communication. It also plays a role in neuron growth and brain development and maturation. Overstimulation of neurons by glutamate is associated with cell injury, leading to its designation as an *excitotoxin*. Glutamate binds two kinds of receptors that are linked to the opening of ion channels in the plasma membrane of neurons. *N*-methyl-D-aspartate (NMDA) receptors have received the most attention, but α-amino-3-hydroxy-5-methyl-4-isoxazolepropionate (AMPA) channels are also thought to contribute to the neurotoxic effects of glutamate. Activation of AMPA receptors results in opening of Na^+ channels in the membrane, which leads to depolarization. Depolarization then affects the NMDA channels by removing an Mg^{2+} ion that usually blocks the NMDA channel (Fig. 44.2). Subsequent binding of glutamate to the NMDA receptor opens it and allows Ca^{2+} to enter the cell accompanied by inflow of water, which results in cytotoxic edema, a rapid swelling of neurons. As previously described, calcium overload mediates a cascade of events leading to cell injury.

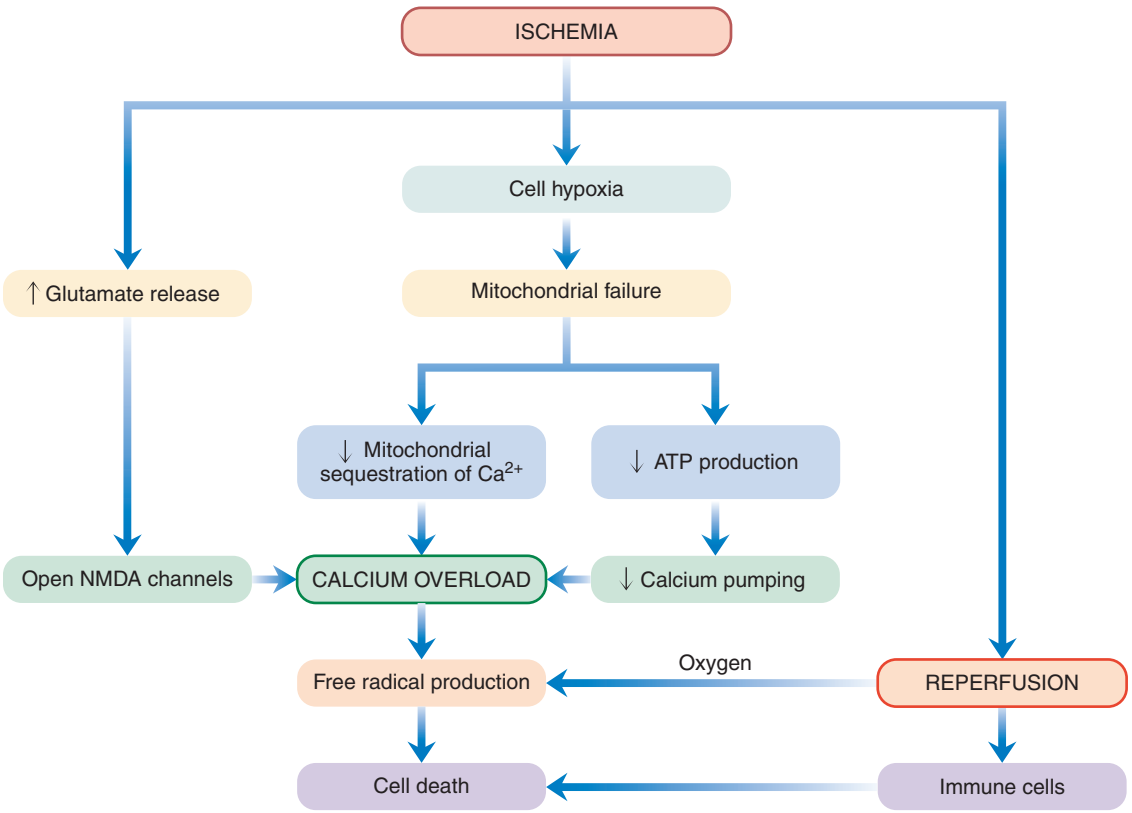


FIG 44.1 Sequence of neuronal cell injury after acute ischemia. Calcium overload is a key event in producing cellular damage.

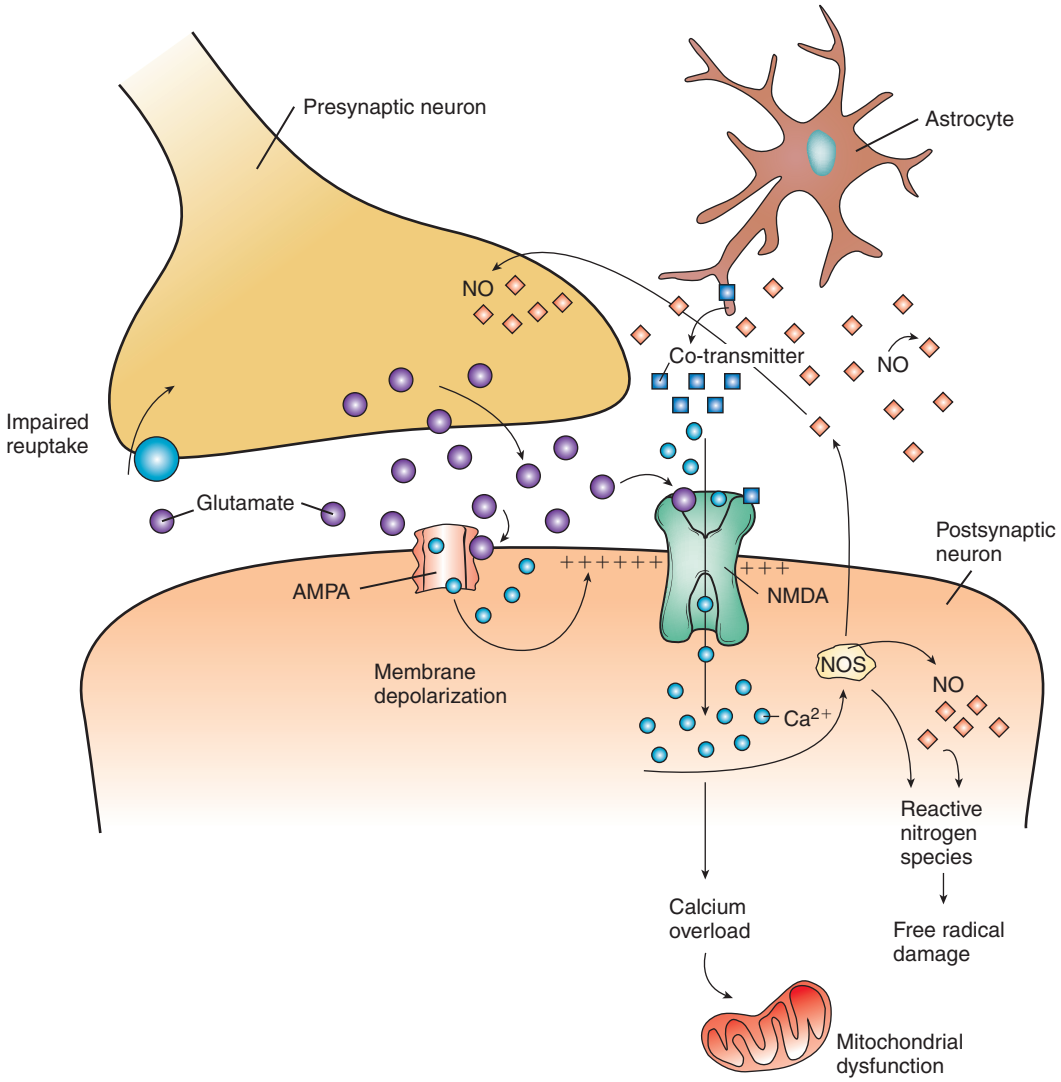


FIG 44.2 Mechanism of glutamate-mediated calcium influx. Impaired removal of glutamate from the synapse by energy-requiring reuptake mechanisms on the presynaptic membrane contributes to excessive glutamate in the synapse. Glutamate binds to the N-methyl-D-aspartate (NMDA) channel, causing it to open and allow calcium influx. A previous depolarization is necessary to remove the $Mg2^{+}$ that normally blocks the channel. Excessive calcium entry impairs mitochondrial function and triggers nitric oxide production. Excessive nitric oxide can increase cell damage through free radical production. *AMPA,* α-Amino-3-hydroxy-5-methyl-4-isoxazolepropionate; *NO,* nitric oxide; *NOS,* nitric oxide synthase.

The amount of glutamate in the synapses is usually tightly regulated by release and reuptake controls. In the presence of neuronal injury, excessive glutamate may be released because of impaired membrane integrity. With concomitant ischemia, reuptake mechanisms fail to remove excess glutamate from the synapse because they are energy-dependent processes. Excess glutamate stimulates nearby neurons, which then take up large amounts of injurious calcium ions. Small neurons in the cerebral cortex and hippocampus are particularly prone to glutamate excitotoxicity, and selective damage in these areas may occur. In addition to the calcium overload mechanism of injury, NMDA receptor activation stimulates nitric oxide (NO) production in neurons. NO is a neurotransmitter, but in excess it may increase the production of reactive nitrogen species (RNS), which function as free radicals to damage cellular components. The potential neuroprotective effects of controlling glutamate release or activity continues to be an area of active research.

Reperfusion Injury

Reestablishing perfusion to an area of prior ischemia is a matter of great urgency if neuronal tissue is to be salvaged. The longer and more severe the period of ischemia, the greater the extent of necrosis. However, previously ischemic cells face new dangers with the return of blood flow. In particular, the return of oxygen brings the potential for oxygen free radical formation, and the flow of blood allows inflammatory cells to invade the area. The secondary injury that occurs after reestablishing blood flow has been termed *reperfusion injury* and has been studied extensively in cardiac tissues. The mechanisms in the brain appear to be similar. During the period of ischemia, brain cells accumulate substrates for oxidative phosphorylation, including the free radical–forming metabolites of adenosine monophosphate (AMP), xanthine, and hypoxanthine. When oxygen reenters the cell, erratic transfer of electrons to oxygen can produce a number of reactive oxygen products

that behave as free radicals, damaging cell structures. These include hydroxyl radicals (OH•), superoxide (O_2–), and peroxide (H_2O_2). Cell membranes may undergo lipid peroxidation in response to free radical damage.

The role of immune mechanisms in reperfusion injury of brain tissue is only partially understood. Previously, the CNS was thought to be relatively shielded from immune cells because of the low permeability of the blood–brain barrier (BBB). However, the BBB is believed to be compromised with ischemia because the capillary endothelial cells are injured. The damaged BBB allows influx of systemic immune cells, thereby creating inflammatory cytokines, including interleukin-1 (IL-1), interleukin-6 (IL-6), and tumor necrosis factor (TNF). The brain's support cells, the microglia, are also activated by the injury, releasing signaling proteins that have a role in inflammation and repair of brain tissue. Trauma and inflammation also promote platelet aggregation in cerebral vessels with subsequent reduction in perfusion and worsening of ischemia. Accumulating laboratory and clinical evidence has implicated neuroinflammation in both acute damage and longer-term ongoing changes in the brain after injury.

Abnormal Autoregulation

Under normal conditions, blood flow through brain tissue is controlled primarily by autoregulation. Cerebral vessels respond to metabolic factors, including pH, carbon dioxide concentration, and oxygen levels. Cerebral blood flow is closely matched to metabolic needs despite wide fluctuations in perfusion pressure. Blood flow is maintained at a fairly constant rate over a range of mean arterial pressure (MAP) from about 60 to 140 mm Hg. Above and below these levels, autoregulation fails. Hypotension predisposes to ischemia, whereas severe hypertension may lead to vascular damage and brain edema.

Appropriate autoregulation is necessary to provide a steady supply of oxygen and nutrients to brain cells and to remove metabolic wastes. Cerebral vessels dilate when mean arterial blood pressure falls or when brain metabolism increases. Anything that interferes with the ability of the vessels to dilate can lead to ischemia. Thrombi, emboli, vasospasm, neutrophil aggregation, and tissue edema may inhibit vasodilating autoregulatory mechanisms. Alternatively, vascular injury may impair vasoconstricting mechanisms and allow hyperperfusion and edema formation. Depending on the cause, autoregulation may be impaired locally, as in an area of thrombosis, or globally, as in generalized cerebral edema.

Autoregulation is influenced by the partial pressures of carbon dioxide ($Paco_2$) and oxygen (Pao_2) in the arterial blood. The response to a change in $Paco_2$ is very brisk: as $Paco_2$ levels fall, cerebral vessels constrict, and as $Paco_2$ levels rise, the cerebral vessels dilate. A rise in $Paco_2$ can increase cerebral blood flow significantly. The response to changes in Pao_2 is much less dramatic.

The autoregulatory response to $Paco_2$ remains robust, except in severely brain-injured patients, and can cause detrimental increases in cerebral blood flow when respiratory compromise leads to hypercapnia. Excessive cerebral blood volume can exacerbate cerebral edema. Conversely, hyperventilation to produce low $Paco_2$ results in prompt vasoconstriction and, often, a reduction in intracranial pressure (ICP). Hyperventilation had been used for many years as standard therapy in the treatment of patients with increased ICP. However, prolonged hyperventilation does more harm than good because it critically reduces cerebral blood flow to responsive vessels and triggers tissue ischemia in these areas.

Loss of matching between oxygen supply and demand occurs when autoregulatory mechanisms fail. Cerebral oxygen demand is correlated with the degree of neuronal activity and may vary widely in different regions within the brain. Excessive levels of catecholamines or excitatory amino acids can significantly increase cerebral metabolism. In the context of impaired blood flow, these neurotransmitters may contribute to ischemia by increasing cerebral oxygen demand. Likewise, seizure activity increases neuronal metabolism and leads to worsening neuronal injury.

Efforts to reduce release of excitatory neurotransmitters through hypothermia, rest, and pain control may be beneficial. Pharmacologic suppression of brain seizures is imperative in the patient with cerebral ischemia. Drug-induced coma, with agents such as barbiturates or general anesthetic drugs, has been advocated to reduce brain metabolism. Hypothermia is a strategy for reducing brain metabolism and protecting the brain from ischemic injury. The effects of hypothermia on patient outcomes and the optimal degree of hypothermia remain controversial. Shivering negates the usefulness of hypothermia by increasing oxygen demand and must be suppressed, usually by pharmacologic means. Further research is ongoing to determine the therapeutic window for effectiveness, appropriate duration, and safe rewarming protocols.

Two related concepts important to the discussion of autoregulation are cerebral edema and increased ICP. Swelling and space-occupying lesions (mass lesions), such as tumors or hematomas, may increase the pressure within the cranium such that blood supply is compromised. Measures to reduce cerebral edema, remove mass lesions, and prevent elevations of ICP help to maintain functional autoregulation.

Increased Intracranial Pressure

ICP is the pressure exerted by the contents of the cranium, and it normally ranges from 0 to 15 mm Hg. Elevated ICP may occur in most types of acute brain injury and is associated with impaired neurologic function attributable to compression of brain structures. In all but very young children, the skull is a rigid, closed system with a set volume and a finite ability to accommodate increasing volume before elevations in ICP occur.

The volume of the cranium is made up of three components: brain tissue, cerebrospinal fluid (CSF), and blood. A relationship known as the *Monro–Kellie hypothesis* describes the compensatory responses to a change in volume in any of the three components. A slight increase in one component can be offset by a reduction in the volume of the other two. An increase in brain volume, as might occur with cerebral edema, can be offset by a reduction in the CSF space and the space occupied by the cerebral vasculature. The ability to accommodate changes in volume without significant increases in pressure is called *compliance.* Intracranial compliance is limited because of the rigid skull; although small increases in intracranial volume may be absorbed, moderate changes result in significant increases in ICP (Fig. 44.3).

Each of the cranial components has a varying capability to compensate for the others. Cerebral blood vessels can reduce their volume through vasoconstriction. The CSF compartment is capable of significant reduction in volume by shunting CSF to the spinal cord or into the venous system via the arachnoid villi. The brain parenchyma has little ability to reduce its volume to compensate for CSF or blood volume expansion. In young children, an increase in ICP may manifest as an increase in head circumference. This occurs because the cranial bones have not yet fused, and the skull can expand to accommodate the increased intracranial volume.

In the healthy brain, transient changes in ICP are common and well tolerated. Sneezing, coughing, straining, and head-dependent positions all cause elevated ICP, but they are without consequence because ICP quickly returns to normal. In the brain-injured person, however, transient elevations in ICP can be very dangerous and poorly tolerated.

Etiology

The most common causes of increased ICP include stroke, trauma, and tumors, but many other primary and secondary disorders can cause

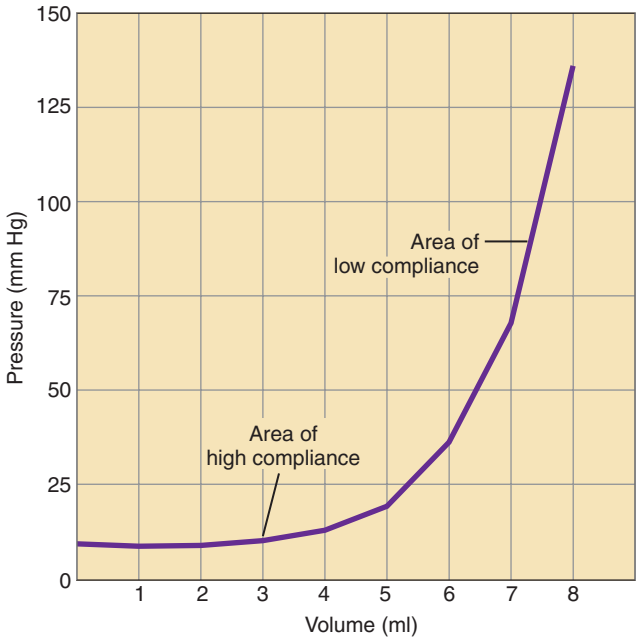


FIG 44.3 A volume–pressure curve showing intracranial compliance. Small increases in volume have little effect on pressure, but larger increases exceed the ability to compensate, and pressure rises dramatically.

BOX 44.1 Common Causes of Increased Intracranial Pressure

Increased Brain Tissue Volume
Tumor
Hemorrhage
Infection
Cytotoxic edema
Vasogenic edema
Ischemia and necrosis

Increased Cerebrospinal Fluid Volume
Obstructive hydrocephalus
Nonobstructive hydrocephalus
Pseudotumor cerebri

Increased Blood Volume
Increased right atrial pressure
Dural sinus thrombosis
High arterial $Paco_2$
Acidosis

significant elevations in ICP (Box 44.1). These disorders have common features in that they all affect the volume of CSF, blood, or brain tissue. An increase in brain tissue volume commonly occurs from conditions that cause cerebral edema. Edema of brain tissues is due to accumulation of fluid in interstitial or intracellular spaces.

Interstitial edema is usually secondary to an increase in capillary pressure, damage to the capillary endothelium from a chemical injury, or a sudden increase in vascular pressure beyond autoregulatory limits. This type of edema has been termed *vasogenic*, and it results in extravasation of electrolytes, proteins, and fluid into the intercellular space. Vasogenic edema is a consequence of stroke, ischemia, and severe hypertension and may occur surrounding brain tumors. The edema is often localized to a particular brain region where the BBB has been disrupted. Thus the swelling may be unilateral, occurring in one brain hemisphere or the other. Unilateral swelling often is poorly tolerated because midbrain structures are compressed and can lead to shifting of cerebral tissue or brain herniation.

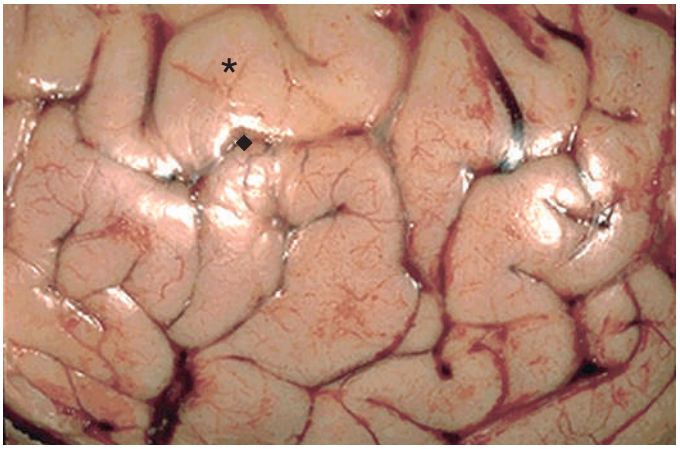


FIG 44.4 Cerebral edema. The cortical surface beneath the meninges of this brain with cerebral edema shows widened, flattened gyri (*) with narrowed sulci (♦). (From Klatt E: *Robbins and Cotran atlas of pathology,* ed 3, Philadelphia, 2015, Elsevier Saunders.)

Intracellular edema, called *cytotoxic edema,* occurs when ischemic tissue swells because of cellular energy failure. A lack of ATP allows Na^+ to accumulate in the cell, creating an osmotic force to draw in water. Cytotoxic edema predominates in cases of global ischemia. Global ischemia occurs when oxygenation of the whole brain is impaired, as would occur with cardiac arrest or severe hypoxemia. Generalized brain edema flattens the gyri and reduces the spaces between them (Fig. 44.4).

In many cases of acute brain injury, vasogenic and cytotoxic edema occur together. Cerebral edema, when severe, can start a cyclic process that promotes further edema of increasing severity and contributes to increased ICP. As edema fluid collects, it compresses local vessels, preventing adequate blood and oxygen from reaching the cells. This results in ischemia, which in turn triggers vasodilation and increased capillary pressure, further fluid leakage into the injured tissue, and increased edema. Vasogenic edema tends to be a delayed process in terms of the secondary effects of brain injury, progressively worsening during the first several days after injury.

In addition to edema, a number of space-occupying processes, such as tumors, hematomas, and abscesses, can increase intracranial volume and contribute to elevated ICP. These mass lesions are often unilateral and may result in severe compression of vital brain structures. Attempts by the brain to accommodate the expanding mass result in typical findings on computed tomography (CT) scans (Fig. 44.5). The ventricles are reduced in size, and midline structures are displaced.

Excessive accumulation of CSF (hydrocephalus) is another important cause of increased ICP. Elevated CSF volume causes the ventricles to enlarge and press on cerebral brain structures (Fig. 44.6). Hydrocephalus may be a primary disorder or may develop as a result of obstruction proximal to the arachnoid granulations (obstructive; noncommunicating hydrocephalus) or at the level of the arachnoid granulations (communicating; nonobstructive hydrocephalus). Obstructive hydrocephalus commonly occurs when a lesion blocks the flow of CSF out of the ventricle, whereas nonobstructive hydrocephalus is common after subarachnoid hemorrhage because residual blood clogs the arachnoid villi and prevents the CSF from being reabsorbed.

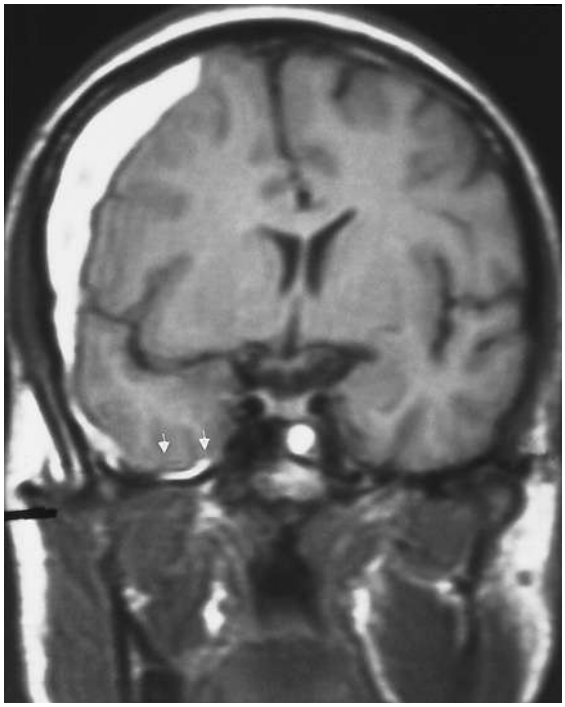


FIG 44.5 Right subacute subdural hematoma on weighted CT scan. Note the shift in midline structures. (From Yousem DM, Grossman RI: *Neuroradiology,* ed 3, St Louis, 2010, Mosby, p 174.)

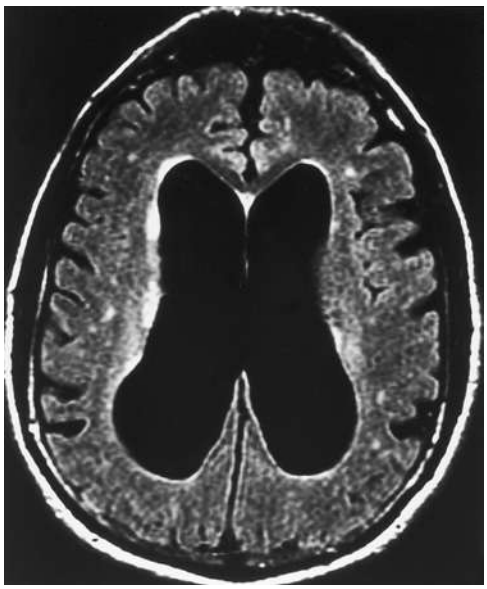


FIG 44.6 Normal-pressure hydrocephalus. (From Yousem DM, Grossman RI: *Neuroradiology,* ed 3, St Louis, 2010, Mosby, p 255.)

Increased intravascular cerebral blood volume is unlikely to be a primary cause of high ICP, but it may contribute to pressure elevations initiated by ischemia or trauma. High $Paco_2$ or loss of autoregulatory controls can lead to vasodilation and increased cerebral blood volume.

Manifestations

Manifestations of elevated ICP include headache, vomiting, and altered level of consciousness (drowsiness). The patient may complain of blurry vision, and evaluation of the fundi may reveal edema of the optic disk (papilledema). As ICP rises to higher levels, the level of consciousness decreases, and pupil responsiveness to light becomes impaired. Eventually the patient will exhibit altered respiratory patterns and will become unresponsive to stimulation and unable to move, verbalize, or open the eyes. Prolonged elevations of ICP are thought to damage brain structures by compressing the blood supply and causing ischemia.

Patients exhibiting manifestations of elevated ICP, or those with significant risk for elevated ICP, may be monitored with a pressure device inserted into the brain parenchyma through an opening in the skull (burr hole). The pressure device is connected to an electrical transducer, and the ICP waveforms can be monitored continuously (Fig. 44.7). In general, a high ICP is associated with poor outcome. Different ICP waveforms are thought to carry different prognoses. The normal ICP waveform is characterized by three pressure peaks called *P1, P2,* and *P3.* These waves are reflections of changes in ICP associated with each arterial pulsation. P1 is the first and generally the tallest peak. It is known as the percussion wave and corresponds to the transmitted systolic blood pressure. P2 has a more variable shape and reflects relative brain volume. The P2 component is often elevated in response to rapidly expanding mass lesions such as hematoma or tumor. P3 is the smallest tidal wave. As the ICP increases, the P2 and P3 rise and eventually surpass P1. Ultimately with continued elevation of ICP the waveform loses its distinct peaks and assumes a triangular morphology (see Fig. 44.7).

Rounded or monotonous waves reflect severe pathologic increases in ICP attributable to changes in cerebral volume. Plateau waves can reach 50 to 100 mm Hg. After the plateau period the ICP slowly decreases, but it usually remains elevated above baseline. These waves reflect a potentially life-threatening situation, and if the pathologic process is not stopped, a cycle of increased ICP followed by vasodilation to maintain constant blood flow through swollen tissues continues, which in turn further increases ICP. An extreme increase in ICP can precipitate an intense reaction by the sympathetic nervous system as it attempts to maintain cerebral perfusion through the compressed blood vessels. This has been termed an *ischemic response* or *Cushing reflex.* The systolic blood pressure can jump to values exceeding 200 mm Hg, accompanied by bradycardia and a widening pulse pressure. The Cushing reflex generally is viewed as a "last ditch" effort by the brain to reestablish cerebral perfusion.

Brain Compression and Herniation

A dreaded complication of elevated ICP is brain compression and herniation. Compression of midbrain and brainstem structures is associated with rapid neurologic demise unless corrected quickly. Important midline structures include the reticular activating system (RAS), which is necessary for maintaining consciousness, and vital regulatory centers for cardiovascular and respiratory control. Radiologic examination by CT scan or other means (e.g., magnetic resonance imaging [MRI]) is useful in evaluating the patient with increased ICP who exhibits a change in neurologic status. CT scans may show midline shifts and herniations when ICP is sufficiently elevated.

Herniation refers to the protrusion of brain tissue through an opening in the supporting dura of the brain. Several types of brain herniation have been described according to their anatomic locations. The brain parenchyma is divided into compartments by the supporting structure of the dura. The dura folds into the space between the cerebral hemispheres to form the falx cerebri and folds in from the lateral aspects to form the tentorium, which separates the cerebellum from the cerebral hemispheres (Fig. 44.8). The most common herniations occur through openings in these structures (Fig. 44.9).

There is a small space between the tip of the falx cerebri and the corpus callosum through which the neurologic lobe of the cerebral

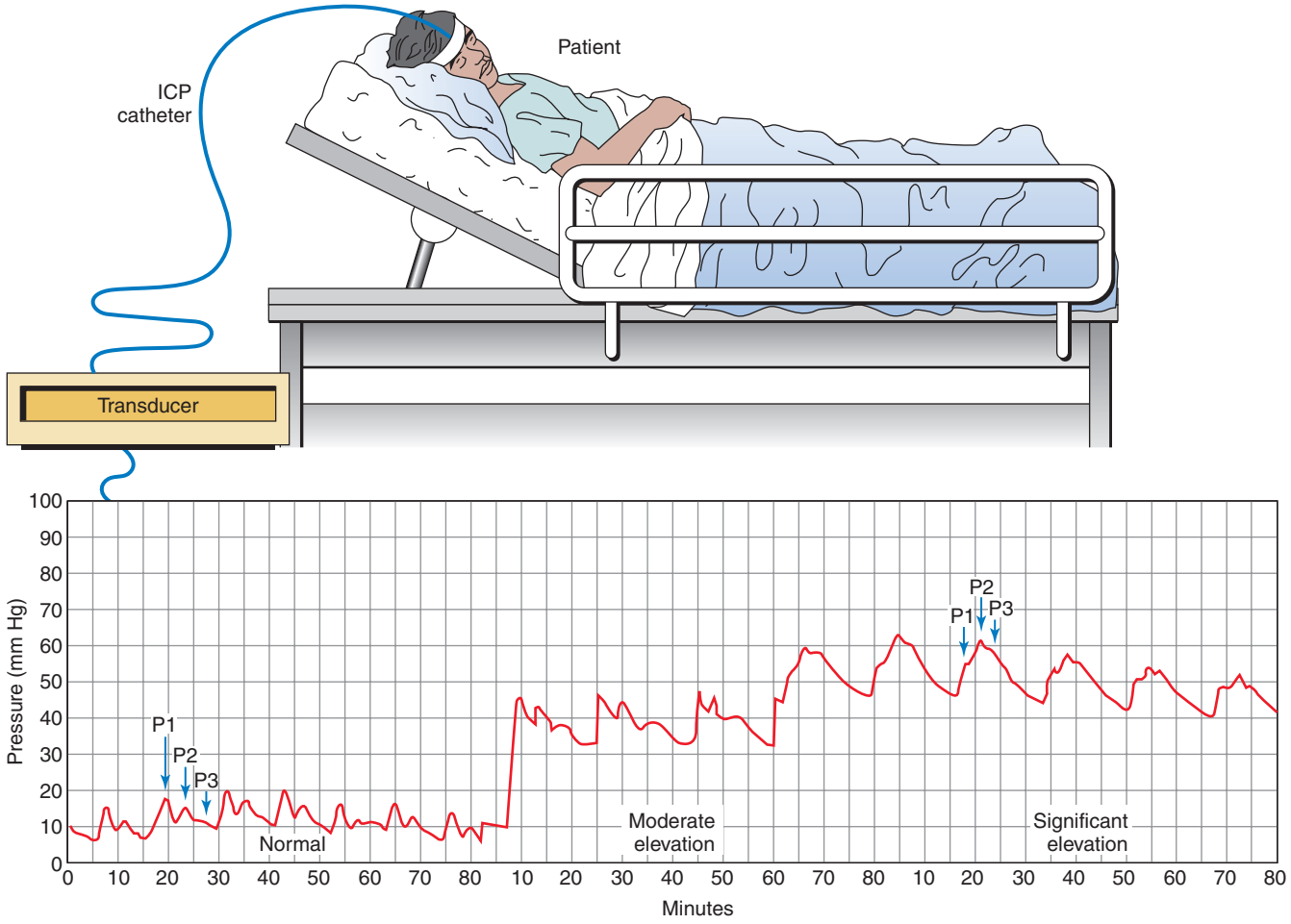


FIG 44.7 Intracranial pressure (ICP) monitoring can be used to continuously measure ICP. The ICP tracing shows normal, elevated, and plateau waves. At high ICP the P2 peak is higher than the P1 peak, and the peaks become less distinct and plateau.

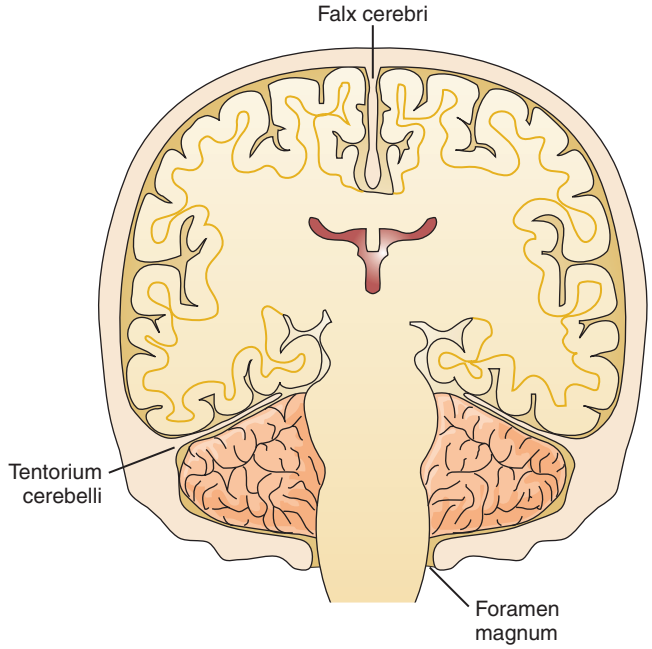


FIG 44.8 Schematic drawing of the normal brain compartments showing the dural folds that form the falx cerebri and the tentorium.

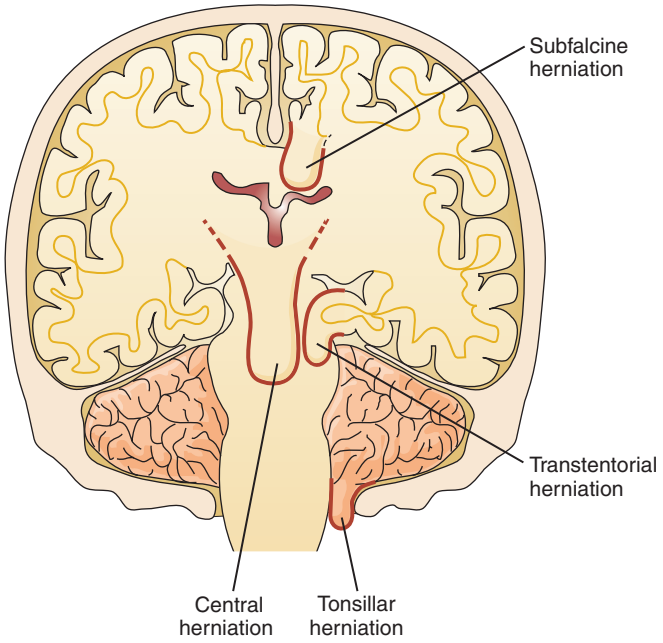


FIG 44.9 Herniations occur when brain tissue is pushed through openings beside the dural folds or the foramen magnum.

FIG 44.10 CT scan of acute hemorrhage with mass effect and subfalcine herniation and shift of the lateral ventricle. (From Yousem DM, Grossman RI: *Neuroradiology,* ed 3, St Louis, 2010, Mosby, p 185.)

cortex can herniate (Fig. 44.10). This is called a *subfalcine hernia.* Subfalcine herniation occurs when a lesion in one hemisphere is large enough to cause a lateral shift across the midline of the intracranial cavity, forcing the neurologic gyrus under the falx cerebri. This results in distortion and compression of the internal cerebral vein. Subfalcine herniation can be asymptomatic and generally carries a better prognosis than other types of brain herniation. The greatest danger results from compression of blood vessels, particularly the ipsilateral anterior cerebral artery, which can cause further cerebral ischemia and edema and contribute to the ICP elevation.

The tentorium is a rigid dural fold that separates the cerebellum and cerebral hemispheres. Midbrain structures pass between the infoldings of the dura in a structure called the *incisura.* With transtentorial herniations, a part of the brain protrudes through this space. Tentorial herniations are of two types: (1) bilateral herniations, which cause central transtentorial herniation, and (2) lateral herniations, in which one hemisphere compresses midbrain structures to the side and herniates through the tentorial opening (see Fig. 44.9).

Central tentorial herniation results from expanding lesions in the frontal, parietal, and occipital lobes that force a downward displacement of the hemispheres and basal nuclei with compression of the diencephalon and adjoining midbrain. Transtentorial herniation can occur rapidly or slowly, depending on the type of lesion. The speed with which the process is recognized is a critical factor in patient survival. Slowly dilating or odd-shaped pupils is an ominous sign that indicates compression of the third cranial nerve and midbrain. Transtentorial herniations are associated with significant intracranial hypertension and may initiate vascular compression and CSF obstruction, which then contribute to the existing problem of ischemia and hypertension.

Uncal herniation is a type of tentorial herniation that typically occurs with expanding lesions in the temporal lobe. As the lobe shifts, the basal edge of the uncus and the hippocampal gyrus bulge over the edge of the incisura (see Fig. 44.9). In the process, the third cranial nerve and the posterior cerebral artery are compressed. The pupil on the same side (ipsilateral) as the lesion often becomes dilated and unresponsive to light (fixed). Flattening of the midbrain interferes with the ascending RAS and depresses the level of consciousness. The ipsilateral cerebral peduncle is also compressed, resulting in contralateral motor dysfunction. Compression of the contralateral cerebral peduncle is also common, leading to the confusing symptom of ipsilateral motor dysfunction.

Tonsillar herniation is less common than the other herniation syndromes and involves the shift of the cerebellar tonsils through the foramen magnum and compression of the medulla and upper cervical cord (see Fig. 44.9). This typically occurs in patients with cerebellar lesions. Because of the proximity of the cerebellum to the brainstem, tonsillar herniation evolves very rapidly and can result in death in a matter of minutes. Signs usually include precipitous changes in blood pressure and heart rate, small pupils, disturbances in conjugate gaze, ataxic breathing, and quadriparesis.

Management

Management of increased ICP is often based on the results of CT or MRI. Processes amenable to surgical intervention can be detected and treated. Removal of excess CSF, tumors, abscesses, and hematomas can dramatically improve ICP. Nonsurgical processes such as cerebral edema, intracerebral bleeding, and infections are managed medically. ICP measurements and determinations of cerebral perfusion pressure (CPP) are used to guide therapy. Controversy regarding the appropriate treatment of increased ICP continues, and the focus of management has shifted from ICP control to management of cerebral oxygenation. The roles of previously established therapies such as hyperventilation, brain dehydration with diuretics, head-up and neutral body positions, and corticosteroid administration have been called into question. Previously discarded therapies, such as hypothermia, hypertonic saline infusion, and drug-induced comas, have been reintroduced. Newer interventions include measurement of cerebral blood flow by transcranial ultrasonography. Measurement of brain tissue oxygenation via jugular venous bulb oximetry and oxygen-15 positron emission tomography (PET) are also being utilized. Cerebral microdialysis measuring cerebral tissue metabolism by intraparenchymal probe is being used in research settings.

Despite these controversies in medical management, the value of careful observation and assessment of neurologic function is unquestioned. Many times subtle changes in neurologic function are detectable early in the process of evolving brain injury. Several tools have been developed to help standardize neurologic examination and are discussed in the following section.

KEY POINTS

- Primary brain injury occurs as a direct result of the initial insult. Secondary injury refers to progressive damage resulting from the body's physiologic response to the initial insult.
- Ischemia is an important mechanism of brain injury that occurs when the blood supply is inadequate to meet metabolic needs. A lack of oxygen results in mitochondrial failure, ATP depletion, and accumulation of intracellular calcium ions.
- Excessive release of excitatory amino acids, like glutamate, is thought to contribute to calcium overload during acute brain injury. Calcium overload is a critical event leading to cell dysfunction, membrane damage, and cell necrosis.
- Reperfusion injury occurs when blood flow is reintroduced to previously ischemic but viable cells. Free radicals are generated, which damage cell structures. Inflammatory cells are recruited to the area and may increase edema, block vessels, and contribute to free radical production.

- Autoregulation of cerebral blood flow achieves appropriate flow to meet metabolic needs despite changes in blood pressure and metabolism. Autoregulation is effective over a range of mean arterial blood pressures from 60 to 140 mm Hg. Hypoxia and high $Paco_2$ result in dilation of cerebral vessels. Hyperventilation with low $Paco_2$ results in cerebral vasoconstriction.
- Pressure in the cranium is a product of the volume of brain tissue, blood, and CSF. Increases in any one component are partially offset by reductions in the others to maintain ICP.
- Brain swelling is a common cause of increased ICP. Edema may result from changes in vascular competency that lead to transudation of fluid into intercellular spaces (vasogenic) or from cellular swelling (cytotoxic) due to a deficiency in cellular ATP.
- Normal ICP ranges from 0 to 15 mm Hg. Transient increases are well tolerated, but chronically increased ICP results in compression of vessels and brain tissue, leading to cellular ischemia and brain damage. High ICP may precipitate herniation of brain tissue through dural compartments.

MANIFESTATIONS OF BRAIN INJURY

Depending on the severity and location of brain injury, a wide variety of clinical manifestations may occur. Patients may present with symptoms ranging from minor headache and visual disturbances to complete loss of consciousness. Patients with significant acute injuries require frequent neurologic assessments to detect changes that may evolve rapidly. Level of consciousness (LOC), cranial nerve reflexes, and brain hemodynamics provide important clues to neurologic status.

Level of Consciousness

A change in LOC is one of the most sensitive indicators of altered brain function. Efforts have been made to standardize the terms used to describe LOC (Table 44.1). In practice, however, it is best to use a full description because of lack of consistency in interpretation of the terms. Consciousness is a state of alertness and attentiveness to one's environment and situation. A fully conscious individual is awake, alert, and oriented to time, person, place, and current circumstances. Consciousness is thought to be dependent on activity in the RAS neurons, which project to the thalamus, and in tracts between the thalamus and cortex. Although consciousness may be suddenly and completely lost, the decline is usually progressive. Cortical neurons are most sensitive, and cognitive and memory functions are impaired early, leading to confusion.

Delirium, a primary disorder of attention, is characterized by an acute onset of severe confusion, motor signs, slurred speech, altered consciousness, and hallucinations. As RAS function is compromised, the patient becomes difficult to arouse and requires increasingly noxious stimuli to produce verbal or motor responses. Eventually complete loss of consciousness may occur, a condition called *coma.* The Glasgow Coma Scale (GCS) can be used to assess LOC with greater reliability among different observers (Box 44.2). Sudden or progressive changes in LOC should prompt a thorough neurologic examination to determine the cause and best course of therapy.

Glasgow Coma Scale

The GCS is a standardized tool developed for the purpose of assessing the LOC in acutely brain-injured patients. It can also be used to evaluate patients with an altered LOC as a result of other neurologic insults such as hemorrhage or craniotomy. Numeric scores are given to arousal-directed responses of eye opening, verbal utterances, and motor reactions. The best response is scored, bilateral responses are recorded for motor reactions, and consistent application of a painful stimulus is required for accuracy. When used correctly, the GCS has a high degree of interrater reliability.

The eye opening response is a simple measure of alertness. Normally, the eyes open spontaneously in response to verbal stimuli. If the eyes do not open in response to verbal stimuli, noxious stimuli, such as compression of the nail beds, may be applied. It is important to be consistent and vigorous enough to achieve the best response from the patient. In patients with acute space-occupying lesions, eye opening is usually depressed in conjunction with impaired response to pain and motor function. Spontaneous eye opening in the acute phase is an encouraging sign, as it implies that the arousal mechanism in the brainstem is intact.

The verbal response on the GCS reflects orientation. A full score (i.e., 5) in this category indicates that the patient is alert and fully oriented: the patient knows his or her name, current location, and the time of day. In the next level (score 4), the patient is awake and can pay attention to a certain degree, but is confused about his or her

TABLE 44.1 Terms Used to Describe Altered Level of Consciousness

Term	Description
Confused	Unable to think clearly or engage in effective problem-solving; orientation to time, place, person impaired; easily aroused by verbal stimuli
Delirious	Restless and disoriented, may have hallucinations; easily aroused, but may have difficulty with attention
Lethargic	Uninterested in surroundings or events; sluggish in thought and motor activities; does not engage spontaneously in activities
Obtunded	Falls asleep unless stimulated; arousable with voice or touch, but quickly returns to sleep
Stuporous	In a deep state of sleep; vigorous stimulation is required to arouse, and a wakeful state is not maintained
Comatose	Unable to be aroused, even with vigorous painful stimuli; motor responses, such as withdrawal or posturing, may occur

BOX 44.2 Glasgow Coma Scale

Eye Opening
4. Spontaneously (eyes open, does not imply awareness)
3. To speech (any speech, not necessarily a command)
2. To pain
1. Never

Verbal Response
5. Oriented (to time, person, place)
4. Confused speech (disoriented)
3. Inappropriate (swearing, yelling)
2. Incomprehensible sounds (moaning, groaning)
1. None

Motor Response
6. Obeys commands
5. Localizes pain (deliberate or purposeful movement)
4. Withdrawal (moves away from stimulus)
3. Abnormal flexion (decortication)
2. Extension (decerebration)
1. None (flaccidity)

identity and does not know the time or the location. If attention is poor and the verbal responses consist of yelling and swearing, it is scored as an inappropriate response (i.e., score 3). Incomprehensible verbalizations are unintelligible sounds or mumbling (i.e., score 2). Absence of verbalization is given a score of 1. As in the eye-opening category, noxious stimuli are applied to achieve the best verbal response from the patient.

Motor response is a powerful predictor of outcome. Motor response is scored as the best level of response the patient is able to perform. Each extremity is evaluated to avoid misinterpretation secondary to muscle paralysis. At the highest level (score 6), the patient can obey a command to move. At the next level down (score 5), the patient does not obey commands, but when a painful stimulus is applied, the patient moves in a purposeful manner to avoid the stimulus; the patient is able to localize the source of pain. As status deteriorates, the patient withdraws only the extremity from the painful stimulus (score 4). This is not considered a purposeful response. Further deterioration results in abnormal posturing movements. *Decorticate* posturing (score 3) is characterized by an abnormal flexor response of the arms and wrists. The legs and feet extend and internally rotate (Fig. 44.11). The level below decorticate posturing is called *decerebrate* or *abnormal extension* (score 2). The arms extend with external rotation of the wrists. The legs and feet extend and rotate internally (see Fig. 44.11). The lowest level of motor response is no response to painful stimuli (i.e., flaccidity in all four limbs, score 1). It is important to emphasize that all limbs must be tested separately, because motor responses may be preserved on one side only, and levels of involvement may vary from side to side. Higher initial scores tend to predict better recovery. On the GCS, the level of coma occurs on a continuum from mild (>12), to moderate (9 to 12), to severe (<8). The lowest total score of 3 indicates likely fatal damage, especially if both pupils fail to respond to light and oculovestibular responses are absent; however, the severity and prognosis are predicted more accurately by also considering diagnostic imaging and other factors.

Cranial Nerve Reflexes

The GCS score alone is not sufficient to accurately determine the status of the patient with an acute insult to the brain. Assessment of the integrity of brainstem function is also important and is indicated by various brainstem reflexes, including the pupil light reflex, oculovestibular reflex, and corneal reflex.

Pupil Reflex

Pupillary assessment provides important information about the function of the brainstem and cranial nerves II and III. The normal pupillary response to light results from an intact afferent cranial nerve II (optic) detecting the light and stimulating the intact efferent cranial nerve III (oculomotor) to constrict the pupils. The response of the pupil to light, in terms of both its shape and the speed of reaction, is a function of cranial nerve III.

The pupillary response is recorded by noting pupil size in millimeters, shape, and reactivity to light. Careful monitoring of the pupillary response to light during the acute phase is critical, because a failing response may be the first indication of brain compression from increasing ICP. Mild dilation of a pupil with sluggish or absent light response is ominous. This phenomenon results from pressure on the oculomotor nerve (cranial nerve III) by lateral displacement of midbrain structures. An oval pupil may be an early indicator of dangerously poor compliance and transtentorial herniation. The oval pupil represents a transitional pupil that can return to normal responsiveness if ICP is controlled.

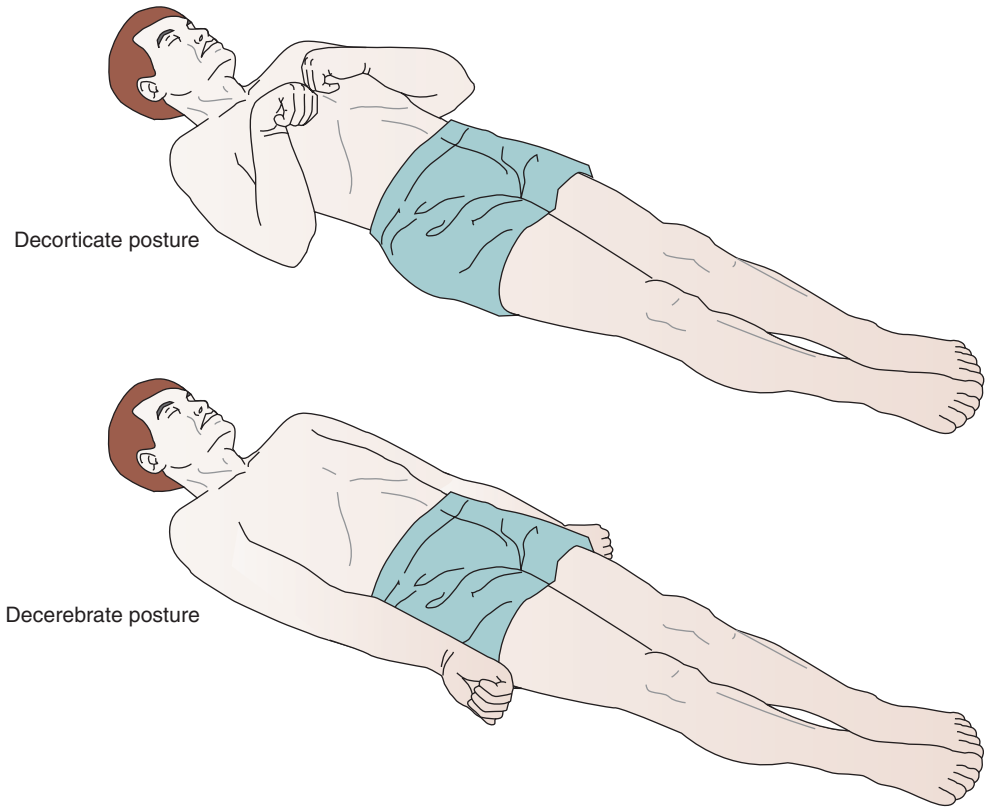


FIG 44.11 Abnormal motor activity with coma. Decorticate posturing is indicated by flexed wrists and arms and extended legs and feet. Decerebrate posturing is indicated by arm and leg extension.

Other pupillary responses indicate damage to the optic nerve. The afferent pupillary defect is a paradoxical response that is detected with the swinging-light test. As the examiner swings a light from the normal eye to the abnormal eye, the abnormal pupil responds by dilating instead of constricting. This occurs because the light signals transmitted to the Edinger–Westphal nucleus in the midbrain through the injured optic nerve are insufficient to maintain constriction triggered by stimulation of the normal eye.

Bilaterally small pupils suggest a destructive lesion in the pons or the presence of certain drugs. Bilaterally fixed and dilated pupils suggest inadequate cerebral perfusion. This could be related to hypotension or increased ICP. If perfusion is not interrupted for too long, a normal pupillary response returns with adequate flow.

Eye movements are important indicators of brainstem function. Cranial nerves III, IV, and VI are responsible for normal eye movements. Abnormalities of eye movement are useful in localizing the site of brain dysfunction. Abnormal eye movements seen in the brain-injured patient can include nystagmus, dysconjugate eye movements, and ocular palsies. Nystagmus is a persistent rhythmic or jerky movement in one or both eyes. Dysconjugate movements occur when the eyes do not move together in the same direction. Ocular palsies occur when one or more cranial nerves are dysfunctional such that motor paralysis of the eye muscles impairs movements in one or more directions.

Oculovestibular Reflex

The oculovestibular reflex normally detects head movements (via receptors in the semicircular canals) and causes appropriate adjustments of eye position such that an object can remain fixed on the retina even though the head is moving. An impaired oculovestibular reflex implies brainstem dysfunction. Two tests can be performed to evaluate this reflex in the unconscious patient: the doll's eyes test and the cold calorics test.

The oculocephalic or doll's eyes test is performed only in patients in whom a lateral spine radiograph has been obtained to rule out spinal injury. The test is performed by holding open the patient's eyelids and rotating the head from one side to the other (Fig. 44.12). If the brainstem is intact, the eyes will turn in a direction opposite to the direction of head rotation. If the eyes do not move in conjugate fashion or are asymmetric, the response is abnormal and brainstem function is impaired. If the eyes remain fixed at the midline and do not move, the response is said to be absent. An absent response indicates severe brainstem impairment.

The oculovestibular response, or cold calorics, is a similar test of brainstem function using cold water instillation into the ear. Cold against the tympanic membrane causes action potentials from the vestibular apparatus to change and simulates the neuronal response to head rotation. If the brainstem is intact, the normal response will be a tonic deviation of both eyes toward the side that is irrigated. Dysconjugate or asymmetric eye movement is abnormal. If there is no eye movement, the response is absent. Testing of the oculovestibular response is one of the essential examinations performed in patients thought to be brain dead. Patients with depression of brain function attributable to metabolic abnormalities usually retain an intact oculovestibular response. Certain drugs, such as barbiturates and high doses of phenytoin, can severely depress the oculovestibular response. In the absence of drug effect, an absent response to cold calorics is a poor prognostic sign indicating minimal chance of brain recovery.

Corneal Reflex

A simple test of cranial nerve function is the corneal reflex. A wisp of cotton is touched to the cornea of the eye to elicit a blink response. Absence of blink is another indicator of severely impaired brain function.

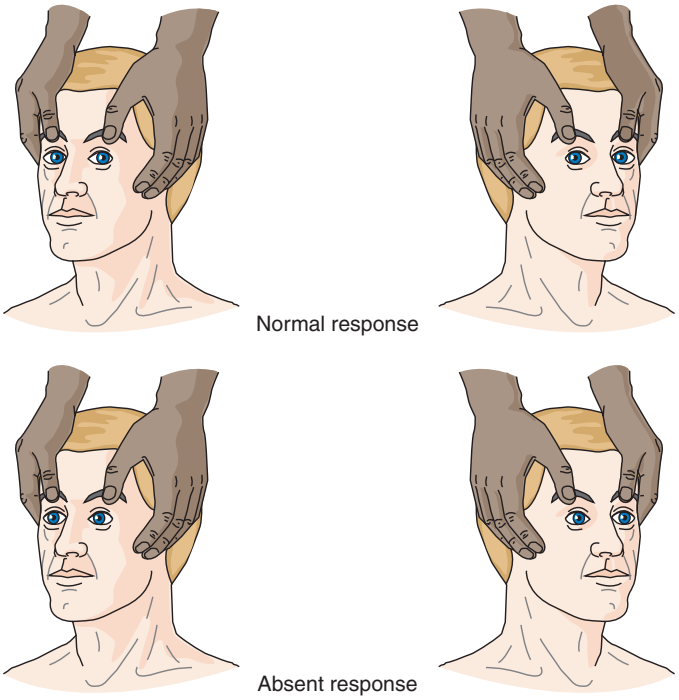


FIG 44.12 Doll's eyes response is indicated by an absent (abnormal) response to the oculocephalic head turning test and indicates brainstem damage. In the normal response, the eyes turn in a direction opposite to head rotation. In the absent response, the eyes stay midline and do not turn when the head is rotated.

KEY POINTS

- A change in LOC often is an early indicator of compromised neurologic status. Normal consciousness is apparent as alertness and orientation to time, person, place, and situation.
- The GCS is used to assess level of coma by scoring alertness (eye-opening response), orientation (verbal response), and motor control (movements). The highest score is given to demonstrating spontaneous eye opening, showing full orientation, and obeying motor commands.
- Pupillary responses indicate the function of the brainstem and cranial nerves II and III. Changes in size, shape, and reactivity of the pupil may be an early indicator of impending brain herniation. Eye movements controlled by cranial nerves III, IV, and VI may be impaired with increased ICP. Nystagmus, dysconjugate gaze, and ocular palsies may be evident.
- An absent doll's eyes response when the subject's head is turned and an abnormal response to activation of the oculovestibular reflex upon instillation of cold water in the ear are very poor prognostic signs.

TRAUMATIC BRAIN INJURY

Traumatic brain injury (TBI) refers to injuries of brain tissues sustained as a consequence of trauma. The term is sometimes used interchangeably with head injury; however, injuries to the cranium do not always result in brain injury. Confusion in terms has led to difficulty in determining actual rates of TBI, but there is no doubt that it is a major public health concern.

EPIDEMIOLOGY

TBI is a leading cause of death and disability in the United States, contributing to the deaths of more than 50,000 people annually. Falls,

sports injuries, firearms, and transportation-related trauma are important causes of TBI. Falls account for the majority of TBI-related hospitalizations for those 0 to 4 years and older than 65 years. Young adults aged 14 to 24 have the highest hospitalization rates due to motor vehicle traffic–related events. Men are three times more likely than women to die as a result of TBI.

It is difficult to quantify the social, medical, and economic impact of TBI. The cost of acute hospital care for a moderately to severely brain-injured patient is substantial, and the long-term care, rehabilitation, and loss of productivity for survivors add significantly to the cost.

Early rescue from the trauma scene and immediate emergency management are important in the effort to reduce morbidity and mortality after TBI. Many TBI fatalities occur within minutes of the traumatic event. Victims who survive until hospitalization require expert monitoring and intervention.

TYPES OF TRAUMATIC BRAIN INJURY

TBI is often characterized according to severity, location of injury, and mechanism of injury. There are different prognoses and management strategies for the different types of TBI.

Severity of TBI usually is based on the patient's GCS score on admission to the hospital or the lowest score in the first 48 hours after admission. A GCS score of 8 or less is defined as a severe injury; moderate injury is defined by a GCS score of 9 to 12; mild injury is associated with a GCS score greater than 12. Injury severity can also be estimated by the degree of brain injury detected on CT examination.

In general, an increased severity score is thought to be associated with a poorer prognosis; however, the predictive value of these tools does have limitations. GCS is used in conjunction with neuroimaging findings, evaluation of coexisting injury/comorbidities, and patient's age to predict outcomes. Injury severity and outcome are somewhat dependent on the physiologic state of the brain before the trauma. Physical factors such as bone thickness, dural stability, brain atrophy, drug effects, previous brain damage, preexisting dementia, and cerebral atherosclerosis have an impact on the outcome of TBI. There is so much individual variation in brain response to injury that making a prognosis statement is often guesswork. 48 to 72 hours after the injury, a better estimate of outcome is possible because the degree of secondary injury will be manifested.

PRIMARY INJURY

Primary injury is the result of the initial trauma on neural tissue. Primary injuries are commonly described as *focal, polar,* or *diffuse.* Although such injuries rarely occur in pure form, they are discussed separately for the sake of simplicity.

Focal injuries (coup) are those that are localized to the site of impact to the skull. The extent of the damage is quite variable. Damage may be superficial or extend deep into the brain matter. Local injury to the brain can result in specific neurologic symptoms, depending on the site. An injury over the motor cortex may result in contralateral weakness of the face and arm, whereas an injury to the frontal lobe can lead to apraxia, impulsive behavior, and poor judgment. However, localized hemorrhages or significant edema may act as space-occupying lesions and result in increased ICP, brain shifting, and herniation. In such cases, symptoms may include a decreased LOC, cranial nerve dysfunction, and contralateral muscle weakness.

Polar injuries (coup contrecoup) occur as a consequence of the brain shifting within the skull and meninges during the course of an acceleration–deceleration movement, resulting in local injury at two opposite poles of the brain. This is commonly the case in motor vehicle accidents in which the head, traveling at the same high speed as the motor vehicle, is abruptly stopped by an obstacle such as the windshield. As a result, the frontal and temporal poles are crushed against the anterior and middle cranial fossae, damaging the tips and inferior surface of the temporal and frontal lobes. Damage may cause bruising or bleeding and, in combination with edema, may result in significant intracerebral mass lesions. Most forces to the head have a lateral rotational component; thus one side of the brain typically is more severely injured than the other. Patients with polar injuries may or may not need significant acute care, depending on the severity of the injury. Polar injuries can, however, be a significant factor in the extent of subsequent cognitive impairment, affecting rehabilitation and long-term recovery.

Diffuse injuries occur when movement of the brain within the cranial cavity causes widespread neuronal damage. The brain is often subject to shifting and rotational forces during injury. The combined force causes stretching and shearing of the axonal white matter, known as *diffuse axonal injury.* Patients with severe diffuse axonal injury frequently are comatose from the time of injury. Coma is a consequence of axonal damage in the cerebral cortex or reticular activating center in the brainstem and can be prolonged. Recovery may be limited to a severely disabled or vegetative state. In addition to falls and motor vehicle accidents, diffuse injury is an unfortunate consequence of vigorous shaking, particularly of babies and the elderly, who have greater mobility of the brain within the skull.

In addition to categorizing primary injuries according to location as focal, polar, or diffuse, they can be differentiated by the mechanism of injury: concussion, contusion, and intracranial hematoma. *Concussion,* otherwise known as mild traumatic brain injury (MTBI), is the most common injury encountered by military personnel and athletes. In this type of injury there is an alteration or LOC (<30 minutes) but no evidence of brain damage on CT.

Symptoms associated with MTBI present immediately and may resolve quickly after the traumatic event. In some instances they may last much longer: headache, nausea, vomiting, dizziness, fatigue, blurred vision. Cognitive and emotional disturbances such as irritability may also be present. The effects of MTBI may be cumulative with additional trauma. Long-term behavioral and cognitive changes are seen in individuals with repeated "mild" injuries. Second impact syndrome may occur in young athletes returning to play/practice before completely recovering from an MTBI. These individuals develop malignant cerebral edema, and this condition is associated with poor neurologic outcomes. Current concussion management guidelines suggest a gradual return to play/activity based on the presence of neurologic symptoms, with the goal of decreasing the risk of second impact syndrome and preventing long-term neurologic consequences (Table 44.2).

Contusion is present when CT or MRI reveals an area of brain tissue damage (necrosis, laceration, bruising). An intracranial hematoma is a localized collection of blood within the cranium resulting from vascular damage.

Intracranial Hematomas

Three types of hematoma can occur after traumatic head injury: epidural (extradural), subdural, and subarachnoid (Fig. 44.13). Hematomas may expand slowly or rapidly, progressively compressing brain structures and increasing ICP. The types of hematoma differ in their clinical presentation and significance. Recognition and prompt management of intracranial hematomas can significantly improve outcome in the patient with TBI.

Epidural Hematoma

Epidural hematomas are collections of blood in the epidural space, which lies between the inner surface of the skull and the dura mater

TABLE 44.2 Graduated Return to Play After Concussion Protocol

Rehabilitation Stage	Functional Exercise	Objective
No activity	Complete physical and cognitive rest	Recovery
Light aerobic exercise	Walking, swimming, or stationary bicycling keeping intensity < 70% maximum heart rate	Increase heart rate
Sport-specific exercise	Skating drills, running drills, passing drills	Add movement
Noncontact training drills	Progression to more complex training drills	Exercise, coordination, and cognitive load
Full contact practice	After medical clearance* participating in normal training activities	Restore confidence and assess functional skills by coaching staff
Return to play	Normal game play	

The athlete is to remain at each level for at least 24 hr and progresses to the next level only if asymptomatic. If neurologic symptoms return with an increase in activity, then they return to the lower level until symptoms again resolve. *Medical clearance should come from a licensed health care provider trained in the management of concussion/head injury.
Adapted from: Harmon KG, Drezner JA, Gammons M, et al. American Medical Society for Sports Medicine position statement: concussion in sport. *Br J Sports Med* 2013;47:15–26. doi:10.1136/bjsports-2012-091941.

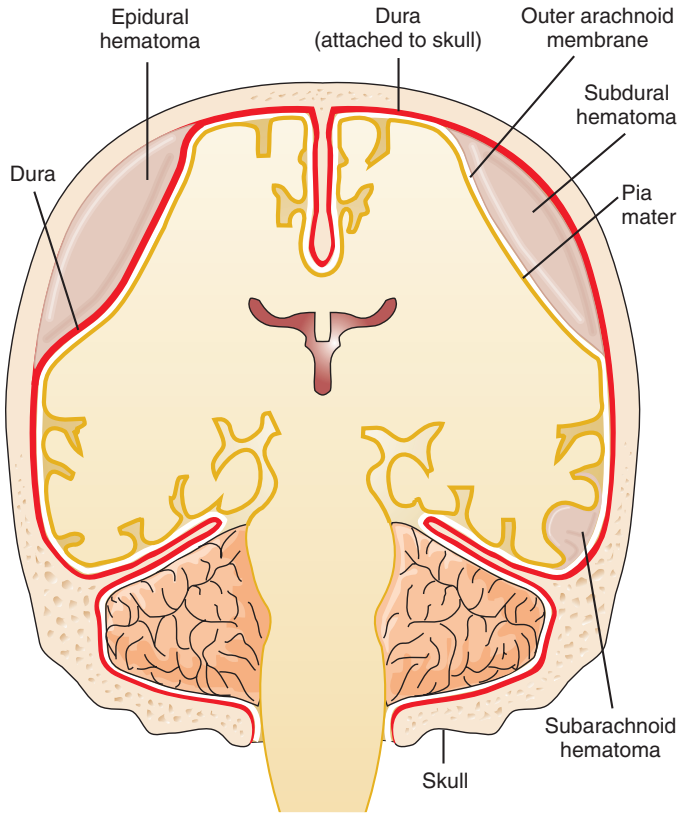


FIG 44.13 Locations of epidural, subdural, and subarachnoid hematomas.

(extradural). Vessels that travel within the dura are susceptible to injury in conjunction with skull fractures. Fracture of the temporal bone commonly disrupts the middle meningeal artery, resulting in an acute epidural hemorrhage. Because the source of bleeding in most epidural hematomas is arterial, the hematoma can expand rapidly, causing acute deterioration of neurologic function. Often, the severity of the primary injury is minor, and the patient may suffer only a brief period of disturbed consciousness followed by a period of normal cognition (lucid interval). Then consciousness rapidly deteriorates as the epidural hematoma expands and compresses brain structures.

Rapid evaluation by CT is recommended to detect the hematoma, and surgical intervention to remove the hematoma is necessary in most cases. Patients with promptly managed epidural hematomas usually have an excellent prognosis because the associated primary injury is minimal.

Subdural Hematoma

Subdural hematomas form in the space between the dura and the outer arachnoid membrane (see Fig. 44.5). The vessels traversing this area are called *bridging veins.* Bridging veins drain venous blood from the surface of the brain, crossing the arachnoid and subdural spaces before emptying into the venous sinuses. Venous sinuses are relatively fixed to the dura and are immobile, whereas the brain, which floats in the CSF, is quite mobile. With a sudden change in head velocity, as occurs with falls and vehicular accidents, the brain moves, the venous sinuses remain stationary, and the bridging veins between the two are stretched and sheared apart.

Because venous blood is under low pressure in the head, the rate of hematoma formation is usually slower than that of an epidural bleed. When subdural bleeds do produce an acute deterioration in neurologic status, the severity of the primary injury is likely to be high. Acute subdural hematomas produce symptoms within 24 hours of injury and have a worse prognosis than epidural or subacute subdural bleeds.

Subacute subdural hematomas can present a diagnostic challenge because the symptoms may be so delayed that the patient does not associate them with a head injury event. Venous bleeding is usually self-limited but may slowly progress and produce symptoms of increased ICP (headache, vomiting, blurred vision) 2 to 10 days after the primary event.

When subdural hematomas are large and sufficiently localized, they are amenable to surgical evacuation. If the subdural hematoma has been present for some time, it enters a chronic stage in which fibroblasts infiltrate the area and granulation tissue forms. Hematomas at this chronic stage are prone to rebleeding from thin-walled capillaries in the new granulation tissue. Chronic subdural hematomas are a common finding at autopsy of elderly individuals. In addition to the increased incidence of falls, the elderly usually have some degree of cerebral atrophy, which makes the brain more mobile within the skull. In the elderly, the likelihood of damage to bridging veins is high even with minor trauma. Alcoholics and those taking anticoagulant medications are also at high risk.

The manifestations of chronic subdural hematoma may be subtle and remain undiagnosed. In the elderly, changes in mentation may be erroneously attributed to dementia. Chronic subdural hematomas are detectable by CT and MRI, and if they are symptomatic they may be managed by surgical removal of the clot and surrounding reactive tissue.

Subarachnoid Hemorrhage

The space between the outer arachnoid membrane and the pia mater is the subarachnoid space. The pia mater is tightly bound to the surface of the brain. The subarachnoid space is filled with CSF. Traumatic subarachnoid bleeding is due to rupture of the bridging veins that pass through the space, in a manner similar to subdural bleeding. Although trauma is an important cause of subarachnoid hemorrhage, it is more commonly associated with rupture of cerebral aneurysms or arteriovenous malformations (AVMs). In that case, bleeding is arterial in origin.

Blood in the CSF manifests with meningeal irritation and a bloody spinal tap. Blood in the subarachnoid space can spread throughout the CSF spaces and may not organize into a confined hematoma. Blood in the CSF produces severe headache in the conscious person and predisposes to secondary vasospasm and ischemia. It also predisposes to clogging of ventricular drainage, which leads to hydrocephalus. Further discussion of manifestations and management of subarachnoid hemorrhage is included in the "Cerebral Aneurysm" section.

SECONDARY INJURY

TBI often initiates mechanisms of secondary injury, resulting in ischemia, increased ICP, and altered vascular regulation. Most of the research on secondary mechanisms of injury has been conducted in the TBI model. Often the damage done by secondary mechanisms far exceeds that of the primary trauma.

In contrast to other types of brain injury, patients with TBI must be carefully evaluated for skull fractures, epidural and subdural hematomas, and injuries to other body systems. Concomitant trauma may complicate the brain injury. For example, uncontrolled hemorrhage can lead to hypovolemia and hypotension, which contribute to brain ischemia. Injuries to the chest can compromise ventilation and produce hypoxemia and hypercarbia, which contribute to cerebral vessel dilation and increased ICP. Attention to other life-threatening injuries may extend the time until radiologic examination of the head can be accomplished, thus delaying surgical management of lesions.

Once the TBI patient's condition is stabilized, many other sources of secondary injury still loom. Brain swelling from both cytotoxic and vasogenic edema may increase for 48 to 72 hours after injury. Ruptured vessels may rebleed or spasm, and CSF drainage can become clogged. Open skull fractures predispose to CNS infections, as do ICP monitors, burr holes, and surgical incisions. Seizures and fever may develop, significantly increasing the brain's metabolic rate and further contributing to brain ischemia. Inflammation and free radical damage continue to cause injury to cells even after ischemia has resolved. Monitoring and managing a patient's course through the myriad perils of secondary injury requires expert knowledge and skill.

TREATMENT

After cardiopulmonary stabilization, the first priority in the TBI patient is radiologic screening of the brain for surgically correctable lesions. Hematomas, depressed skull fractures, and bleeding vessels require prompt surgical intervention. Surgical decompression by performing a craniotomy or craniectomy and insertion of ICP monitors and/or CSF drainage devices may be done. Further therapy is individualized, seeking to maintain ICP, cerebral blood flow, and cerebral oxygen utilization within optimal ranges. Treatment recommendations are controversial, but in patients with acceptable cerebral blood flow, maintenance of normal body temperature, normal $Paco_2$, normal serum glucose levels, and normal intravascular volume is suggested. Nursing measures such as maintaining neutral head alignment with mild elevation (30 degrees) of the head of the bed are standard. In the acute period of injury, elevated ICP can be managed with administration of mannitol (osmotic diuretic), hypertonic saline, sedation, hypothermia, and mild hyperventilation. Repeat radiologic examination may be indicated to determine whether a new surgical lesion has developed. If a new lesion is not present and the patient continues to exhibit high ICP, more aggressive measures may be attempted, including drainage of CSF from ventriculostomy, decompressive craniotomy, moderate hyperventilation, and barbiturate coma; however, despite these interventions, the outcome is likely to be poor. Prevention of seizure activity is also important.

Patients with open head injuries may be treated with prophylactic antibiotics to prevent CNS infection. Sometimes fractures at the base of the skull are not visible on the routine CT scan but allow drainage of CSF into the nasal sinuses. Head-injured patients who have drainage of clear fluid from the ears or nose should be evaluated for basilar skull fracture. CSF drainage differs from normal nasal mucus because it has a high glucose content and tends to separate into layers (halo) on tissue paper. Other findings with basilar skull fracture are bilateral periorbital hematomas (black eyes, "raccoon sign") and bruising under the ear (Battle sign). The presence of basilar skull fracture increases the risk of CNS infection.

Although morbidity and mortality after acute TBI remain high, significant advances in prehospital, hospital, and rehabilitative care continue to occur. The care received during the first hour of injury can dramatically affect outcome. Efforts to improve prehospital management are likely to have the greatest impact on patient outcomes.

KEY POINTS

- Most head injuries are incurred in motor vehicle accidents, falls, and sports accidents. Young males ages 14 to 24 years are the most common victims. The seriousness of head injury can be classified according to GCS scores as mild (>12), moderate (9 to 12), or severe (<8).
- Injury that is directly due to the initial impact is called *primary injury*. Primary injuries are classified as focal, polar, or diffuse. Focal injuries are localized to the site of skull impact. Polar injuries are due to acceleration–deceleration movement of the brain within the skull, resulting in double injury. Diffuse injury is due to movement of the brain within the skull, resulting in widespread axonal injury.
- Disruption of the vasculature can result in intracranial hemorrhage. Epidural hematomas are associated with skull fracture and progress rapidly because they are arterial in origin. Subdural hematomas are associated with shearing of bridging veins and may develop slowly. Traumatic subarachnoid hemorrhage is also due to trauma to bridging veins.
- Secondary injury is a consequence of the body's response to the primary injury. Mechanisms are similar to those described for nontraumatic brain injury. In addition, concomitant injuries and cardiopulmonary impairment may contribute to the severity of secondary injury.
- The management of brain injury is directed primarily to detecting and managing surgical lesions and reducing brain damage from secondary injury. Normovolemia, normothermia (or mild hypothermia), normal glucose level, and normal Pao_2 and $Paco_2$ values are recommended for most patients. In those with high ICP, diuretics, hyperventilation, and drug-induced coma may be tried. Open head injuries constitute a risk for CNS infections, and prophylactic antibiotics may be used.

CEREBROVASCULAR DISEASE AND STROKE

Cerebrovascular diseases cause abnormalities of cerebral perfusion, including transient ischemic attack (TIA), ischemic stroke, and hemorrhagic stroke. *Stroke* is a term applied to cerebrovascular events that result in a localized area of CNS infarction and was previously termed

cerebrovascular accident (CVA). The term *brain attack* has been popularized to educate the public about the importance of seeking care early, as is recommended for heart attack.

The symptoms of stroke usually are sudden in onset and may include the following: (1) numbness or weakness of the face, arm, or leg, especially affecting only one side of the body; (2) confusion, trouble in speaking or in understanding others; (3) visual disturbances in one or both eyes; (4) dizziness, loss of balance, and difficulty with walking; and (5) severe headache. Persons experiencing any of these symptoms, even temporarily, should seek medical care immediately.

EPIDEMIOLOGY

Approximately 750,000 new strokes occur each year in the United States, and 150,000 of those individuals die. It is the fifth leading cause of death and the leading cause of disability in the United States. The incidence of stroke increases with age, with two thirds of all strokes occurring after age 65. It is more common in men than in women and in African Americans than in Caucasians. The incidence of strokes has declined in recent decades due to improvements in the treatment of hypertension and efforts to decrease smoking and manage other modifiable risk factors.

Risk factors for stroke are similar to those for other atherosclerotic vascular diseases and include hypertension, diabetes, hyperlipidemia, cigarette smoking, advancing age, and family history. A previous stroke significantly increases the risk for suffering a subsequent stroke. Cardiac disease complicated by atrial fibrillation is an important risk factor for embolic types of strokes. Strokes can be categorized according to cause as ischemic and hemorrhagic strokes. Ischemic strokes are by far the most common and include thrombotic and embolic types.

ISCHEMIC STROKE

Ischemic strokes result from sudden occlusion of a cerebral artery secondary to thrombus formation or embolization. Thrombotic and embolic strokes are grouped together because the clinical presentation and treatment are similar. However, etiologic risk factors and preventive measures are different and are discussed separately.

Thrombotic strokes are associated with atherosclerosis and hypercoagulable states. Risk reduction strategies for thrombotic stroke are those aimed at reducing atherosclerosis and platelet aggregation. Significant atherosclerotic plaques in the carotid arteries are sometimes evident as carotid bruits. Assessment for the presence of carotid bruits in all individuals older than 50 years may help identify persons at risk so that prevention strategies can be initiated early.

Emboli usually are from a cardiac source, although disruptions in carotid artery plaques may lead to downstream embolization. Cardiac sources include thrombi formed in the cardiac chambers (mural thrombi) and thrombi or vegetations on valve leaflets. Because atrial fibrillation allows stagnation of blood in the left atrium, it is associated with a high risk of mural thrombi, which can dislodge and travel to the cerebral circulation. Patients with chronic atrial fibrillation commonly receive anticoagulant medications to prevent this occurrence.

Sudden blockage of a cerebral artery by a thrombus or embolus produces acute ischemia in the territory served by the artery. Insufficient blood flow to brain tissue results in oxygen deprivation and rapid cerebral deterioration. Neurologic deficits become evident after just 1 minute of insufficient oxygen. If the ischemia continues for several minutes, irreversible cellular damage can occur. With further progression, the local area becomes infarcted and necrotic. Surrounding the infarct is a much larger area of ischemic but viable cells, called the *penumbra*. The penumbra receives some partial or collateral flow and may recover if the ischemia is mild or perfusion is restored in a timely manner. Salvaging the penumbra is the aim of early thrombolytic therapy; however, treatment must be instituted within 3 hours of stroke onset to be maximally effective.

Transient ischemic attacks (TIAs) are episodes of stroke signs/symptoms with a duration of less than 24 hours and usually less than 1 hour. TIAs were previously assumed to be due to temporary interruptions of blood flow with thrombolysis of clot before permanent tissue damage had occurred. However, with improvements in neuroimaging, many individuals who have brief, resolving neurologic symptoms were indeed found to have infarction of brain tissue. The revised definition of TIA is not based on time of symptoms but on the absence of evidence of infarct. If damage is seen on imaging, the episode is classified as a stroke regardless of the duration of neurologic symptoms. TIAs are important warning signs of thrombotic disease and carry a significant risk for subsequent stroke. Approximately 30% of patients with a stroke have had a prior TIA.

Patients who present with TIAs should undergo evaluation to determine the origin of their symptoms. Unless contraindicated, these patients are started on daily aspirin therapy to prevent thrombus formation. Carotid endarterectomy or angioplasty may prevent stroke in a subset of patients experiencing TIAs who have carotid artery plaques occluding more than 70% of the arterial lumen.

Manifestations of ischemic stroke are related to the cerebral vasculature involved (Fig. 44.14) and the area of brain tissue the vessel supplies. The middle cerebral artery is most commonly occluded, resulting in damage to the lateral hemisphere. Contralateral hemiplegia, hemisensory loss, and contralateral visual field blindness are usual. If the dominant hemisphere is affected, global aphasia will occur. Occlusions of smaller branches of the middle cerebral artery produce more limited neurologic findings. Occlusions of the other cerebral arteries have different neurologic manifestations depending on the brain area they normally perfuse (Table 44.3).

Occlusion of the small penetrating arterioles can produce small lesions called *lacunar infarcts.* The basal ganglia, pons, cerebellum, and internal capsule are common sites of lacunar infarcts. These lesions are sometimes not observable on CT scan. The prognosis for recovery from a lacunar infarct is usually good, and neurologic manifestations are more circumscribed, often affecting purely motor or sensory functions.

HEMORRHAGIC STROKE

Intracerebral hemorrhage is a hemorrhage within the brain parenchyma and usually occurs in the context of severe and often long-standing hypertension. It carries a 30% mortality rate. In contrast, subarachnoid hemorrhage occurs under the arachnoid membrane and above the pia mater. Two common structural abnormalities that can cause subarachnoid hemorrhage (cerebral aneurysms and AVMs) are discussed in the "Cerebral Aneurysm and Arteriovenous Malformation" section. Most intracerebral hemorrhagic strokes occur in the basal ganglia or thalamus. If the hemorrhage is large, it may significantly increase ICP, which can lead to herniation and death. The prognosis for hemorrhagic stroke depends on the patient's age, the location and size of the hemorrhage, and the speed at which the hemorrhage produces brain distortion and shift. The degree of secondary injury and associated morbidity and mortality are significantly higher for hemorrhagic stroke than for ischemic stroke.

TREATMENT

Initially, after the patient's airway, respiratory status, and cardiovascular function are assured, an assessment of stroke severity and associated

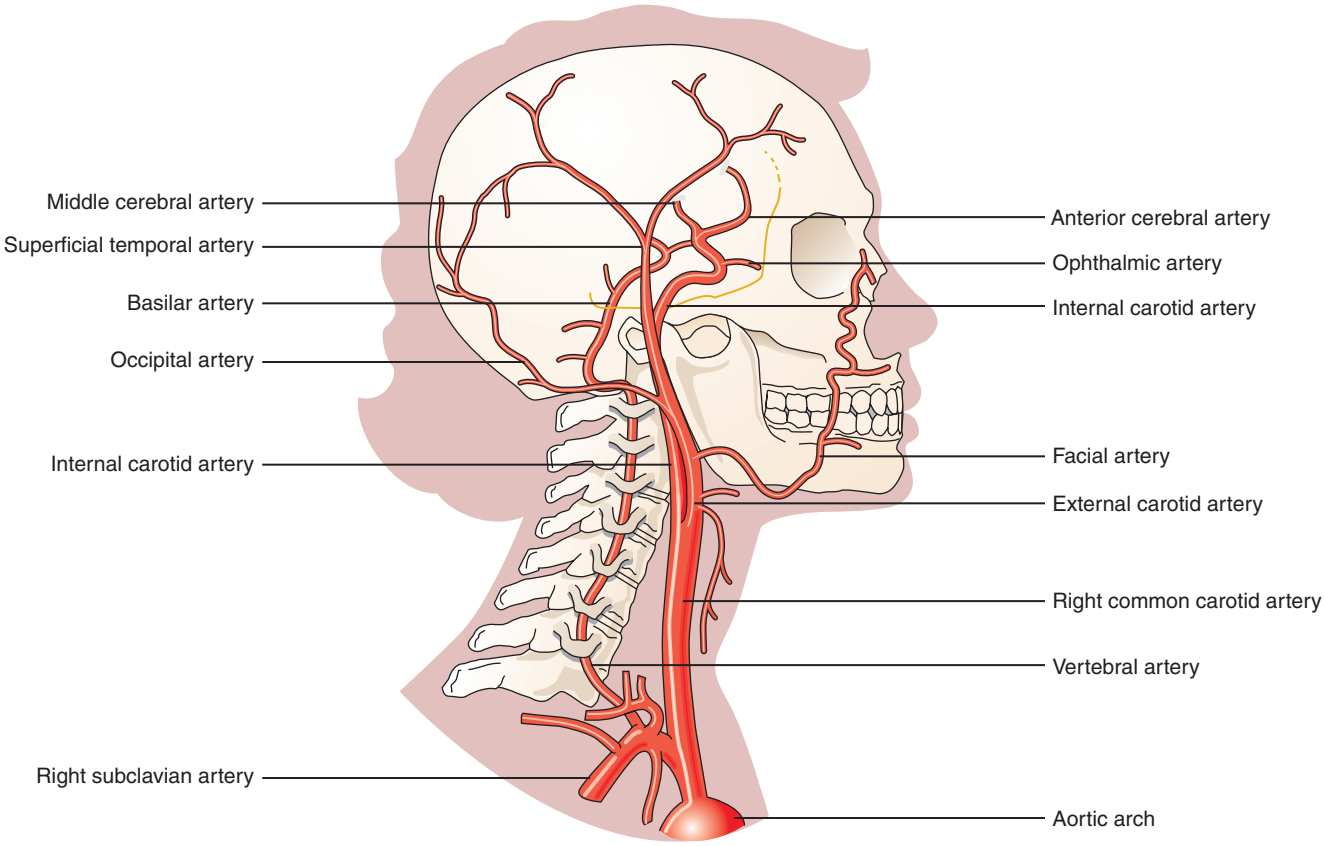


FIG 44.14 Cerebral vasculature showing fields of perfusion.

TABLE 44.3 Manifestations of Ischemic Stroke According to Location of Arterial Blockage

Cerebral Artery	Territory of Perfusion	Clinical Manifestations
Anterior cerebral	Medial aspect of frontal lobes	Contralateral hemiparesis; contralateral sensory loss; impaired cognition and decision-making; aphasia (left-sided stroke); incontinence
Middle cerebral	Most of lateral cerebral hemisphere, internal capsule, and basal ganglia	Contralateral hemiplegia; contralateral sensory loss; aphasia (left-sided stroke); homonymous hemianopsia; altered consciousness; neglect syndrome
Posterior cerebral	Occipital lobe and medial aspect of temporal lobe	Visual defects including homonymous central blindness and color blindness; memory impairment
Basilar and vertebral	Thalamus, cerebellum, and brainstem	Sensory loss; mild hemiparesis; disturbances of gait, speech, swallowing, and vision

neurologic deficits is made. An initial noncontrast CT scan or diffusion weighted brain MRI is used to determine the type of stroke and treatment, because treatment pathways differ between ischemic and hemorrhagic stroke. Patients who have experienced a hemorrhagic stroke secondary to hypertensive disease often have extremely high blood pressure. Returning their blood pressure into normotensive ranges could result in ischemia. In these circumstances it is best to keep the patient mildly hypertensive with the goal of normalizing the blood pressure once the patient is medically stable. Patients who have experienced a hemorrhagic stroke are at risk for increased intracranial pressure and are assessed and treated similar to patients with a TBI. Thus the remaining discussion will focus on treatment of ischemic stroke.

The goals of therapy for acute ischemic stroke are to minimize infarct size and preserve neurologic function. Aspirin may be administered after negative findings of hemorrhagic stroke on CT scan. The administration of 325 mg of aspirin immediately affects platelet aggregation and may help inhibit thrombus size. Thrombolytic therapy is most effective in limiting infarct size if it is initiated early.

It is critical to prevent further cerebral hypoxia or ischemia after ischemic stroke. Thus volume depletion, hemoconcentration, hypotension, and arterial obstruction must be avoided. As with the hemorrhagic stroke patient, careful blood pressure management is critical. Overhydration can result in cerebral edema in the ischemic area of the brain and raise ICP. Patients should have their ability to swallow evaluated *before* they take any food or liquids orally. Injury to cranial nerves V, VII, IX, X, or XII can place a patient at risk for respiratory aspiration and further compromise the individual's health and potential for recovery.

Anticoagulation therapy may be used in ischemic stroke, especially if the event is progressive. A stroke is termed *progressive* if an initial focal deficit worsened or fluctuated before hospital admission or deteriorated on serial examinations after admission. Patients who receive thrombolytic therapy should not receive anticoagulation therapy because the risk of bleeding is high. Throughout the course of therapy, it is essential to monitor clotting parameters and recognize the potential for hemorrhage into the ischemic area. Even in the absence of thrombolytic or anticoagulant therapy, a significant number of ischemic strokes

convert to hemorrhagic lesions. A sudden change in neurologic function should prompt reevaluation by CT. Management of the stroke patient also must include efforts to prevent stroke recurrence. Evaluation and management of risk factors are essential parts of prevention. The survivor of a stroke is at high risk for a subsequent stroke if precipitating factors are still present. Patients who have experienced thrombotic strokes are also at significant risk for other vascular events, such as myocardial infarctions. Secondary prevention varies according to the cause of the stroke. For hemorrhagic strokes, careful monitoring and control of blood pressure is essential. For ischemic strokes of embolic origin, identifying and removing the source of emboli is crucial. Usually the source is the heart, and therapy includes control of dysrhythmias, implementation of anticoagulation therapy, and administration of antiplatelet drugs such as aspirin. The most common cause of embolic stroke is atrial fibrillation. Conversion of this dysrhythmia to normal sinus rhythm can sometimes be accomplished with antidysrhythmic agents or electrical cardioversion. Measures to improve left ventricular function may help reduce atrial pressure and correct atrial fibrillation. Patients with long-standing atrial fibrillation require anticoagulation therapy. When abnormal valves are suspected to be the source of emboli, evaluation for surgical replacement may be indicated.

Secondary prevention for thrombotic stroke includes lifestyle modifications to address modifiable risk factors, including smoking cessation and lowering of serum lipid levels. In addition, the long-term daily use of aspirin or other antiplatelet agents (e.g., clopidogrel bisulfate) has been recommended. Some patients may benefit from surgical removal of carotid artery plaque by endarterectomy or angioplasty. Placement of rigid tubes, called *stents,* in the area of plaque removal may be helpful in preventing reocclusion.

STROKE SEQUELAE

Recovery after stroke depends on the size and location of the cerebral infarct, comorbid conditions, and rehabilitative efforts. Stroke rehabilitation begins during the acute hospitalization phase and continues after the patient has returned to the community. Many patients have residual or permanent deficits in motor, sensory, language, and cognitive functions that necessitate intensive strategies to maximize the likelihood of return to a productive life.

Motor and Sensory Deficits

Motor impairment from a stroke is initially characterized by flaccidity, which is a decrease in or absence of muscle tone in the affected extremities. Most commonly, motor paralysis is contralateral to the side of the brain in which the stroke occurs. Thus a stroke on the right side of the brain results in left-sided body paralysis, whereas left brain strokes result in right-sided body paralysis. Foot drop, outward rotation of the leg, and dependent edema are common features in the lower extremity. In the upper extremity, the arm may separate from the shoulder if not supported. Muscles in the affected limbs tend to atrophy from lack of tone and use. Many of the complications can be limited with therapeutic interventions, including performing frequent range-of-motion exercises, elevating edematous limbs, wearing elastic stockings, and maintaining body alignment.

Starting at about 6 weeks after the stroke, recovery of motor function is evident by the onset of spasticity. Spasticity is the resistance of muscle groups to passive stretch with an increase in tone. Increased flexor tone is usually seen in the upper extremities and increased extensor tone in the lower extremities. Performing passive or active range-of-motion exercises and maintaining proper body position are critical to maintenance of function, because uncontrolled spasticity can result in contractures of the limbs, including adduction of the shoulder, pronation of the forearm, and flexion of the fingers. In the lower extremity, the patient may have problems with hip and knee extension. If spasticity in a paretic extremity is not evident within 3 months, motor function is not likely to return to the affected limb.

Sensory impairment occurs in the same locations as the motor paralysis. A lack of sensory information from the paralyzed side contributes to the phenomenon of *neglect* (also called *hemiattention*). The patient seems unaware that the affected body parts belong to him or her. Loss of the visual field on the paralyzed side also contributes to neglect. Contralateral field blindness is called *homonymous hemianopsia* (also called *hemianopia*) because the same side of the retina in each eye is blinded (Fig. 44.15). Patients with neglect may crush, burn, or otherwise injure the neglected body parts without realizing it. Poor hygiene of affected extremities may be apparent. Neuropsychological studies have shown that objects in the field of neglect are usually ignored. For example, when asked to draw the numbers on the face of a clock, all 12 are drawn on one side. Self-portraits may be conspicuous for the distortion or omission of structures on the neglected side. Neglect is associated with a high risk for falls and other injuries.

Language Deficits

Aphasia is an integrative language disorder that occurs with brain damage to the dominant cerebral hemisphere (usually left) and involves all language modalities. Characteristics of aphasia include a reduced vocabulary, reduced verbal attention span, and reduced ability to use learned linguistic rules. Aphasia is associated with lesions in the primary language centers (Broca and/or Wernicke areas) as well as in adjacent cortical areas. Aphasia is categorized according to the location of the lesion and the linguistic deficit. The following is a brief description of those categories.

Broca aphasia, also known as *verbal motor* or *expressive aphasia,* results from a lesion in the third frontal convolution of the left hemisphere in most persons. Patients speak with poorly articulated and sparse vocabulary and in the simplest grammatical constructions.

Wernicke aphasia, also known as *sensory, acoustic,* or *receptive aphasia,* is characterized by impaired auditory comprehension and speech that is fluent but empty of content. This form of aphasia is caused by lesions in the posterior portion of the first temporal gyrus of the left hemisphere. Speech is frequently circumlocutory or tangential and contains paraphasic errors and jargon. Word finding and naming difficulties are a prominent feature of this disorder. Patients with Wernicke aphasia are unable to monitor their own language production and cannot comprehend or monitor the language production of others.

Anomic aphasia results from lesions in the parietotemporal area in proximity to the angular gyrus. This is a fluent aphasia with intact grammatical structure. Patients have greater word-finding difficulties than those with Wernicke aphasia but do not make paraphasic errors and have intact comprehension. However, their speech is typically constructed of simple words.

Conduction or central aphasia is associated with increased paraphasic errors and a reduced ability to repeat words. It is associated with a lesion in the arcuate fasciculus in the left hemisphere. Patients are well aware that they are making language errors. However, the more they struggle to find the correct words, the more likely they are to repeat paraphasic errors.

Cognitive Deficits

Patients experience impairments of cognition attributable to diffuse cortical or subcortical injuries that affect the ability to be alert, to concentrate or attend to stimuli, to remember, and to reason. Cognitive impairment varies according to the area of brain affected and the severity of the injury. Injuries that disturb an individual's ability to maintain

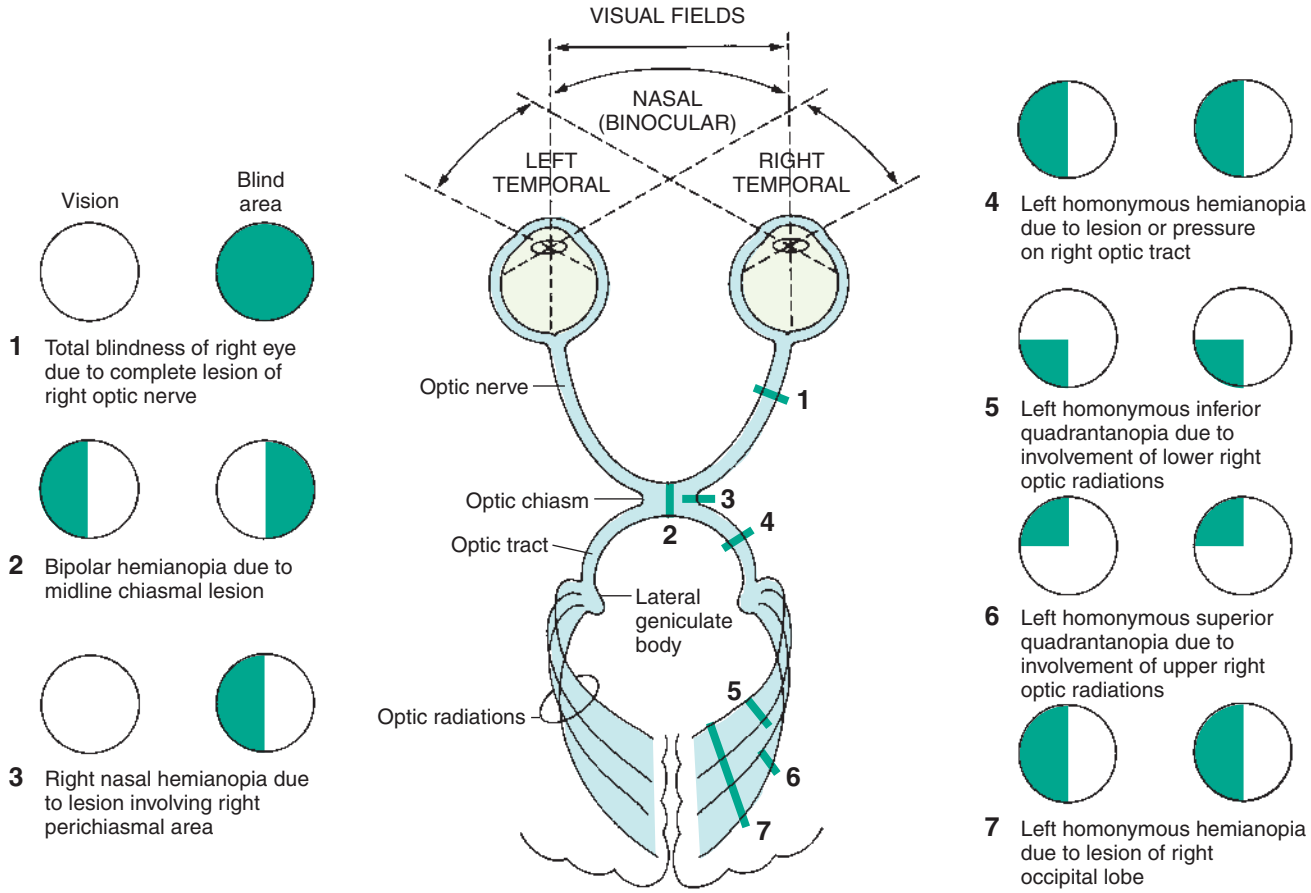


FIG 44.15 Homonymous hemianopsia (also called hemianopia). A right-sided brain stroke may cause lesions that disturb visual fibers and result in blindness in the left visual field. The optic pathway from the other side remains intact. (From Black JM, Hawks JH: *Medical-surgical nursing: clinical management for positive outcomes,* ed 8, Philadelphia, 2008, Saunders, p 1850.)

an alert status are the most severe. Increasing cognitive skill is necessary for the function of memory and the ability to learn and associate, to discriminate, to separate, and to categorize various stimuli. The highest levels of cognitive function include analysis, synthesis, and reasoning abilities.

Cognitive impairment is commonly evidenced as language deficit, impaired spatial relationship skills, short-term memory impairment, and poor judgment. Patients who do not retain the ability to learn are unlikely to benefit from rehabilitative services.

KEY POINTS

- Stroke is the sudden onset of neurologic dysfunction attributable to cerebrovascular disease. The most common cause of stroke is thrombosis, followed by embolization and intracranial hemorrhage.
- Thrombi form at atherosclerotic plaques, causing sudden occlusion of an already narrowed vessel. Emboli are usually a consequence of clots from within the heart chambers caused by disease or dysrhythmia. Hemorrhagic stroke is usually associated with uncontrolled hypertension.
- Stroke symptoms depend on the area of brain affected, which in turn depends on the vessel occluded: internal carotid, anterior cerebral, middle cerebral, or posterior cerebral artery. Common manifestations include contralateral motor and sensory loss, aphasia, and contralateral visual field loss.
- Treatment is aimed at limiting the size of the brain infarction, supporting bodily functions, and initiating aggressive rehabilitation strategies. Acute therapy with thrombolytic agents may limit infarct size in patients with ischemic stroke.
- Stroke is associated with long-term deficits in motor, sensory, language, and cognitive abilities. Initially, affected muscles are flaccid, with spasticity occurring after about 6 weeks. Prevention of contractures is a major concern. Aphasia may be described as expressive or receptive. Most individuals with aphasia have impaired integrative ability involving all language modalities. Concentration, memory, and reasoning may be impaired.

CREBRAL ANEURYSM AND ARTERIOVENOUS MALFORMATION

Structural abnormalities of the cerebral arteries predispose individuals to intracerebral bleeding and hemorrhagic stroke. Cerebral aneurysms and AVMs are the two most common causes of subarachnoid hemorrhage. Early recognition and surgical management of these conditions are necessary to prevent significant mortality and morbidity associated with rupture.

CEREBRAL ANEURYSM

Etiology

An aneurysm is a lesion of an artery that results in dilation and ballooning of a segment of the vessel. Aneurysm rupture occurs in about 10 in

every 10,000, or about 30,000 individuals per year in the United States. The prevalence is higher in women than in men, and rupture most often occurs between the ages of 30 and 60 years. However, unruptured aneurysm is detected on autopsy in approximately 2% of the population. Thus, other factors are likely to be important in precipitating aneurysm rupture. High blood pressure, acute alcohol intoxication, and recreational drug use (especially cocaine) have been implicated. The annual risk of rupture in persons with aneurysms is 1% to 2%. Larger aneurysms and those located in the posterior circulation are more prone to rupture.

Pathogenesis and Manifestations

Although the exact pathogenesis is not understood, saccular aneurysms are believed to result from congenital defects of the medial layer of the artery. This structural weakness permits gradual ballooning at the site as a consequence of arterial pressure effects over years. A common location for saccular aneurysms is arterial bifurcations, where turbulent blood flow might have a greater impact on a weakened vessel wall. Ninety-five percent of cerebral aneurysms are located in the circle of Willis; 10% to 20% of affected individuals have more than one aneurysm.

Saccular aneurysms (berry aneurysms) are round and are the most common (Fig. 44.16). The aneurysmal sac is composed of thickened intima and adventitia layers, with the medial layer having abruptly ended at the sac edge (Fig. 44.17). Rupture of the aneurysm generally occurs from the dome of the sac or at the edge of the atheromatous plaque.

The development of aneurysms is a multifactorial interaction of acquired factors, such as atherosclerosis or hypertension and congenital predisposition, and aneurysm development is associated with various vascular abnormalities. Multiple conditions have been associated with cerebral aneurysms, including autosomal-dominant inherited polycystic kidney disease, Marfan syndrome, Ehler–Danlos syndrome, lupus, and bacterial endocarditis, among others.

Warning leaks may occur before an aneurysm ruptures and often produce severe headache, which is typically described by the patient as "the worst headache I have ever had." The frequency of true "warning leaks" is unknown, but rapid evaluation of patients presenting with new onset of severe generalized headache is warranted. Patients may also complain of photophobia (visual sensitivity to light), nausea/vomiting, and stiff neck. A stiff and painful neck results from meningismus caused by the irritating properties of blood in the CSF. After rupture, the onset of symptoms is very rapid. Sudden injection of blood into the subarachnoid space raises ICP and distorts intracranial structures. Secondary cerebral vasospasm, a pathologic narrowing of the major vessels around the area of rupture, typically occurs from day 4 to day 14. This process significantly reduces cerebral blood flow and results in increased cerebral ischemia and possibly infarction. Vasospasm is due to the presence of blood in the CSF. The next most serious consequence of the initial rupture is rebleeding. The risk for rebleeding is highest in the first 14 days. Patients are also at risk for developing hydrocephalus from clogging and obstruction of CSF flow through the ventricular system.

Diagnostic procedures for detecting a ruptured aneurysm include CT or MRI to confirm a subarachnoid hemorrhage. If the scan is negative but suspicion of subarachnoid hemorrhage is high, a lumbar puncture for CSF analysis can be done. Blood in the CSF is indicative of subarachnoid hemorrhage. A cerebral angiogram is obtained to demonstrate the location of aneurysms in preparation for surgical management.

Treatment

The primary treatments for aneurysms are surgical stabilization by clipping or placement of endovascular coils for embolization. Prognosis is favorable if the aneurysm is detected and managed before significant rupture occurs. In most cases, the aneurysm is not diagnosed until after subarachnoid or intracerebral hemorrhage has occurred, so mortality is higher. Early surgery in stable patients with subarachnoid hemorrhage is associated with a lower overall mortality. Aneurysm clipping is accomplished by placement of a permanent vascular clip at the neck of the aneurysm. Coil devices may be inserted under radiographic guidance to thrombose or sclerose the area.

In patients experiencing subarachnoid hemorrhage as a consequence of a ruptured aneurysm, the complications of cerebral vasospasm and hydrocephalus must be monitored and managed. Vasospasm can be managed by keeping blood volume and blood pressure at normal to high levels. Calcium channel blockers may be used to reduce vasospasm. In addition to hemodynamic monitoring, careful and frequent neurologic assessments are essential to monitor stability and indicate the first signs of deterioration so that rapid intervention can be undertaken.

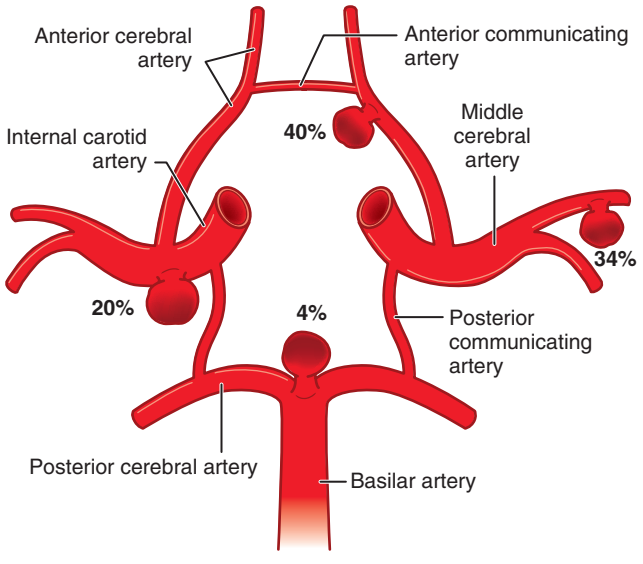


FIG 44.16 Saccular (berry) aneurysms are most commonly found in the circle of Willis, particularly at arterial bifurcations.

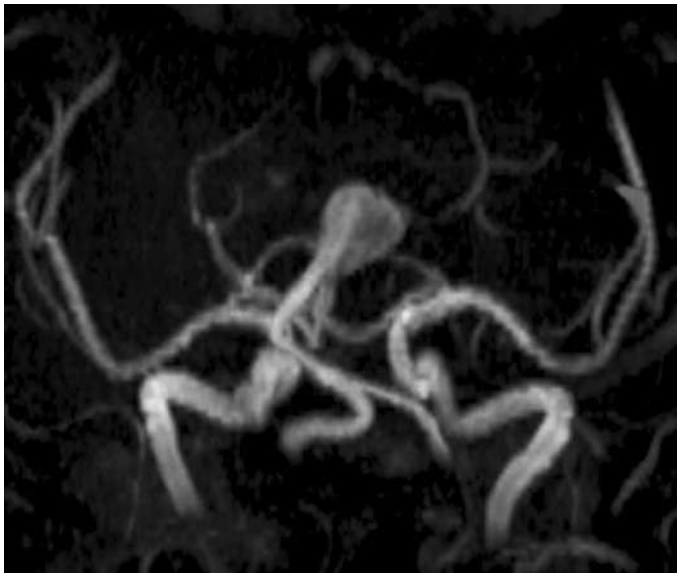


FIG 44.17 Arteriogram of the circle of Willis showing a large saccular aneurysm arising from the base of the basilar artery. (From Dutton JJ: *Radiology of the orbit and visual pathways,* Philadelphia, 2010, Elsevier.)

ARTERIOVENOUS MALFORMATION

Etiology

AVMs are the second most common cause of spontaneous subarachnoid hemorrhage, but can also cause intracerebral hemorrhage. The majority of AVMs are diagnosed in patients before the age of 40, with approximately 50% presenting with hemorrhage. AVMs are vascular lesions thought to be congenital; however, they are rarely diagnosed in the pediatric population. AVMs are more frequent in men and rarely follow a familial pattern.

The risk of clinically recognizable hemorrhage from an AVM is 2% to 4% per year. The risk of rebleeding is as high as 17%, especially in the first year.

Pathogenesis and Manifestations

In the normal vascular system, the capillaries are situated between the arterioles and the venules. In an AVM, the capillary system fails to develop appropriately, and arterial blood is shunted directly into the venous system. Exposure of the high-capacitance venous system to the high pressure of the arteries causes the vessels to progressively enlarge, as do the arteries and veins that feed and drain the lesion. The blood vessels that comprise the AVM proliferate and enlarge over time. AVMs were once considered purely a congenital abnormality; however, numerous endothelial growth factors and vascular substances have been identified that continue to affect the structure and function of these abnormal vessels and contribute to the risk of rupture and hemorrhage.

Because of their abnormal structure and the high vascular pressure, AVMs are vulnerable to hemorrhage. Hemorrhage is the initial manifestation in 50% of cases of AVM; 30% are manifested by seizures and the rest by varying degrees of vascular steal syndrome. The abnormal shunting of blood into the AVM, called *vascular steal syndrome,* causes progressive neurologic dysfunction as a result of ischemia in normal tissue.

Treatment

Once the AVM and vascular bed are evaluated, the AVM may be surgically removed. Alternatively, gamma knife or stereotactic radiosurgery can be used to deliver precisely aligned beams of gamma radiation to shrink the abnormal vascular tissue. For deep or very large AVMs, other approaches (e.g., irradiation and glue embolization) may be used. Supportive therapy for AVMs that rupture and cause subarachnoid hemorrhage is similar to that described for ruptured cerebral aneurysms.

KEY POINTS

- Cerebral aneurysms and AVMs are the two most common causes of subarachnoid hemorrhage. Aneurysm is most common and has a higher mortality rate.
- Blood in the subarachnoid space is associated with headache, stiff neck, and secondary cerebral vasospasm. Vasospasm, which leads to cerebral ischemia, is an important cause of morbidity and mortality.
- Aneurysms are congenital weaknesses in the arterial walls that lead to dilation and ballooning of the wall. Treatment includes surgical stabilization by clip ligation and aggressive management of secondary vasospasm.
- AVMs are congenital malformations in which arterial blood is shunted directly into the venous system, causing high venous pressure. The AVM enlarges and may compress adjacent structures or rupture. Surgical management, radiation, or glue embolization to occlude the AVM may be done to prevent bleeding.

CENTRAL NERVOUS SYSTEM INFECTIONS

Infections of the CNS include meningitis, encephalitis, and abscesses. Organisms gain access to the CNS by various portals of entry. These include via the bloodstream, by direct extension from a primary site (e.g., sinuses), by extension along peripheral and cranial nerves, and through maternal-fetal exchange. Factors contributing to infections include such conditions as immunocompromised status, debilitation, poor nutrition, radiation therapy, steroid therapy, and contact with vectors. Meningitis and cerebral abscess are most commonly associated with bacterial infections, whereas encephalitis is usually viral.

MENINGITIS

Meningitis is the most common sequela to microbial invasion of the CNS. Most frequently, meningitis is bacterial in origin, but it can also be viral or fungal. Persons with AIDS have an increased susceptibility to infection and have an increased prevalence of meningitis of viral, fungal, or parasitic origin.

Etiology

The bacterium most frequently involved in causing meningitis in adults is *Streptococcus pneumoniae. Haemophilus influenzae* type B (HIB) incidence has fallen dramatically since the introduction of the HIB vaccines. The bacteria that cause meningitis usually reach the CNS by way of the bloodstream or by extension from cranial structures, such as the paranasal sinuses or ears. Some of the organisms responsible for causing meningitis may be normal inhabitants of the nasopharynx. Pathogens can also gain access to the CNS through breaks in the barrier system, as occur with penetrating head wounds or skull fractures or after neurosurgery in which the dura is penetrated. The mortality rate is 3% to 7% for meningitis caused by *H. influenzae, Neisseria meningitides,* or group B Streptococcus. It is 20% for *S. pneumonia.* Moderate-to-severe neurologic deficits can occur in up to 25% of survivors.

Pathogenesis and Clinical Manifestations

Bacterial meningitis is a pyogenic infection that invades the leptomeninges and the subarachnoid space. Because of its involvement in the subarachnoid space, the infection travels readily around the brain and spinal cord. The accumulation of inflammatory exudate frequently results in obstructive hydrocephalus and exudative invasion into the sheaths of the blood vessels and spinal and cranial nerves.

The combination of headache, fever, stiff neck, and signs of cerebral dysfunction (confusion, delirium) is the classic presentation of meningitis. Deterioration in LOC is progressive and often rapid. Patients who deteriorate rapidly often demonstrate dramatic tachypnea. About one third of patients experience seizures. Cranial nerve involvement is also common and is most often seen as ocular palsies, facial weakness, deafness, and vertigo.

The diagnosis of meningitis is usually made by lumbar puncture. Typical CSF findings are shown in Table 44.4. Gram stain of the CSF will reveal the causative organism in most patients. In addition to the causative organism, classic CSF findings include white blood cell (WBC) counts between 1000 and 10,000/mm^3 with a predominance of neutrophils. The CSF glucose level is reduced and often extremely low, and patients with bacterial or fungal meningitis have increased protein levels.

Treatment

Recovery from bacterial meningitis depends largely on how quickly effective treatment is started. Treatment includes general supportive care, intravenous antibacterial drug therapy targeting the specific

TABLE 44.4 Typical Cerebrospinal Fluid Findings in Bacterial Meningitis

CSF Variable	Typical Findings	Normal
White blood cell count	1000–5000 cells/mm^3 (up to 10,000) (high)	<5 cells/mm^3
Neutrophils	≥90% (high)	60%–80%
Protein	80–500 mg/dL (high)	30 mg/dL
Glucose	≤40 mg/dL (low)	50–80 mg/dL
Gram stain	Positive (60%–90% of cases)	Negative
Culture	Positive (70%–85% of cases)	Negative
CSF opening pressure	>20 cm H_2O (high)	<15 cm H_2O
Latex agglutination	May be + in *S. pneumoniae, N. meningitidis, H. influenza, E. coli* or group B streptococcus	Negative
PCR	Detects bacterial DNA	Negative

CSF, Cerebrospinal fluid; *PCR,* polymerase chain reaction.

pathogen, and management of any complications. Complications from meningitis can include visual impairment, optic neuritis, deafness, headache, seizures, personality changes, motor weakness, hydrocephalus, endocarditis, and pneumonia. Much of the damage to CNS structures is not a direct result of the pathogen; rather, it is the immune response that is injurious. Antibiotic therapy, with resultant bacterial cell wall lysis, can increase the immune-mediated injury. This has led some investigators to recommend the use of corticosteroids during the antibiotic treatment phase. However, treatment with dexamethasone remains controversial.

Prevention strategies include public education promoting prompt and appropriate management of sinusitis, mastoiditis, ear infections, and pneumonia. Strict aseptic techniques for all procedures involving a break in the CNS barrier system may help prevent nosocomial CNS infections. Vaccination against *N. meningitidis* provides short-term protection (a few years) and may be useful before situations in which exposure is more likely, such as during the college years.

ENCEPHALITIS

Etiology

Encephalitis, an inflammation of the brain, can be caused by a variety of agents, including viruses, bacteria, fungi, and parasites. Viral causes account for the vast majority of encephalitis cases. There are an estimated 20,000 cases of acute viral encephalitis annually in the United States. Death occurs in 5% to 20% of affected individuals, and another 20% are left with residual neurologic deficits of varying severity. In herpes simplex encephalitis, approximately 50% of patients die or are left with impairment. Western/eastern equine and West Nile infections only cause death or neurologic deficit in 5% to 15% of patients. Herpes viruses are by far the most common viral cause. It is almost always associated with herpes simplex virus type 1 (HSV-1), which is also the cause of herpetic lesions of the oral mucosa; however, it can also be caused by HSV-2. Western equine and West Nile viruses are arthropod-borne viruses transmitted to humans primarily through the bites of infected mosquitoes or insects.

Pathogenesis and Manifestations

HSV-1 encephalitis occurs sporadically in healthy and immunocompromised adults. The HSV-1 virus lies dormant in the trigeminal nerve and is reactivated, or it infects the nose and travels along the olfactory tracts. HSV-2 can be transmitted to neonates during vaginal delivery. Once across the BBB, the virus enters the neurons and disrupts cellular functioning, causing bleeding and inflammation. HSV forms intense hemorrhagic necrosis of the inferior and medial temporal lobes and mediorbital parts of the frontal lobe.

Clinical manifestations of HSV encephalitis typically evolve over several days and commonly include fever, headache, seizure, confusion, stupor, and coma. Hallucinations, personality changes, and psychotic behavior also may occur. Lumbar puncture shows increased opening pressure with elevated numbers of WBCs. Rarely do red blood cells (RBCs) appear in the CSF despite the hemorrhagic nature of the lesions. CSF protein level may be elevated, and glucose level will likely be normal. CT and MRI may show characteristic scattered hemorrhages with surrounding edema.

West Nile virus is transmitted by mosquitos and ticks, and spread primarily to mammals and other hosts such as birds. It is primarily a disease of summer months. It most often infects the very young, those older than 50, and those with immune compromise. Once inside the CNS these viruses infect neurons and cause severe immunopathology and apoptosis. The mechanism used by these viruses to cross the BBB and invade the CNS is unclear.

If symptoms occur (80% are asymptomatic), onset is generally rapid and includes malaise, mild headache, and often nausea and vomiting. A moderately elevated temperature develops, and the headache usually becomes more severe. In an uncomplicated infection, symptoms persist about 14 days and gradually resolve. In severe cases, lethargy progresses to stupor alternating with extreme restlessness. In the rare fatal cases the progression of the disease is rapid, culminating in coma and death.

Treatment

In general, the management of encephalitis is supportive and symptomatic. As with all severe illnesses, respiratory and cardiovascular support is imperative. Patients with encephalitis must be carefully hydrated because they frequently show signs and symptoms of excessive antidiuretic hormone secretion and water retention. Those with moderate-to-severe disease require careful and ongoing neurologic assessment. Seizures are a common complication in encephalitis secondary to hypoxia, tissue destruction, toxic encephalopathy, inflammatory vasculitis, and hyponatremia. All patients with moderate-to-severe illness should be monitored for intracranial hypertension. Although there is no definitive drug treatment, steroids may be given to control edema, anticonvulsants to prevent seizures, analgesics to relieve headaches, and antipyretics to control hyperthermia. Patients in whom herpes simplex encephalitis has been diagnosed should be treated with antiviral medications such as acyclovir.

BRAIN ABSCESS

Etiology

A brain abscess is a localized collection of pus within the brain parenchyma. Pyogenic (pus-producing) pathogens reach the brain by a number of routes, including (1) penetrating wounds, (2) direct extension or retrograde thrombophlebitis of an infected neighboring structure (e.g., mastoiditis, sinusitis), or (3) bloodborne dissemination from a distant infected site (e.g., the lungs). Most brain abscesses are bacterial. The most common infective organisms are streptococci, enterobacteria, and anaerobes.

Pathogenesis and Manifestations

Brain abscess presents as a space-occupying lesion in the brain. Most patients experience symptoms 1 to 4 weeks after the initial infection. The abscess has a focal infected core in which the central portion contains

FIG 44.18 CT scan of a cerebral abscess showing typical ring with decreased core density and an edematous area surrounding the abscess. (From Yousem DM, Grossman RI: *Neuroradiology,* ed 3, St Louis, 2010, Mosby, p 199.)

an abundance of neutrophils and tissue debris (pus). The peripheral portion of the abscess consists of inflammatory granulation tissue. Around the abscess is perifocal edema with proliferation of surviving astrocytes. In the chronic phase, the core of the abscess is liquefied, and the peripheral portion forms a collagenous capsule that in turn is surrounded by fibrous gliosis. A CT scan typically shows an outer ring surrounding a low-density core (Fig. 44.18).

Treatment

Management of a brain abscess depends on its location and accessibility, and usually involves drainage or excision. A critical feature in management is the administration of intravenous antibiotics, which is required for several weeks. Recently the treatment of patients with brain abscess has become increasingly challenging because of the increase in unusual bacterial, fungal, and parasitic infections, particularly in immunosuppressed patients. Postinfection care must address residual neurologic deficits of cognitive, motor, or sensory function.

KEY POINTS

- Meningitis is usually a consequence of bacterial infection in the CNS. Infection may be introduced through the bloodstream or by invasion from infected sinuses or ears. Fever, stiff neck, and headache are common. Seizures may occur. The diagnosis is based on an elevated CSF WBC count and the presence of bacteria in the CSF.
- Obstructive hydrocephalus is a serious complication of meningitis that leads to increased ICP. Antibiotics are used for treatment.
- Encephalitis is inflammation of the brain that is most commonly due to viral infection. Common causes of viral encephalitis in the United States include West Nile virus, western equine encephalitis, and herpes simplex virus. Management is based on symptoms and may include steroids, anticonvulsants, analgesics, and antipyretics. Antiviral agents (e.g., acyclovir) are helpful in the treatment of herpes simplex encephalitis.
- Brain abscesses are usually due to pus-forming bacteria. Abscesses may be asymptomatic at first, later showing manifestations of a progressive space-occupying lesion. Drainage or excision and antibiotics are indicated.

SUMMARY

Acute disorders of brain function are characterized by rapidly progressing neurologic deficits and life-threatening complications. The cellular pathophysiologic process is similar for most types of brain injury and includes mechanisms of ischemia, cellular calcium overload, and free radical and immune-mediated damage. The development of increased ICP with compression of vital brain structures is a potential complication of all types of brain injury.

Efforts to minimize brain damage focus on recognizing and managing secondary brain damage. Careful monitoring and management of body temperature, blood pressure, volume status, and respiratory function are essential. Efforts to reduce brain ischemia are important because it is thought to be a critical factor in acute brain injury.

The acute brain injury disorders presented in this chapter, including TBI, stroke, vascular rupture, and CNS infections, are all largely preventable. Efforts at prevention are paramount because often the outcome of acute brain injury is poor.

RESOURCES

Traumatic Brain Injury

Andrews P, Sinclair H, Rodriguez A, et al: Hypothermia for intracranial hypertension after traumatic brain injury. *NEJM* 373(25):2403–2412, 2015. doi:10.1056/nejmoa1507581.

Borlongan C, Acosta S, de la Pena I, et al: Neuroinflammatory responses to traumatic brain injury: etiology, clinical consequences, and therapeutic opportunities. *NDT* 97, 2015. doi:10.2147/ndt.s65815.

Cdc.gov: CDC | Get the Facts | Traumatic Brain Injury | Injury Center. 2015. Available at: http://www.cdc.gov/traumaticbraininjury/get_the_facts.html. Accessed 6 December 2015.

Cushing H: *Studies in intracranial physiology and surgery*, London, 1926, Oxford University Press, pp 19–23.

Hall JE: *Guyton and Hall textbook of medical physiology*, ed 13, Philadelphia, 2016, Elsevier.

Hemphill J, III, Smith WS, Gress DR: Chapter 275. Neurologic critical care, including hypoxic-ischemic encephalopathy, and subarachnoid hemorrhage. In Longo DL, Fauci AS, Kasper DL, et al, editors: *Harrison's principles of internal medicine*, ed 18, 2012. From: http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=331&Sectionid=40727062. Accessed 28 November 2015.

Hinson HE, Rowell S, Schreiber M: Clinical evidence of inflammation driving secondary brain injury: a systematic review. *J Trauma Acute Care Surg* 78(1):184–191, 2015.

Lai T, Zhang S, Wang Y: Excitotoxicity and stroke: identifying novel targets for neuroprotection. *Prog Neurobiol* 115:157–188, 2014.

Mohseni-Bod H, Drake J, Kukreti V: Management of raised intracranial pressure in children with traumatic brain injury. *J Pediatr Neurosci* 9(3):207, 2014. doi:10.4103/1817-1745.147572.

Sheriff F, Hinson H: Pathophysiology and clinical management of moderate and severe traumatic brain injury in the ICU. *Semin Neurol* 35(01):42–49, 2015. doi:10.1055/s-0035-1544238.

Stroke

Aminoff MJ, Kerchner GA: Nervous system disorders. In Papadakis MA, McPhee SJ, Rabow MW, editors: *Current medical diagnosis & treatment 2015*, New York, 2014, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=1019&Sectionid=57668616. (Accessed December 7, 2015).

Greenberg DA, Aminoff MJ, Simon RP: Chapter 13. Stroke. In Greenberg DA, Aminoff MJ, Simon RP, editors: *Clinical neurology*, ed 8, New York, 2012, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=398&Sectionid=39812250. (Accessed December 3, 2015).

Ling H, et al: Neurological consequences of traumatic brain injuries in sports. *Mol Cell Neurosci* 2015. http://dx.doi.org/10.1016/j.mcn.2015.03.012.

Sacco R, Kasner S, Broderick J, et al: An updated definition of stroke for the 21st century: a statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke* 44(7):2064–2089, 2013. doi:10.1161/str.0b013e318296aeca.

Strokeassociation.org: About Stroke. 2015. Available at: http://www.strokeassociation.org/STROKEORG/AboutStroke/About-Stroke_UCM_308529_SubHomePage.jsp. Accessed 5 December 2015.

Aneurysm and A-V Malformation

Darsaut T, Magro E, Gentric J, et al: Treatment of brain AVMs (TOBAS): study protocol for a pragmatic randomized controlled trial. *Trials* 16(1):2015. doi:10.1186/s13063-015-1019-0.

Ninds.nih.gov: Cerebral aneurysms fact sheet. 2015. Available at: http://www.ninds.nih.gov/disorders/cerebral_aneurysm/detail_cerebral_aneurysms.htm#3098_4. Accessed 7 December 2015.

Ropper AH, Samuels MA, Klein JP: Chapter 34. Cerebrovascular diseases. In Ropper AH, Samuels MA, Klein JP, editors: *Adams & Victor's principles of neurology*, ed 10, New York, 2014, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=690&Sectionid=50910885. (Accessed December 7, 2015).

Schatlo B, Fathi A, Fandino J: Management of aneurysmal subarachnoid hemorrhage. *Swiss Med Wkly* 2014(144):W13934, 2014.

Smith WS, Johnston S, Hemphill J, III.: Cerebrovascular diseases. In Kasper D, Fauci A, Hauser S, et al, editors: *Harrison's principles of internal medicine*, ed 19, New York, 2015, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=1130&Sectionid=79755261. (Accessed December 7, 2015).

Statler M: Evaluation and treatment of acute cerebral ischemia. *Clin Advis* 16(13):34–39, 2013.

Brain Infections

Kim H, Su H, Weinsheimer S, et al: Brain arteriovenous malformation pathogenesis: a response-to-injury paradigm. *Acta Neurochir Suppl* 111:83–92, 2011.

Roos KL, Tyler KL: Chapter 381. Meningitis, encephalitis, brain abscess, and empyema. In Longo DL, Fauci AS, Kasper DL, et al, editors: *Harrison's principles of internal medicine*, ed 18, New York, 2012, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=331&Sectionid=40727197. (Accessed December 7, 2015).

Ropper AH, Samuels MA: Chapter 32. Infections of the nervous system (bacterial, fungal, spirochetal, parasitic) and sarcoidosis. In Ropper AH, Samuels MA, editors: *Adams and Victor's principles of neurology*, ed 9, New York, 2009, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=354&Sectionid=40236345. (Accessed December 7, 2015).

Ropper AH, Samuels MA: Chapter 33. Viral infections of the nervous system, chronic meningitis, and prion diseases. In Ropper AH, Samuels MA, editors: *Adams and Victor's principles of neurology*, ed 9, New York, 2009, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=354&Sectionid=40236346. (Accessed December 7, 2015).

Winkelmann ER, Luo H, Wang T: West Nile virus infection in the central nervous system [version 1; referees: 3 approved]. F1000Research 2016, 5(F1000 Faculty Rev):105 (doi: 10.12688/f1000research.7404.1).

45

Chronic Disorders of Neurologic Function

Joni D. Marsh

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- How are the various types of seizures recognized, classified, and treated?
- How is Alzheimer dementia diagnosed and managed?
- What are the similarities between Alzheimer dementia and vascular dementia?
- What are the proposed neurotransmitter alterations in Parkinson disease, and how are drugs used to restore balance?
- What are the similarities and differences between multiple sclerosis and amyotrophic lateral sclerosis?
- How are congenital disorders, such as cerebral palsy, hydrocephalus, and spina bifida, manifested in the newborn?
- How does the level of spinal cord injury relate to expected functional losses and clinical manifestations?
- What are the roles of immune mechanisms in Guillain-Barré syndrome, amyotrophic lateral sclerosis, and multiple sclerosis?
- What are the causes of facial paralysis in Bell palsy, and how is this condition different from other chronic disorders of neurologic function?

CHAPTER OUTLINE

Brain and Cerebellar Disorders, 916
Seizure Disorder, 916
Etiology, 916
Pathogenesis, 916
Clinical Manifestations, 916
Diagnosis Treatment, 917
Dementia, 918
Etiology, 918
Pathogenesis, 918
Clinical Manifestations, 920
Diagnosis and Treatment, 920
Parkinson Disease, 920
Etiology, 920
Pathogenesis, 921
Clinical Manifestations and Treatment, 921
Cerebral Palsy, 922
Etiology and Pathogenesis, 922
Clinical Manifestations, 922
Treatment, 923
Hydrocephalus, 923
Etiology, 923
Pathogenesis and Clinical Manifestations, 924
Treatment, 924
Cerebellar Disorders, 924
Etiology and Clinical Manifestations, 924

Spinal Cord and Peripheral Nerve Disorders, 925
Multiple Sclerosis, 925
Etiology, 925
Pathogenesis, 925
Clinical Manifestations and Treatment, 925
Spina Bifida, 927
Etiology and Pathogenesis, 927
Clinical Manifestations, 927
Treatment, 931
Amyotrophic Lateral Sclerosis, 928
Etiology and Pathogenesis, 928
Clinical Manifestations and Treatment, 928
Spinal Cord Injury, 928
Etiology, 929
Pathogenesis, 929
Clinical Manifestations, 931
Treatment, 931
Guillain-Barré Syndrome, 931
Etiology and Pathogenesis, 931
Clinical Manifestations and Treatment, 931
Bell Palsy, 932
Etiology and Pathogenesis, 932
Clinical Manifestations and Treatment, 932

Patients experiencing neurologic dysfunction from chronic disease states present a challenge to health care professionals, who must strive to maximize the patient's function and quality of life. This chapter focuses on common chronic disabilities of neurologic function, including those primarily affecting the brain such as seizures, dementia, Parkinson disease, cerebral palsy, and hydrocephalus. Disorders of the spinal cord or peripheral nervous system include multiple sclerosis (MS), spina bifida, and spinal cord injury. Guillain-Barré syndrome and Bell palsy are examples of disorders affecting the peripheral nervous system.

BRAIN AND CEREBELLAR DISORDERS

Seizure Disorder

Seizures are a transient neurologic event of paroxysmal abnormal or excessive cortical electrical discharges that are manifested by disturbances of skeletal motor function, sensation, autonomic visceral function, behavior, or consciousness. Symptoms are not constant, and the length of time between seizure episodes is extremely variable. A seizure may occur only once in a person's lifetime. *Epilepsy,* or *seizure disorder,* refers to recurrent seizures. Seizures are a component of many diseases. Approximately 50 million people worldwide have epilepsy.

Etiology

Seizures have many causes, and under the right circumstances anyone can experience a seizure. A seizure disorder can be acquired as a consequence of cerebral injury or other pathologic process, including structural lesions such as tumors, blood clots, or infection. Other causes include metabolic and nutritional disorders such as electrolyte and water imbalance, hypoxia, acidosis, pyridoxine deficiency, acute withdrawal from alcohol, therapeutic medication overdose or medication adverse effect, and exposure to toxins such as heavy metals or street drugs. If seizures develop as a result of a structural change such as head injury or stroke, the onset is not predictable. In some cases, seizures may not develop for months or years after the structural change has occurred. In some cases, no explanation for the seizure disorder can be found. These individuals are classified as having idiopathic seizures.

A seizure event is often triggered by specific stimuli, usually unique for each individual. Physical inducements include specific sensory stimuli such as flashing lights, loud noises, and rhythmic music. Fever, physical exhaustion, sleep deprivation, fatigue, inadequate nutrition, hormonal changes of the menstrual cycle, hyperventilation, injury, and drugs can also prompt seizure activity. Psychosocial factors include family and environmental stress, shock, and emotional stress.

Pathogenesis

Seizures are due to an alteration in membrane potential that makes certain neurons abnormally hyperactive and hypersensitive to changes in their environment. These physiologically abnormal neurons form an epileptogenic focus (i.e., an area of the brain from which the seizure emanates). The epileptogenic focus functions autonomously, emitting excessively large numbers of paroxysmal electrical discharges. Results from animal studies suggest that oxidative stress and free radical damage may be a cause or consequence of these electrical abnormalities. Nerve cells in this area can recruit neurons in adjacent areas, as well as synaptically related neurons in distant areas of the brain, greatly increasing the number of neurons involved in the seizure activity. Recruitment can also incorporate neurons in the opposite hemisphere. Clinical symptoms become evident when a sufficient number of neurons have been excited. Seizures are classified according to clinical symptoms and the electroencephalographic (EEG) features. Clinical manifestations depend on the area of the brain involved, the area of origin, and the areas to which the seizure spreads.

BOX 45.1 Classifications of Seizures

Generalized Seizures: Entire Brain Surface Is Affected During Seizure

- Absence (petit mal)
- Atypical absence
- Myoclonic
- Atonic (drop attack)
- Clonic
- Tonic
- Generalized tonic-clonic (grand mal)

Partial Seizures: Part of Brain Surface Is Affected During Seizure

- Simple partial: There is no impairment of consciousness during the seizure.
- Complex partial: There is impairment of consciousness during the seizure.
- With secondary generalization: Onset begins as simple partial and then progresses to impairment of consciousness.

Clinical Manifestations

Seizures may be classified as partial, in which only part of the brain surface is affected (also known as *focal seizures*), or generalized, in which the whole brain surface is affected during the seizure (Box 45.1).

Generalized seizures. Episodes in which the entire brain is involved from the onset of the seizure are referred to as *generalized seizures*. Involvement of the thalamus and reticular activating system results in loss of consciousness. Metabolic or toxin-induced seizures tend to be generalized. This category includes the following: absence (petite mal), atypical absence, myoclonic, atonic (drop attack), or tonic-clonic (grand mal) seizures.

Absence or petite mal seizures usually occur only in children and are sometimes identified in children manifesting poor academic performance. They are very brief (2 to 10 seconds), and episodes are characterized by staring spells that last only seconds. Onset and termination of attacks are abrupt. During the spell, the individual is unaware of the surrounding environment and is usually motionless; however, it is not unusual for the person to continue walking or performing a routine motor task. If the seizure activity occurs during conversation, the individual may pause or miss a few words. Absence seizures almost always occur during childhood and rarely begin earlier than age 4 or after puberty. Atypical absence seizures have accompanying myoclonic jerks and automatisms (such as lip smacking or repetitive semipurposeful movements) with the staring spell. The EEG patterns are unique to each syndrome. Myoclonic seizures are extremely brief and are characterized by a single jerk or multiple jerks of one or more muscle groups. Atonic seizures or drop attacks are characterized by a sudden and complete loss of muscle tone. Falls and injuries are common with this type of seizure activity. Myoclonic episodes may also be associated with atonic seizures. Tonic-clonic seizures involve stiffening and repetitive jerking of muscle groups.

Tonic-clonic or grand mal seizures are characterized by a sudden loss of consciousness followed by muscle rigidity (tonic phase). The individual falls, and initial motor signs include opening of the mouth and eyes, extension of the legs, and adduction of the arms. There may be tongue biting or a high-pitched cry while the whole musculature is in spasm and air is forced out of the lungs through closed vocal cords. Respiration is arrested, and cyanosis may occur. Bowel and bladder incontinence frequently occurs. The tonic phase may last 10 to 15 seconds and is followed by clonic activity, in which there is often violent but rhythmic muscular contractions. During this phase the eyes roll, the face grimaces, and the pulse rate accelerates. Salivation increases, and the patient may become diaphoretic. The clonic phase usually lasts 1 to 2 minutes with a gradual decline in the amplitude of the clonic jerks.

The individual remains apneic until the end of the clonic phase that is marked by a deep inspiration.

During the terminal or postictal phase, the individual may regain consciousness or drift into a deep comalike state. Disorientation and confusion are common. If allowed, the individual may sleep for several hours. Other findings include headache, drowsiness, nausea, muscle soreness, no memory of the seizure event, and retrograde amnesia. During the seizure, the person is at risk for injury from the initial fall, as well as from the muscle contractions of the clonic phase.

A potentially life-threatening situation known as status epilepticus occurs in some seizure disorders. Status epilepticus is a continuing series of seizures without a period of recovery between episodes. It can occur with all types of seizures, but is of greatest concern in tonic-clonic seizures. Irreversible brain damage and possible death from hypoxia, cardiac dysrhythmias, or lactic acidosis can occur if the airway is not maintained and seizure activity is not halted. Evidence suggests that nonconvulsive status epilepticus is associated with cellular injury in the area of the seizure focus.

Partial seizures. Partial seizures are those in which activity is restricted to one brain hemisphere. They are further divided into three categories: (1) simple partial, (2) complex partial, and (3) partial seizures that are secondarily generalized.

In simple partial seizures, the individual does not have a change in level of consciousness. The symptoms may be motor, sensory, or autonomic or any combination of the three. Motor symptoms may be limited to one part of the body. Sensory seizures may result in tingling or numbness that spreads, or "marches," to different parts of the limb or body (depending on the location of the seizure activity in the brain) or may involve the special senses, producing auditory (buzzing sounds), olfactory, or visual manifestations (flashing lights). Autonomic symptoms may include pupillary (pupil dilation), skin (diaphoresis, flushing), or respiratory changes.

Complex partial seizures have many different combinations of cognitive, affective, and psychomotor symptoms. Either loss or alteration of consciousness may occur when the seizure begins. After the attack, the individual may feel drowsy or confused. At the onset of impairment of consciousness, the individual often displays automatisms. Aggressive behavior may be displayed as well, especially if bystanders attempt to restrain the individual. Complex partial seizures often last several minutes and may be followed by a postictal state.

Partial seizures that are secondarily generalized are the third subtype of partial seizures. This category comprises seizures that begin as simple partial seizures and then progress to involve both brain hemispheres. Once generalized, these seizures are clinically similar to primary generalized seizures.

Aura/prodrome. Some people may have a subjective sense of an impending seizure. This prodromal period may be characterized by any one of several phenomena such as a type of myoclonic jerking, headache, lethargy, mood alterations, palpitations, or epigastric sensations, which may precede the actual seizure by several hours. In about half of cases there is some type of movement or odd sensory experience (visual, auditory, olfactory, or gustatory) that occurs seconds before consciousness is lost and that is remembered by the individual after recovery from the seizure. This experience is known as an *aura.* Although the individual may interpret the aura as an indication that a seizure is about to occur, in fact, it is the beginning of the seizure episode. Auras can be significant, because they may be a clue to the location of the epileptogenic focus.

Diagnosis and Treatment

The diagnosis and management of seizure disorders are based on the patient's history, physical, and neurologic examination results, as well as the results of EEG studies. Electroencephalograms may be normal between seizures, so activation techniques (sleep deprivation, hyperventilation) may be used to elicit the pathologic mechanism. Laboratory studies are frequently used to investigate possible metabolic abnormalities, as well as therapeutic serum levels in those already using anticonvulsant drugs. Lumbar puncture may be utilized when there is a suspicion of a central nervous system (CNS) infection. Initial studies ruling out structural causes may include computed tomography (CT) or magnetic resonance imaging (MRI).

Treatment of an individual experiencing a seizure is concentrated on maintaining an airway and protecting the individual from injury. Recording the course of the seizure episode is useful for identifying the location of the epileptogenic focus and for noting any change in the patient's seizure pattern. These data are useful in treatment planning. The information recorded should include the time of onset and duration of the seizure, precipitating factors, presence of a prodrome or aura, and sequence of seizure activity. Autonomic signs, level of consciousness, and postictal state are also essential to document.

Long-term treatment depends on the cause of the seizure disorder. In seizures resulting from a metabolic abnormality, infection, or tumor, the precipitating source is removed. If the seizures are due to irreversible or unidentifiable factors, anticonvulsant medications specific to the type of seizure are the best management. The decision to treat after one seizure is controversial when an identifiable cause has not been found. The objective of therapy is to achieve seizure control with a minimum of side effects. Medication is continued until there have been no seizures for at least 2 years and is then gradually withdrawn. If seizures continue despite treatment at a maximal dose of a single medication, a second agent is added and the dosage is increased depending on patient tolerance. The first drug is then gradually discontinued. Anticonvulsant medication is a form of control, not a cure.

Treatment also includes patient education in the avoidance of activating factors (e.g., stress, loud noise, alcohol). Patients should be advised to avoid situations that could be dangerous or life threatening if seizures should recur (e.g., driving or swimming). State laws defining when patients with seizure disorders are allowed to resume driving vary widely. Compliance to the treatment plan is sometimes difficult because of side effects of pharmacologic interventions. However, most patients are able to achieve optimal seizure control and lead active and productive lives. For some patients with seizure disorder uncontrolled by medications, surgical excision of the seizure focus may be an option. Neurostimulation is an appropriate therapy for certain patients with refractory seizures.

KEY POINTS

- Seizure disorder is characterized by recurrent episodes of abnormal electrical impulses in the brain. Some individuals appear to have a lower-than-normal threshold for seizure activity. Seizure activity may occur in anyone, given the right conditions. Head injury, meningitis, brain tumors, and metabolic disorders (electrolyte imbalance, fever, acidosis) may predispose an individual to having seizures.
- Initiation of seizure activity may occur in a particular brain area (the epileptogenic focus). Nearby and distant neurons may then be recruited into the seizure. When sufficient neurons are involved, the seizure becomes clinically evident as involuntary movement or unusual sensations.
- Seizures are classified as partial or generalized. Partial seizures involve a part of the brain; generalized seizures involve the entire brain at the onset. Partial seizures are further classified as simple, in which consciousness is retained, and complex, in which consciousness is impaired. Seizures may

begin as partial and then generalize to affect the entire brain. Generalized seizures include absence, myoclonic, atonic, and tonic-clonic types. Consciousness is always impaired in generalized seizures.
- Status epilepticus is a serious condition in which seizures occur continuously, resulting in intense brain metabolism. Ischemic brain damage may result.
- Management of a seizure in progress is aimed at maintaining the individual's airway and protecting the person from trauma. Close attention is given to the quality and progression of seizure activity. Anticonvulsant medications are used to suppress seizure activity.

Dementia

Dementia is not a specific disease but rather a syndrome associated with many pathologic processes. It is characterized by progressive deterioration and continuing decline of memory and other cognitive changes. Personality and behavior changes accompany the cognitive deterioration. Judgment, abstract thinking, and complex task performance are all affected. The onset of dementia may be insidious, and the affected individual may initially appear uninterested or lacking initiative. Many demented patients have agnosia, or lack of insight into their cognitive deficiencies.

Alzheimer disease accounts for 60% to 80% of all dementias, whereas vascular dementia is the second most common cause. An estimated one in nine people older than 65 have Alzheimer disease. It affects nearly one third of those 85 years of age and older.

Etiology

Multiple causes/types of dementia exist, and a full discussion of each is beyond the scope of this chapter. Some examples of dementia-causing illness include alcoholism, intracranial tumor, normal-pressure hydrocephalus, Parkinson disease, Lewy body disease, Huntington disease, multiple sclerosis, Pick disease, Creutzfeldt-Jakob disease, and bovine spongiform encephalopathy (mad cow disease). Unfortunately, these also can occur in combination, causing severe disease. Because Alzheimer- and vascular-type dementias are the first and second most common causes of dementia, they will be discussed in detail. The subsequent discussion of treatment of individuals with dementia will be more general because the care issues are similar, regardless of the type.

It is important to consider all potential causes of cognitive change when dealing with patients with mental status change. Both delirium and depression in the elderly can cause signs and symptoms that resemble those of dementia. Delirium is a global mental dysfunction or an acute confusional state that includes disturbed consciousness, decreased awareness of the environment, inability to maintain attention, disrupted sleep–wake cycles, drowsiness, restlessness, emotional lability, incoherence, and hallucinations. Symptoms of delirium tend to have an abrupt onset and may fluctuate, often becoming worse at night. Delirium can result from numerous causes such as medication/polypharmacy, metabolic abnormalities, nutritional deficiencies, and infection, among others. Delirium may occur more frequently in individuals with an underlying dementing illness.

Pathogenesis

For decades Alzheimer research has focused on the hallmark structural changes of intracellular neurofibrillary tangles and extracellular amyloid (senile) plaques (Fig. 45.1). However, more recently, research has been focused on the role of neuroinflammation. Neurofibrillary tangles are composed of helical filaments formed from hyperphosphorylated protein tau, also known as *neural thread protein*. In the CNS, neural thread proteins bind and help stabilize microtubules (the cell's internal support structure or skeleton). Inflammatory changes, lipid abnormalities, and aging are among the processes thought to be responsible for activating the phosphorylating enzymes altering the structure of the tau proteins. The presence of neurofibrillary tangles is well correlated with dementia; however, neurofibrillary tangles are not specific just to Alzheimer disease and are found in other neurodegenerative disease processes. The second, but more specific, change in the brain of patients with Alzheimer disease is the deposition of extracellular amyloid plaques.

It is not known whether amyloid plaques cause Alzheimer disease or result from it. The number of senile plaques seems to correlate with disease severity. In plaques, β-amyloid is a protein fragment snipped from a larger protein—amyloid precursor protein (APP)—during metabolism. APP is a member of a large family of proteins that are associated with cell membranes. During metabolism, APP becomes embedded in the membrane of the nerve cell, partly inside and partly outside the cell. While APP is embedded in the cell membrane, proteases cleave APP apart. β-Amyloid is produced only when the cleavage happens at the wrong place in APP.

After β-amyloid is formed, it is not known how it moves through or around the nerve cells. In the final stages of its journey, it joins with other β-amyloid filaments and fragments of dead and dying neurons to form the dense, insoluble plaques that are a hallmark of Alzheimer disease in brain tissue.

Along with these characteristic structural changes, inflammation is another finding of Alzheimer disease. Cytokines, complement

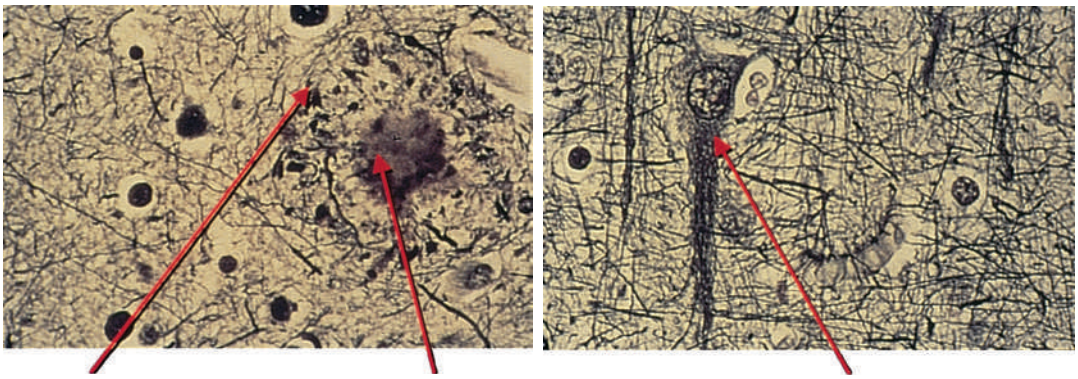


FIG 45.1 Amyloid plaques and neurofibrillary tangles. (Courtesy James King-Holmes and Science Photo Library.)

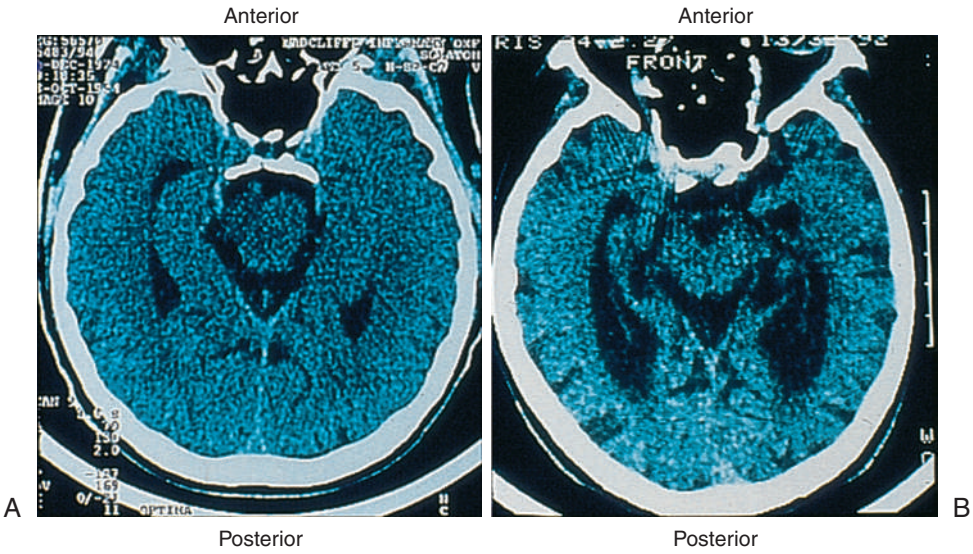


FIG 45.2 Axial (horizontal) CT scan section through the temporal lobes. **A,** Normal. **B,** Alzheimer disease. (Courtesy James King-Holmes and Science Photo Library.)

defense proteins, acute-phase reactants, signs of microglial activation and scavenger attack proteins, and other signs of inflammation are found in the brains of affected individuals. Although it is clear that inflammation is present, it has been more difficult to show whether the inflammation is contributing to the structural aberrations and damage or whether the disordered proteins may be causing a secondary immune response.

As a result of these changes, diffuse neuronal damage and brain atrophy occur. The brain of a patient with advanced Alzheimer disease often weighs up to 20% less than a normal brain. The temporoparietal and anterior frontal regions of the brain are chiefly affected, exhibiting enlarged sulci and ventricles and atrophic gyri (Fig. 45.2).

Much interest exists in the neurotransmitter systems in relation to Alzheimer disease. Damage in Alzheimer disease involves changes in three mechanisms: nerve cell communication, metabolism, and repair. Several studies have found abnormalities in the cholinergic system, including reduced activity of choline acetyltransferase (the enzyme necessary for acetylcholine synthesis) and decreased acetylcholine synthesis (Fig. 45.3). Some researchers believe that β-amyloid may be responsible for lower choline levels in nerve cells and decreased acetylcholine levels. The degeneration of cells in the nucleus basalis, a band of gray matter in the ventral portion of the medulla oblongata, has also been linked to diminished levels of acetylcholine in the cerebral cortex. Along with alterations in acetylcholine, other neurotransmitters are also affected. Imbalances in the activity of glutamate, dopamine, and serotonin contribute to the behavioral signs and symptoms of Alzheimer disease.

Vascular dementia results from a single cerebrovascular insult (such as cerebral infarction), from multiple lacunar infarcts, or from microvascular pathology. Microvascular insults may not show any localizing clinical symptoms and may be found incidentally on brain imaging. However, the presence of these also does not automatically mean a diagnosis of vascular dementia. The symptoms of vascular dementia may be similar to those of Alzheimer dementia. Most research/reports present vascular dementia and Alzheimer dementia as entirely separate entities; however, there is increasing evidence that particularly in elderly patients, the brain lesions associated with both often coexist. A common link between these diseases may be activation of innate immune responses, including those mediated by microglia in the CNS.

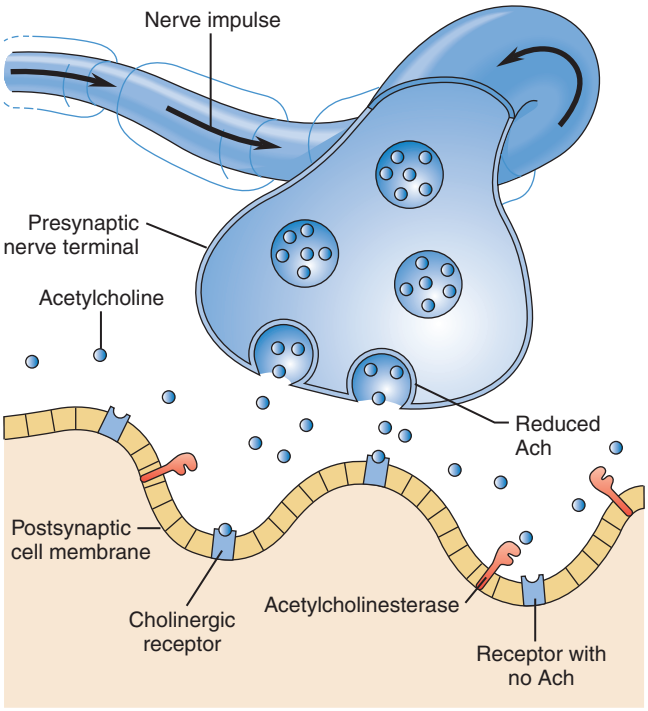


FIG 45.3 Cholinergic synapse.

The primary risk factors for the development of Alzheimer disease include age and family history. Epidemiologic studies show that individuals who have an affected first-degree relative with Alzheimer disease have higher risk of developing the disease than those without. The risk is greater if there are individuals in more than one generation with the disease. Three main genes have been identified and are thought to be responsible for autosomal-dominant familial Alzheimer disease. *Presenilin 1, presenilin 2,* and *APP* are believed to increase the amount of β-amyloid protein. These genes are rare, accounting for only 1% of Alzheimer disease cases, usually the early-onset variant with the development of signs and symptoms between the ages of 30 and 60. The major gene associated with late-onset Alzheimer disease is apolipoprotein e4 *(APOe4).*

This protein is closely tied to the innate immunity and is the strongest known risk factor for sporadic late-onset Alzheimer disease. Carrying one e4 allele increases the risk for Alzheimer disease by two- or threefold, whereas two alleles virtually assures the development of Alzheimer disease in all of those who survive into their 80s. Interestingly, this same allele contributes to an elevated low-density lipoprotein fraction in the serum.

Lifestyle has also been linked to the risk of Alzheimer disease. Head trauma, diabetes, and depression have been linked to an increased incidence of the disease, as well as lower educational level, poor diet, and inactive mental and physical lifestyle.

Risk factors for vascular dementia include those for stroke, hypertension, and diabetes. Clearly both vascular dementia and Alzheimer disease include a complex interplay of genetic, environmental, and lifestyle factors.

Clinical Manifestations

Regardless of when patients first present with dementia, it is likely that brain disease has been present for quite some time. Most patients experience a gradual onset with a chronic progressive decline in cognitive functioning. There is memory loss, especially in short-term memory, whereas long-term memory may be preserved. Thinking ability declines, and there is a decreasing ability to function at work and in social settings. Anxiety and agitation are common. As the disease progresses, individuals have increasing difficulty with judgment, problem-solving, and communication. Assistance may be necessary for completing activities of daily living (ADLs). Difficulty with eating and swallowing and weight loss are common. Loss of bladder and bowel control and eventual complete loss of the ability to ambulate occur in the late stages.

Diagnosis and Treatment

The initial evaluation of a patient thought to have dementia of any type begins with a complete history and physical examination. This should address the patient's overall general health and any coexisting medical conditions. All manageable causes for dementia or delirium should be ruled out. It is recommended that the evaluation include a complete blood cell count, chemistry panel, thyroid function, vitamin B_{12} levels, and syphilis serology. Other testing such as Lyme serology, human immunodeficiency virus (HIV), urinalysis with culture/sensitivity, heavy-metal assays, sedimentation rate, and other vitamin levels may be warranted in certain patient situations. Other evaluations such as a chest x-ray and lumbar puncture may also be helpful. Neuroimaging may include CT and magnetic resonance imaging/magnetic resonance angiography, which may identify vascular disease, normal-pressure hydrocephalus, tumors, abscesses, or subdural hematoma. Positron emission tomography) scans are not routinely recommended at this point, but show promise in identifying Alzheimer disease when combined with a history of genetic risk. Mental status examinations, the clock drawing test, and tests of functional status are recommended. A current list of the patient's medications, including over-the-counter medications, must also be reviewed. Medications with anticholinergic actions/side effects are a common cause of changes in cognitive functioning in the elderly.

Early diagnosis and intervention are key in the management of dementia. The financial and legal ramifications of dementia can be devastating to patients and their families and caregivers. If the diagnosis is made before the onset of severe cognitive disability, the patient can be involved in decisions regarding long-term care, power of attorney, and living will issues. Early diagnosis is also vital to initiating therapy as early as possible.

Currently, two classes of drugs are approved by the Food and Drug Administration (FDA) for the treatment of Alzheimer disease. The first class is the acetylcholinesterase inhibitors: donepezil (Aricept), rivastigmine (Exelon), and galantamine (Reminyl). These agents are indicated for use in patients with mild-to-moderate Alzheimer disease. Although not a cure, the acetylcholinesterase inhibitors have been shown to stabilize cognitive function and slow progression of the illness.

The second class of drugs used in the treatment of Alzheimer disease is known as the *N-methyl-D-aspartate (NMDA) receptor antagonists.* Currently only one drug in this class is available in the United States. Memantine (Namenda) is indicated for the treatment of moderate-to-severe Alzheimer-type dementia. This drug blocks stimulation by the neuroexcitatory transmitter glutamate. Again, this medication is not a cure, but slows progression of the disease. Studies using combination therapy of acetylcholinesterase inhibitors and NMDA antagonists have demonstrated modest improvement in cognitive functioning.

Treatment for vascular dementia is aimed at reducing risk of additional cerebral damage, including controlling blood pressure, blood sugar, and serum lipids, along with cessation/avoidance of tobacco. Acetylcholinesterase inhibitors and NMDA receptor antagonists have not been shown to consistently improve functioning in patients with vascular dementia.

Many other medications, although not approved for use in treating Alzheimer disease and vascular dementia, are used to manage the symptoms, such as depression, sleep disturbance, agitation, and psychosis. These medications include antidepressants, anxiolytics, antipsychotics, and mood stabilizers.

Other treatments for dementia include optimal management of other coexisting illnesses, interventions aimed at wellness, regulation of optimal nutritional intake, and protection from injury. In early stages of the disease, most patients are cared for at home, often by family members. It is important that the home environment be safe and that there be measures in place to control wandering. Consistent routines and familiar surroundings allow the patient to feel more comfortable and experience less confusion. As the disease progresses, the individual with dementia may have to be placed in an alternative living situation such as a nursing home or assisted-living program. Caring for the caregivers of patients with dementia is important.

KEY POINTS

- Dementia refers to progressive degeneration of cognitive function attributable to organic causes. In many instances the cause is unknown. There are limited definitive treatments for dementia, none of which is a cure. It is important to first rule out manageable causes of mental impairment.
- The dementia of Alzheimer disease is characterized by degeneration of neurons in the temporal and frontal lobes, atrophy of the brain, and the formation of amyloid plaques and neurofibrillary tangles. The synthesis of brain acetylcholine is deficient. The cause of Alzheimer disease remains unknown, although genetic factors and environmental triggers are suspected.
- The behavioral problems of individuals with Alzheimer disease progress from forgetfulness to total inability for self-care. Depression and psychosis may be significant.

Parkinson Disease

Parkinson disease is a disorder of mobility that affects 1 million Americans. It is estimated that 60,000 new cases are diagnosed each year. The incidence increases with age, but an estimated 4% of those diagnosed are younger than age 50.

Etiology

Parkinson disease may be idiopathic or acquired. Idiopathic Parkinson disease has no demonstrable cause. Common causes of acquired

parkinsonism include infection, intoxication, and trauma. Typically, parkinsonism attributable to drug toxicity evolves rapidly, unlike the slow, insidious onset of the idiopathic form of the disease. Side effects of drugs of the phenothiazine class (e.g., chlorpromazine, prochlorperazine, and thioridazine) and butyrophenone class (e.g., haloperidol) may manifest in a parkinsonian syndrome at toxic levels. Discontinuing the medication generally results in improvement in the symptoms. However, the additional use of the anticholinergic antiparkinsonian drugs may aid in more rapid recovery. The rest of this discussion of Parkinson disease refers to the most common idiopathic type.

Pathogenesis

Parkinson disease results from degeneration of the pigmented dopaminergic neurons found in the substantia nigra (Fig. 45.4) and, to a lesser extent, neurons elsewhere in the brain. Cytoplasmic inclusions known as *Lewy bodies* may be found in the surviving neurons. Incidentally, Lewy bodies are found along with amyloid plaques at autopsy in the brains of some patients with a severe form of dementia. This suggests a possible link with Alzheimer disease. The exact cause of this degeneration is unknown, but mitochondrial dysfunction from oxidative stress, genetics, and environmental toxins have been implicated. Adaptive immunity may play a part in the progression of Parkinson disease by reacting to the abnormal proteins and causing neuroinflammation.

Several different gene groups have been identified as having a role in the development of Parkinson disease. In particular, identification of a mutation in the α-*synuclein* gene has become the focus of much interest. Although mutations of the gene are a rare cause of Parkinson disease, α-*synuclein* is abundant in neurons, specifically in presynaptic terminals, and is a major component of Lewy bodies. Another gene identified in the development of Parkinson disease is *PINK1/parkin.* Mitochondrial quality control is mediated by *parkin.* When functioning normally, it selectively recognizes and eliminates damaged mitochondria from the cell by autophagy. A third area of genetic mutation receiving attention involves the *LRRK2* mutations. The mechanism for cell death related to this mutation is not completely understood, but animal studies demonstrate that it is important in maintaining neurite length and branching.

Although genes have received much attention in Parkinson disease research, environmental factors have also been studied. High caffeine intake has been found to have an inverse relationship to the risk of developing the disorder. Long-term exposure to the pesticide rotenone and the herbicide paraquat have been linked to increasing risk for Parkinson disease. Whatever the cause of the degeneration of the dopaminergic cells, 75% to 80% of the neurons have died before any symptoms of the disease become apparent.

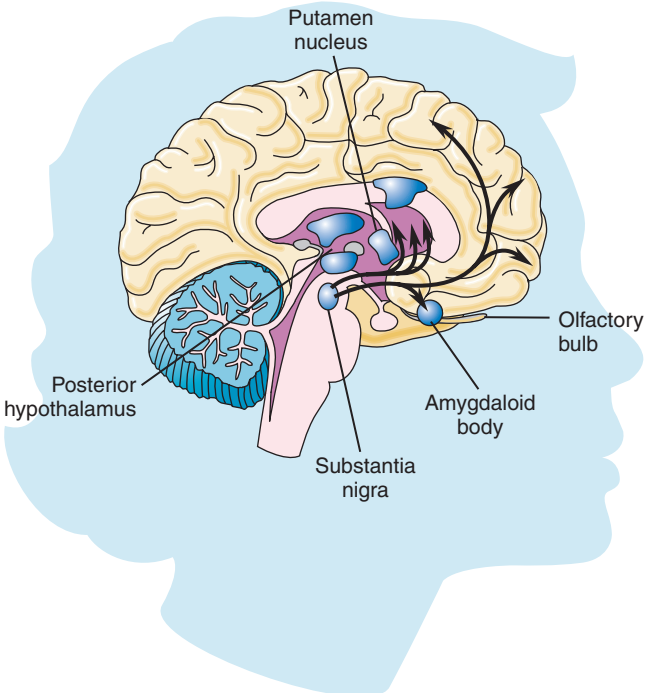


FIG 45.4 Dopaminergic neurons and their pathways in the human brain.

Clinical Manifestations and Treatment

Because of the insidious onset, earlier evidence of Parkinson disease may be discovered in a thorough health history. Frequently, the very early signs of the disorder (loss of flexibility, aching, and fatigue) are overlooked by the patient or are attributed to the aging process. Initially, symptoms are usually worse on one side of the body and then progress to involve both sides. Tremor is often the first symptom recognized that prompts patients to seek treatment. The tremor is generally at rest, unilaterally affecting distal extremities. Hand tremors may be described as pill-rolling movements. Attempts to passively move the extremities are met with cogwheel rigidity. As Parkinson disease progresses, the tremor will often become bilateral/more generalized. Additional early signs of the disease include bradykinesia, rigidity, hypokinesia, loss of facial expression, and infrequent eye blinking (Fig. 45.5). Again, these symptoms may be overlooked by patients but are usually apparent to observant family members.

As the disease progresses, additional functional changes are noted. The patient's handwriting may become small (micrographia) and cramped, with evidence of tremor. Speech may become low in volume, monotonous, and dysarthric. There may be a mumbling quality to the speech. The effects of bradykinesia are evident in the patient's swallowing function, ability to initiate activity, and level of mobility. Swallowing becomes delayed, so much so that the individual may drool, and patients are at risk for aspiration. The effect of the disease on the ability to initiate activity is evident when the individual rises from a chair or begins to walk from a standing-still position. However, many people with Parkinson disease are able to act quickly in times of emergency, such as fire. This phenomenon is known as *paradoxical kinesia.*

Additional difficulties in mobility are evident from the lack of spontaneous position changes while the individual is sitting in one position, from the decreased or absent arm swing while the individual is walking, and from the shuffling gait. Impairment of postural reflexes presents particular safety problems for the individual with Parkinson disease in maintaining balance, as evidenced by propulsive or retropulsive gaits. Involvement of the autonomic nervous system may result in orthostatic hypotension, which adds yet another risk to the individual's health. Because of these various impairments, falls are a common problem. Depression is present in many patients with Parkinson disease. Daytime sleepiness is also common, as well as sleep disturbances and restless leg syndrome. Dementia is prevalent in patients with Parkinson disease.

There is no known cure for Parkinson disease. Treatments are aimed at slowing the progression of the disease and managing symptoms. The mainstay of Parkinson therapy has been aimed at increasing the level of dopamine in the CNS. Dopamine precursors such as levodopa are one approach to increasing dopamine levels. Dopamine itself cannot be used because it does not cross the blood–brain barrier efficiently. Outside the CNS, levodopa is metabolized to dopamine and then to

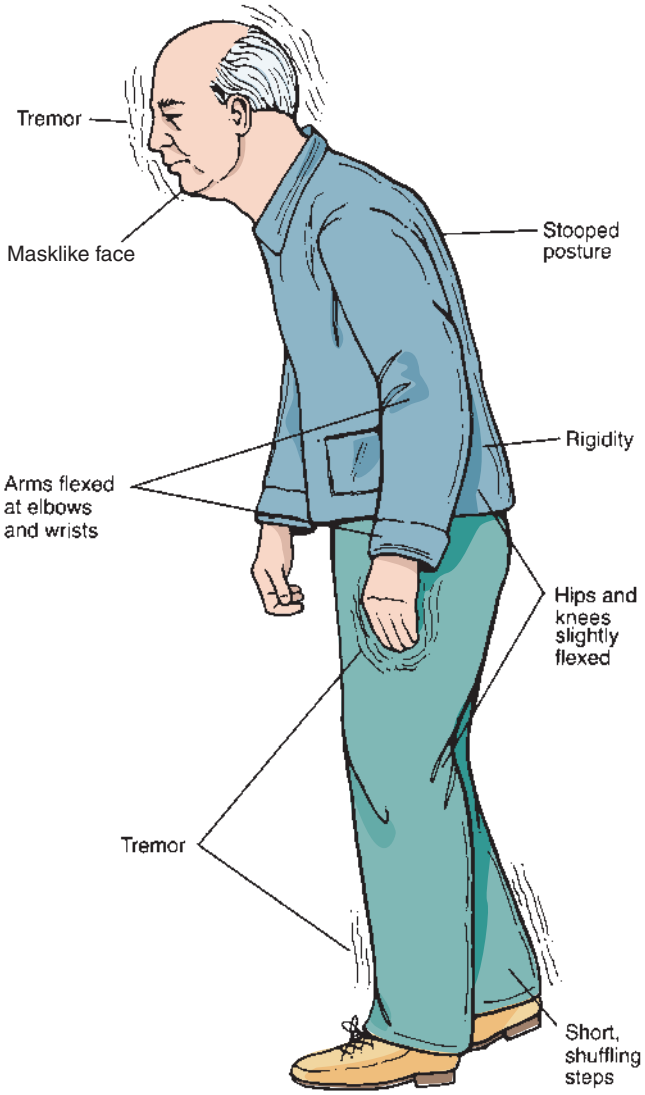


FIG 45.5 Clinical manifestations of Parkinson disease. (From Monahan FD, Neighbors M: *Phipps' medical-surgical nursing: health and illness perspectives,* ed 8, Philadelphia, 2007, Saunders, p 1446.)

adrenaline and noradrenaline, which can cause altered blood pressure. To minimize these side effects, levodopa is combined with carbidopa. This agent blocks the conversion of levodopa to dopamine in the periphery, allowing it to cross the blood–brain barrier. Long-term use of levodopa has been associated with "on-off" phenomena (in which the action of the drug suddenly stops, leaving the patient with sudden onset of symptoms) or abnormal movements called *dyskinesias.* Other medications aimed at increasing the level of dopamine in the CNS include dopamine receptor agonists (pramipexole, ropinirole), medications to slow the metabolism of dopamine (monoamine oxidase inhibitors such as selegiline, catechol-*O*-methyltransferase inhibitors), and medications shown to be helpful with akinesia and dyskinesia (such as amantadine). Anticholinergic medications may help with tremor, rigidity, or drooling. Surgical options for the management of Parkinson disease have received much recent attention. Tissue transplantation of embryonic stem cells in an attempt to increase the level of dopamine in the CNS is very controversial and has not been consistently shown to improve the symptoms of Parkinson disease. Deep brain stimulation involves the surgical implantation of a high-frequency thalamic electrical stimulator that interrupts the tremor-causing nerve impulses, minimizing dystonia. This can decrease the need for Parkinson medication.

KEY POINTS

- Parkinson disease may be idiopathic or a consequence of the use of certain drugs. Dopamine deficiency in the basal ganglia (substantia nigra, caudate, and putamen) is associated with symptoms of motor impairment. Difficulty initiating and controlling movements results in akinesia, tremor, and rigidity. Tremor occurs at rest, and hand tremors may be described as pill-rolling movements. Attempts to passively move the extremities are met with cogwheel rigidity. There is a general lack of movement, loss of facial expression, drooling, propulsive gait, and absent arm swing.
- Treatment is aimed at restoring brain dopamine levels or activity by administration of dopamine precursors, dopamine agonists, monoamine oxidase inhibitors, and anticholinergics. Antidepressant therapy may also help alleviate depression, and surgical procedures may be helpful for motor symptoms.

Cerebral Palsy

Etiology and Pathogenesis

Cerebral palsy refers to a diverse group of crippling syndromes that appears during childhood and involves permanent, nonprogressive damage to the developing brain. Such damage occurs during fetal development; before, during, or shortly after birth; or during early infancy. Damage occurs in the upper motor neurons that control voluntary and involuntary muscle movement. The symptoms of this damage remain for life. The majority of these children will survive until at least early adulthood. Although cerebral palsy is not considered a progressive disorder, these adolescents and adults are challenged by a variety of health problems and functional decline, including chronic pain, scoliosis, and respiratory dysfunction, along with a host of other symptoms.

Cerebral palsy is classified on the basis of neurologic signs and symptoms, with the major types involving spasticity, ataxia, or dyskinesia or a combination of these symptoms. Cerebral palsy is one of the most common crippling disorders of childhood, occurring in 0.1% to 0.2% of children and up to 15% of those born prematurely. Etiologic factors include prenatal infections or diseases of the mother; mechanical trauma to the head before, during, or after birth; or exposure to nerve-damaging poisons or a period of reduced oxygen supply to the brain. Neonatal hypoglycemia, kernicterus, prematurity, and low birth weight are also risk factors. Often the cause is multifactorial, and in many cases a single cause cannot be identified.

Clinical Manifestations

Spastic cerebral palsy manifests with hypertonia, prolonged primitive reflexes, exaggerated deep-tendon reflexes, clonus, rigidity of the extremities, scoliosis, and contractures. This type of cerebral palsy is the most common. Spastic paralysis often affects one entire side of the body (hemiplegia), both legs (paraplegia), both legs and one arm (triplegia), or all four extremities (quadriplegia). A "scissors" gait and toe walking are common. Dyskinetic/athetoid cerebral palsy manifests with extreme difficulty in purposeful movement and fine-motor coordination. Movements are slow, jerky, uncontrolled, and abrupt, resulting from injury to the basal ganglia or extrapyramidal tracts. The uncontrolled movements may increase during times of stress and disappear during sleep. Ataxic cerebral palsy is associated with gait disturbances and instability. The infant with this type of cerebral palsy

may have hypotonia at birth, but stiffness of the trunk muscles develops by late infancy. Persistence of truncal stiffness affects the child's gait and ability to maintain equilibrium. Pure ataxic cerebral palsy is rare. This palsy denotes maldevelopment of the cerebrum or its pathways, which if severe may be associated with significant cognitive impairment. More typically, a child will have a mixed disorder with clinical manifestations of each of the types.

Children with cerebral palsy often have neurologic complications such as seizures, intellectual difficulties ranging from mild to severe, and visual problems. Other associated clinical manifestations include hearing impairment, communication disorders, respiratory problems, bowel and bladder problems, and orthopedic disabilities.

Treatment

There is no cure for cerebral palsy, and the goal of management is to increase functionality. Treatment varies according to the nature and extent of brain damage. As a result of problems with muscle spasticity and contracture, muscle relaxation is a large part of therapy. Muscle relaxant medications are commonly used. Selective dorsal rhizotomy, a neurosurgical procedure in which a portion of the dorsal roots of the lumbar spine are cut, may reduce spasticity of the lower extremities. Botulinum toxin type A (Botox) is used to reduce pain and increase joint range of motion. Anticonvulsant drugs are necessary when seizures are among the symptoms of the disorder. Orthopedic surgery, casts, braces, and traction may be useful to correct some types of associated disability. A comprehensive rehabilitation program including early muscle training and special exercises may help the child with cerebral palsy lead a more productive life.

KEY POINTS

- Cerebral palsy refers to a diverse group of crippling syndromes that appear during childhood and involve permanent, nonprogressive damage to motor control areas of the brain.
- Cerebral palsy may be classified on the basis of neurologic signs and symptoms, with the major types involving spasticity, ataxia, or dyskinesia, or a mix of two or more of these three symptoms.
- Etiologic factors include prenatal infections or diseases of the mother; mechanical trauma to the head before, during, or after birth; or exposure to nerve-damaging poisons or a period of reduced oxygen supply to the brain.
- Treatment varies according to the nature and extent of brain damage. Muscle relaxants, anticonvulsant drugs, orthopedic surgery, casts, braces, and traction are among the therapies used.

Hydrocephalus

Etiology

Hydrocephalus is a condition caused by abnormal accumulation of cerebrospinal fluid (CSF) in the cerebral ventricular system. Fig. 45.6 illustrates the normal flow of CSF. Hydrocephalus is generally associated with a congenital defect, usually a neural tube defect. Viral infections or other neurotoxic agents acquired during pregnancy have been implicated with the congenital forms. It also occurs occasionally in adults and elderly persons as a consequence of mass lesions, trauma, hemorrhage, or infections such as meningitis. There are three types of hydrocephalus: (1) normal-pressure hydrocephalus, (2) obstructive/

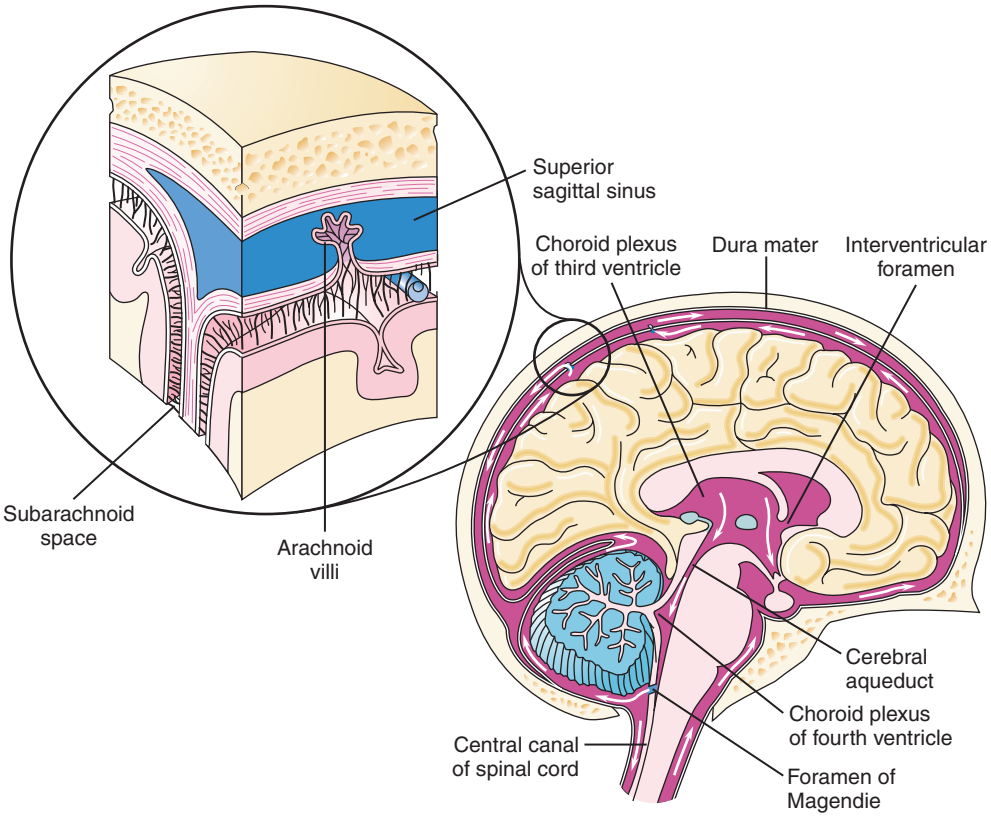


FIG 45.6 Ventricular system of the brain and distribution of cerebrospinal fluid (CSF). CSF is formed in the ventricles, passes to the subarachnoid space outside the brain and spinal cord, and moves through small valvelike structures into the large veins of the head.

noncommunicating hydrocephalus, and (3) nonobstructive/communicating hydrocephalus.

Normal-pressure hydrocephalus is a condition in which CSF volume increases without change in intracranial pressure because brain tissue has been lost. The cause of normal-pressure hydrocephalus remains unknown, but it is thought to be from an abnormality of the normal absorption of CSF. Ventricles become distended, compressing brain tissue and the cerebral vessels. There is no net change in intracranial pressure. Patients with this form of hydrocephalus demonstrate a triad of symptoms: gait instability, urinary incontinence, and dementia. If the problem is identified quickly, symptoms may improve/resolve with appropriate treatment.

Obstructive/noncommunicating hydrocephalus is most common in children and attributable to an abnormality of the cerebral aqueduct or a lesion in the fourth ventricle. The cause is usually a congenital abnormality, such as stenosis of the foramina of the fourth ventricle or spina bifida cystica.

Nonobstructive/communicating hydrocephalus (sometimes referred to as *acquired communicating hydrocephalus*) is identified by an abnormality in the capacity to absorb fluid from the subarachnoid space. There is no obstruction to the flow of fluid between the ventricles. Infections, trauma, and tumors have been identified as etiologic factors. In premature infants this usually results from an intraventricular hemorrhage.

Pathogenesis and Clinical Manifestations

Usually the obstructive type of hydrocephalus is caused by a block in the aqueduct of Sylvius, resulting from premature closure before birth in affected babies or from a brain tumor at any age (Fig. 45.7). As fluid is formed by the choroid plexus in the two lateral and the third ventricles, the volumes of these three ventricles increase greatly. This flattens the brain into a thin shell against the skull. In neonates, the increased pressure also causes the entire head to swell because the skull bones have not fused. Developmental delays may be noted.

The communicating type of hydrocephalus is usually caused by blockage of fluid flow in the subarachnoid space around the basal regions of the brain or blockage of the arachnoid villi themselves. Fluid therefore collects both inside the ventricles and on the outside of the brain. If it occurs in infants when the skull is still pliable and can be stretched, the head swells tremendously.

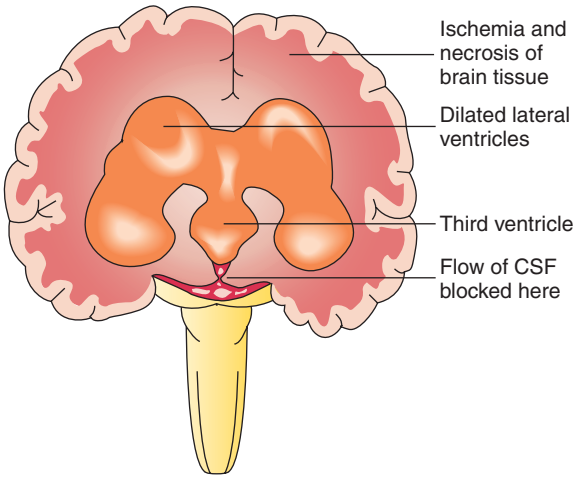


FIG 45.7 Hydrocephalus. *CSF,* Cerebrospinal fluid. (From Gould BE, Dyer R: *Pathophysiology for the health professions,* ed 5, Philadelphia, 2014, Saunders.)

Treatment

Medical treatment has been used with only limited success in controlling the secretion of CSF and relieving hydrocephalus. The most effective treatment is surgical correction employing a shunting technique. The basic components of the shunt are a ventricular catheter, a valve, and a distal catheter. Multiple perforations along the ventricular catheter permit the drainage of fluid from the ventricle. The valve is constructed so that fluid will flow in one direction only, and some valves have a pumping chamber to facilitate drainage. The distal catheter may be positioned at any of a number of sites, the most common being the peritoneal cavity (ventriculoperitoneal shunt) (Fig. 45.8). The shunt thus extends all the way from one of the ventricles to the peritoneal cavity where the fluid can then be absorbed and excreted. Another surgical approach is endoscopic third ventriculostomy. This involves making a hole in the third ventricle to allow free flow of CSF into the basal cisterns for reabsorption. This surgery is used for obstructive hydrocephalus. It should be emphasized that the correlation between degree of hydrocephalus and impaired cognitive function often results from additional complications, such as severe congenital malformations, acute or chronic infections, or progressive brain tumors.

KEY POINTS

- Hydrocephalus is a condition characterized by abnormal accumulation of CSF in the cerebral ventricular system.
- There are three types of hydrocephalus: (1) normal-pressure hydrocephalus, caused by an increased volume of CSF without a change in CSF pressure; (2) obstructive hydrocephalus, attributable to an obstruction to the flow of CSF; and (3) communicating hydrocephalus, in which absorption of CSF is abnormal.
- The most effective treatment for hydrocephalus is surgical correction employing a shunt.

Cerebellar Disorders

The cerebellum performs three general functions in the control of skeletal muscles: (1) together with activity of the cerebral cortex, it coordinates the activities of muscle groups to produce skilled movement; (2) it functions below the level of consciousness to maintain posture and make movements smooth, steady, efficient, and coordinated; and (3) it controls skeletal muscles to maintain balance (see Chapter 43). Fig. 45.9 illustrates the cerebrum and cerebellum working together to coordinate muscle movement. Impulses from the motor control areas of the cerebrum travel down the corticospinal tract and through peripheral nerves to skeletal muscle tissue. Simultaneously, the impulses go to the cerebellum. The cerebellum compares the motor commands of the cerebrum with information coming from receptors in the muscle. In effect, the cerebellum compares the intended movement with the actual movement. Impulses then travel from the cerebellum to both the cerebrum and the muscle tissue to adjust or coordinate the movements to produce the intended action.

Etiology and Clinical Manifestations

Cerebellar disorders may have myriad causes. Abscess, hemorrhage, tumors, trauma, viral infection, and chronic alcoholism have been implicated. Identification and eradication of the causal agent determine treatment and prognosis. The clinical manifestations of cerebellar disorders primarily include ataxia (muscle incoordination), hypotonia, intention tremors, and disturbances of gait and balance. Disturbances of gait and balance vary, depending on the muscle groups involved. The

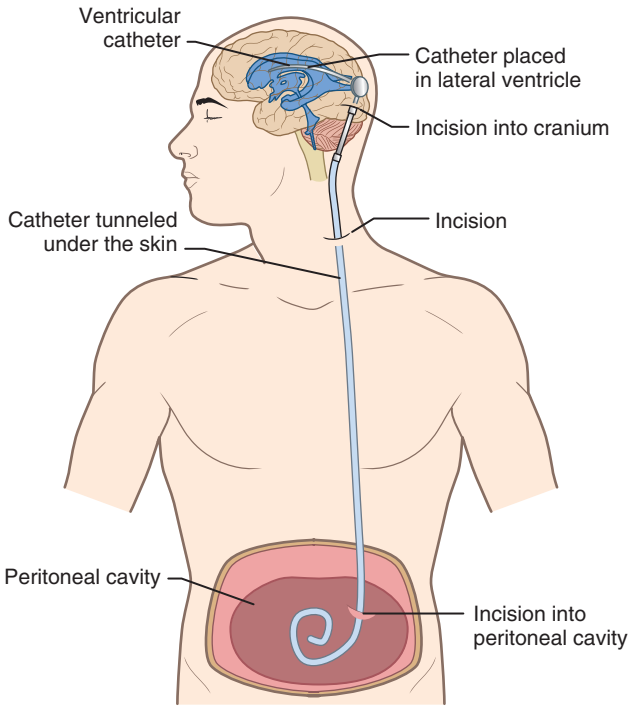


FIG 45.8 Ventriculoperitoneal shunt placed for chronic hydrocephalus.

walk, for instance, is often characterized by staggering or lurching and by a clumsy manner of raising the foot too high and bringing it down with a clap. Loss of cerebellar function does not result in paralysis.

KEY POINTS

- The cerebellum is responsible for coordinated control of muscle action, excitation and inhibition of postural reflexes, and maintenance of balance.
- Etiologic factors in cerebellar disorders may include the following: abscess, hemorrhage, tumors, trauma, viral infection, or chronic alcoholism.
- Clinical manifestations of cerebellar disorders primarily include ataxia, hypotonia, intention tremors, and disturbances of gait and balance.

SPINAL CORD AND PERIPHERAL NERVE DISORDERS

Multiple Sclerosis

Etiology

Multiple sclerosis (MS) is a chronic demyelinating disease of the CNS that causes significant disability in young adults. It is thought to be an autoimmune disorder that results in inflammation and scarring (sclerosis) of the myelin sheaths covering nerves. It is estimated that more than 2.3 million people are affected by MS worldwide. The age of onset ranges from 20 to 50 years, and MS is two to three times more common in women than in men. Epidemiologic studies show that MS occurs at a higher rate among individuals from Caucasian Northern European descent and those who live in northern latitudes. Several studies indicate that those who were born and spent the early years of life (first 15 years) in northern areas carry an increased risk of MS even if they migrate south at some time later in their lives.

MS is an unpredictable disease with a wide variety of clinical presentations. Symptoms can vary daily, and the disease may cause only mild disability with occasional exacerbations. In some individuals, however, MS may cause extreme, progressive disability. Despite great advances in research, the exact cause of MS is unknown. Genetics may have a role. Nontwin first-degree relatives have a 2.5% to 5% higher risk of developing MS.

Pathogenesis

In MS, the demyelination of nerves can occur anywhere in the CNS. There does not seem to be any predictable pattern in the timing or location of the lesions. However, structures most frequently affected are the optic nerves; the oculomotor nerves; and the corticospinal, cerebellar, and posterior column systems. Fig. 45.10 illustrates demyelination. Myelin facilitates nerve conduction; the inflammation and scarring that occur with MS slow or interrupt the conduction of nerve impulses. The triggering event for this process is not understood. It is theorized that an exposure to a viral infection or environmental toxin initiates the autoimmune attack in a genetically predisposed individual. Both humoral and cellular immune factors have been implicated in demyelination. Antibodies to specific myelin proteins have been found in both the serum and the CSF of MS patients. T-cell lymphocyte–mediated injury to the myelin has also been implicated in causing the autoimmune damage and sustaining inflammation. Tumor necrosis factor α, an important proinflammatory cytokine, is elevated in active lesions, serum, and CSF of MS patients.

Clinical Manifestations and Treatment

Symptoms of MS vary widely and depend on the location of damage to the myelin. They include impaired visual acuity or blurred vision, diplopia, weakness, numbness, tingling, extreme fatigue, imbalance, movement disorders, spasticity, coordination difficulties, and gait disturbance. Bladder and/or bowel difficulties, vertigo, pain, and paresthesia may also be present. Neurobehavioral symptoms may include depression, emotional lability, and sexual dysfunction, as well as memory and cognitive impairment (Table 45.1). In later stages of the disease, spastic paralysis of the limbs may be present because of upper motor neuron damage. Symptoms may be exacerbated by heat, infection, trauma, and stress. Relapses are also common in the postpartum period after pregnancy. There are four main categories to classify the clinical course of MS (Table 45.2).

There is no conclusive diagnostic test for MS. The diagnosis is based on clinical characteristics, imaging studies, and laboratory evidence. The current diagnostic criteria for MS require documentation of two or more episodes of symptoms and two or more signs that reflect pathology in anatomically different areas of the CNS. Advances in neuroimaging techniques have become quite useful in the diagnosis of MS. MRI of the brain and spinal cord may show the presence of demyelination (plaques). CNS lesions that are disseminated in time and space with no better explanation is one diagnostic criterion for MS. Evoked potential recording of nerve stimulation in the visual and other nerve pathways may be helpful. Laboratory tests may show mild lymphocytosis and elevated serum protein levels, especially after an acute relapse. Elevated levels of immunoglobulin G (IgG) in the CSF with the presence of discrete bands of IgG (oligoclonal bands) may also be present.

There is no cure for MS. Treatment centers not only on managing the symptoms of the disease, but also on minimizing the damage inflicted by the autoimmune attack on myelin. Treatment for MS can be divided into three main categories: treatment for acute attacks, use of disease-modifying agents, and symptom management. Acute attacks are considered a sudden worsening of symptoms not thought to be related to heat, infection, or fever and are managed with corticosteroids such as prednisone. These are used to reduce edema and the inflammatory

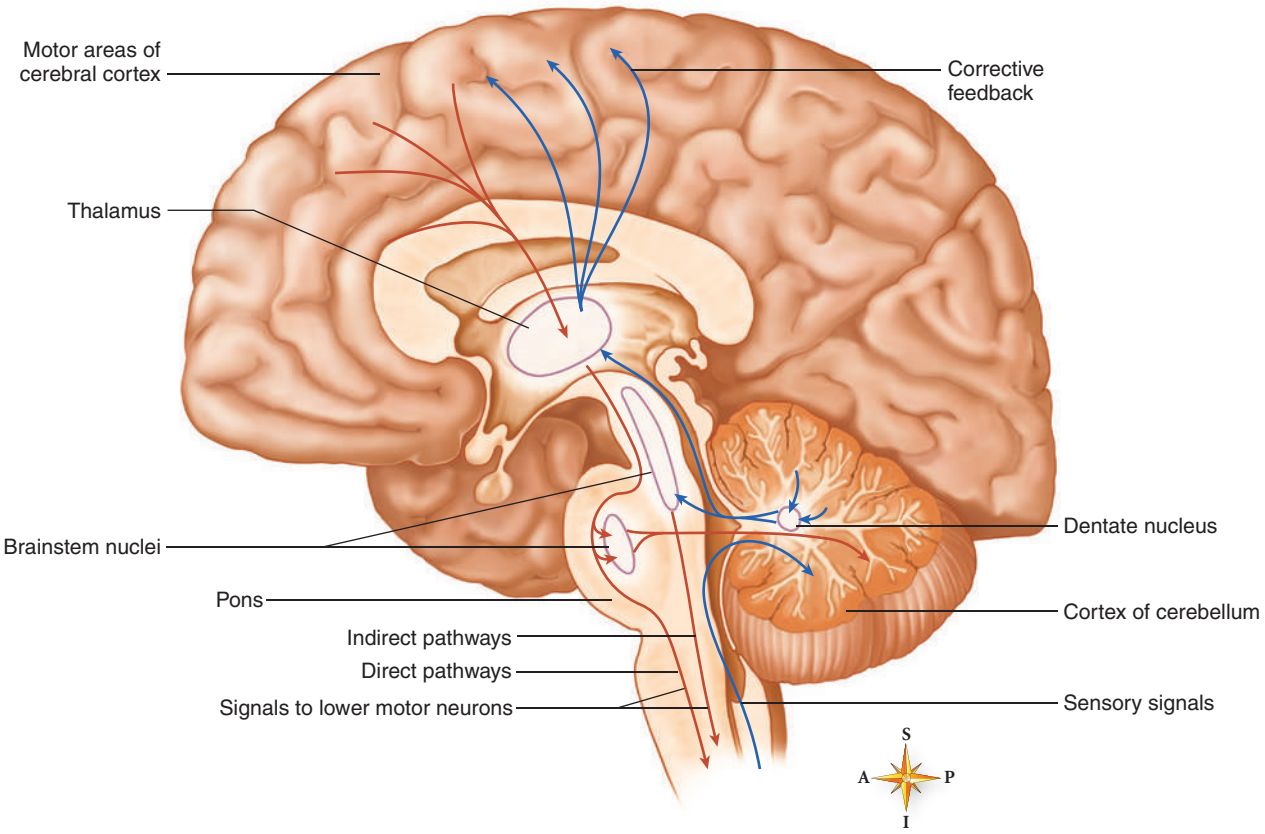


FIG 45.9 Coordinating function of the cerebellum. Impulses from the motor control areas of the cerebrum travel simultaneously to skeletal muscle tissue and to the cerebellum. The cerebellum, which also receives and evaluates sensory information, compares the intended movement with the actual movement. It then sends impulses to both the cerebrum and the muscles, thus coordinating and smoothing muscle activity. (From Patton K, Thibodeau G: *Essentials of anatomy & physiology,* ed 1, St Louis, 2012, Mosby.)

response. Recovery may be hastened by the use of these agents; however, the extent of recovery is unchanged. Disease-modifying therapy is aimed at reducing autoimmune destructive activity. These agents reduce the activity of immune cells or inflammatory cytokines.

Management of symptoms frequently requires participation from multiple disciplines, including medicine, nursing, speech pathology, neuropsychiatry, social services, and vocational services. Treatment with an array of medications such as antispasmodics, anticholinergics, antidepressants, and antimicrobials helps to manage symptoms. Treatment also includes avoidance of complications such as urinary tract infections, constipation/impactions, respiratory tract infections, and pressure sores.

Research in MS continues to examine the immune system role. Viruses such as the Epstein-Barr virus have been implicated. Low levels of vitamin D have also been implicated as an associated risk factor, because this is a common finding in people living in higher latitudes.

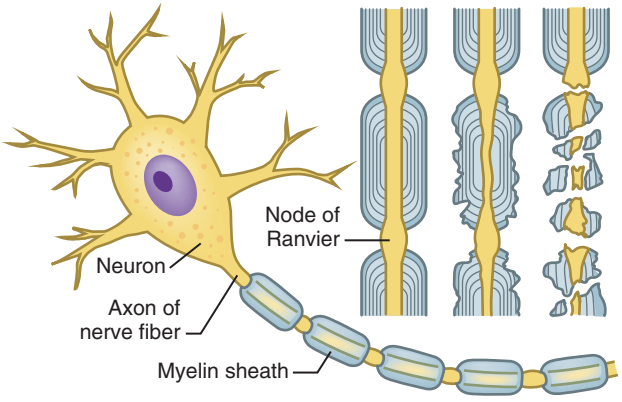


FIG 45.10 Changes in the nerve sheath as seen in multiple sclerosis. Myelin is made by oligodendrocytes and coats nerves, facilitating nervous impulse. In patients with multiple sclerosis, the myelin degenerates in patches, causing nerve transmission to become erratic.

KEY POINTS

- MS is a demyelinating disease of the CNS that primarily affects young adults. The risk of contracting MS is greater for persons living in higher latitudes. The cause of MS is unknown, but immunologic abnormalities and environmental factors are suspected.
- Demyelination can occur throughout the CNS but most frequently affects the optic and oculomotor nerves and spinal nerve tracts.
- In most cases symptoms are slowly progressive, and the disease is marked by exacerbations and remissions.
- Symptoms include double vision, weakness, poor coordination, and sensory deficits. Bowel and bladder control may be lost. Memory impairment is common.
- Management is symptomatic. Short-term steroid therapy may be helpful during acute exacerbations, and immune-modifying drugs may slow the progression of symptoms.

Spina Bifida

Etiology and Pathogenesis

Spina bifida is a developmental anomaly characterized by defective closure of the bony encasement of the spinal cord (neural tube) through which the spinal cord and meninges may or may not protrude. If the anomaly is not visible, the condition is called *spina bifida occulta.* If there is an external protrusion of the saclike structure, the condition is called *spina bifida cystica* and is further classified according to the extent of neural involvement (e.g., meningocele, meningomyelocele, or myelomeningocele) (Fig. 45.11).

Both environmental factors and genetics appear to be a factor in the etiologic development of neural tube defects. These include vitamin deficiency (folate), anticonvulsant drugs, and chromosomal abnormalities. Maternal obesity and diabetes are also risk factors. Changes in anticonvulsant drug therapy may be recommended for pregnant women or for those considering pregnancy. Supplementation of folic acid before conception and during pregnancy also appears to decrease the prevalence of neural tube defects.

Clinical Manifestations

In spina bifida occulta, the posterior vertebral laminae have failed to fuse. The defect is extremely common and occurs to some degree in 10% to 20% of the population. The vast majority of these vertebral defects are located in the lumbosacral regions, most commonly in the fifth lumbar vertebra and the first sacral vertebra, and may be detected prenatally through ultrasound and α-fetoprotein testing.

Spina bifida occulta may be manifested by changes in the skin and body hair: either very coarse or silky hair along the spine; a midline dimple, with or without a sinus tract; a cutaneous port-wine angioma; and/or a subcutaneous mass typically representing a lipoma or dermoid cyst. Spina bifida occulta usually causes no serious neurologic problems. Common lumbosacral defects can cause gait disturbances, positional deformities of the feet, or bladder/bowel dysfunction. These symptoms become evident in childhood during periods of rapid growth.

In the meningocele form of spina bifida cystica, a saclike cyst filled with CSF protrudes through the spinal defect but does not involve the spinal cord. Meningoceles occur with equal frequency in the cervical, thoracic, and lumbar areas. A myelomeningocele or meningomyelocele deformity contains meninges, CSF, and a portion of the spinal cord that protrude from the vertebral defect in a cystlike sac. These defects most often occur in the lumbar or lumbosacral region of the spine, because these are the last areas of the neural tube to close during fetal development. These defects may be detected in prenatal ultrasound and with α-fetoprotein testing. The bony prominences of the unfused neural arches are palpable at the lateral borders of the defect. The sac includes a transparent membranous covering that may have neural tissue attached to its inner surface. This membrane may be intact at birth or leak CSF, thereby increasing the risk of infection and neural damage. These infants are delivered via cesarean section to decrease the trauma to the exposed neural tissue, and surgical closure is attempted soon after delivery. Affected infants often suffer from permanent neurologic damage resulting in motor weakness or paralysis and sensory deficit below the level of the spinal defect, bowel and bladder dysfunction, scoliosis, hydrocephalus, and seizures. Often the problems worsen as the child grows and the cord ascends within the vertebral canal, pulling primary scar tissue and thereby tethering the cord. Surgical closure of the defect in utero before delivery may be attempted in some cases.

Treatment. Treatment for this common disorder is based on the severity of the defect and neurologic dysfunction. Supplementation with folic acid during the period before conception has been shown to significantly decrease the risk of having a child with a neural tube defect. It is recommended that all women of childbearing age take folic acid daily for prevention. In the United States common foods are supplemented with folic acid to decrease the incidence of this deformity.

TABLE 45.1 Multiple Sclerosis Symptoms

Area of Dysfunction	Symptoms
Cranial nerve dysfunction	Blurred central vision, faded colors, blind spots (optic neuritis)
	Diplopia
	Dysphagia
	Facial weakness, numbness, pain
Motor dysfunction	Weakness
	Paralysis
	Spasticity
	Abnormal gait
Sensory dysfunction	Paresthesias
	Lhermitte sign (electric shock–like sensation radiating down spine into extremities)
	Decreased proprioception
	Decreased temperature perception
Cerebellar dysfunction	Dysarthria
	Tremor
	Incoordination
	Ataxia
	Vertigo
Bowel and bladder dysfunction	Fecal urgency, constipation, incontinence
	Urinary frequency, urgency, hesitancy, nocturia, retention, incontinence
Cognitive dysfunction	Decreased short-term memory
	Difficulty learning new information
	Word-finding trouble
	Short attention span
	Decreased concentration
	Mood alterations (depression, euphoria)
Sexual dysfunction	Women: decreased libido, decreased orgasmic ability, decreased genital sensation
	Men: erectile, orgasmic, and ejaculatory dysfunction
Fatigue	Overwhelming weakness not overcome with increased physical effort

Adapted from Ropper AH, Samuels MA, Klein JP: Chapter 36. Multiple sclerosis and other inflammatory demyelinating diseases. In: Ropper AH, Samuels MA, Klein JP, editors: *Adams & Victor's principles of neurology,* 10e. New York, 2014, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid = 690&Sectionid = 50910887. Accessed January 3, 2016.

Hauser SL, Goodin DS: Multiple sclerosis and other demyelinating diseases. In: Kasper D, Fauci A, Hauser S, et al., editors: *Harrison's principles of internal medicine,* 19e, New York, 2015, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid = 1130&Sectionid = 79756278. Accessed May 22, 2016.

TABLE 45.2 Clinical Course of Multiple Sclerosis

Course	Characteristics
Relapsing-Remitting	
Most common form; approximately 85% of MS patients have this form	Clearly defined exacerbations (relapses) with acute decline in neurologic function; followed by periods of partial/complete recovery and remissions; remissions may last months to years
Primary-Progressive	
Relatively rare; approximately 10% of MS patients have this form	Slow but almost continuous decline in neurologic function; plateaus or temporary minor improvements may occur; relapses/remissions not present; severe disability develops early
Secondary-Progressive	
Before use of disease-modifying drugs; approximately 50% of relapsing-remitting patients develop this form	Begins as relapsing-remitting; followed by steady decline in neurologic function, with or without occasional relapses, remissions, or plateaus
Progressive-Relapsing	
Relatively rare; approximately 5% of patients have this form	Progressive from outset, but with clear exacerbations, with or without recovery

Hauser SL, Goodin DS: Multiple sclerosis and other demyelinating diseases. In: Kasper D, Fauci A, Hauser S, et al., editors: *Harrison's principles of internal medicine,* 19e, New York, 2015, McGraw-Hill.

KEY POINTS

- Spina bifida is a developmental anomaly characterized by defective closure of the bony encasement of the spinal cord (neural tube) through which the spinal cord and meninges may or may not protrude.
- If the anomaly is not visible, the condition is called *spina bifida occulta.* If there is an external protrusion of the saclike structure, the condition is called *spina bifida cystica* and is further classified according to the extent of neural involvement (e.g., meningocele, myelomeningocele).
- The natural history of myelomeningocele supports an early and aggressive operative approach before significant clinical deterioration begins. A cesarean section before rupture of amniotic membranes and onset of labor may decrease the degree of paralysis.
- Folic acid supplementation taken before conception and during pregnancy appears to decrease the prevalence of neural tube defects.

Amyotrophic Lateral Sclerosis

Etiology and Pathogenesis

Amyotrophic lateral sclerosis (ALS) is a progressive degenerative disease affecting both the upper and the lower motor neurons characterized by muscle wasting and atrophy of the hands, arms, and legs and in some cases cognitive changes and dementia. ALS leads to progressive paralysis and death, usually within a few years of onset. ALS is also known as *Lou Gehrig disease* after the famed "Iron Man" of the New York Yankees who died from the disease. ALS affects an estimated 12,000 to 15,000 Americans at any given time. It most commonly strikes between the ages of 50 and 75 with a higher incidence in men than women. The majority of ALS cases occur at random; however, 5% to 10% of cases are familial. Although ALS is classified as a single disease entity, emerging evidence suggests that it is a clinical syndrome with several possible causes. Smoking, increasing age, and male gender are the only risk factors identified at this point.

Several genes have been implicated in ALS familial cases; however, in the vast majority of cases, there is no identifiable genetic cause. The pathologic hallmark of ALS is the presence of misfolded protein aggregate inclusions in affected motor neurons and glial cells. Similar to other neurodegenerative diseases such as Alzheimer and Parkinson diseases and MS, neuroinflammation is a prominent feature in the process of neuronal loss in ALS. Other proposed mechanisms include mitochondrial dysfunction, defective axonal transport, and axonal retraction. The death of peripheral motor neurons in the brainstem and spinal cord leads to denervation and subsequent atrophy of the corresponding muscle fibers.

Clinical Manifestations and Treatment

Most patients with ALS demonstrate muscle weakness and atrophy. The earliest symptoms may be muscle twitching, cramping, and stiffness. Often the hands or upper extremities are affected first. The weakness is progressive and eventually affects the muscles that control speech, swallowing, and breathing. Finding hyperreflexia in a weak atrophied extremity is highly suggestive of ALS. Most individuals die of respiratory failure within 3 to 5 years from the onset of symptoms. Despite the marked physical disability, most patients maintain their sensory and cognitive functions.

ALS is a diagnosis of exclusion, based on the patient's clinical signs and symptoms. Electromyography (EMG), nerve conduction studies, MRI, and serum laboratory testing may be used to rule out other causes of weakness, such as MS, brain and spinal tumors, HIV, and Lyme disease.

One FDA-approved treatment for ALS is riluzole (Rilutek), a glutamate inhibitor. This medication is not a cure, but its use can prolong life for several months and may delay the need for mechanical ventilation. Patients with ALS benefit from a multidisciplinary approach to care to prevent complications from immobility, as well as to address both physical and psychological needs.

KEY POINTS

- ALS is a progressive disease affecting both the upper and the lower motor neurons. The cause of ALS remains unknown. Weakness and wasting of the upper extremities usually occur, followed by impaired speech, swallowing, and respiration.
- ALS usually strikes between the ages of 50 and 75, and it is more common in men than in women. The mean survival time is about 3 years from the time of diagnosis.
- Clinical manifestations include weakness, atrophy, cramps, stiffness, and irregular twitching of muscle fibers.
- Diagnosis is based on clinical signs and symptoms, EMG results, nerve conduction studies, MRI studies, and serum laboratory testing.
- Riluzole (Rilutek) is a glutamate inhibitor, which may be helpful in management of ALS.

Spinal Cord Injury

Spinal cord injuries are among the most devastating and costly problems faced by patients and their families. Marked changes in lifestyle are required for survivors. Medical advances in the emergent management of spinal cord injuries and their associated complications have been

FIG 45.11 Photographs of infants with spina bifida cystica. **A,** Spina bifida with meningomyelocele in the lumbar region. **B,** Spina bifida with myeloschisis in the lumbar region. Note that the nerve involvement has affected the lower limbs. (In Moore KL, Persaud TVN: *The developing human: clinically oriented embryology,* ed 8, Philadelphia, 2008, Saunders. Courtesy Dr. Dwight Parkinson, Department of Surgery and Department of Human Anatomy and Cell Science, University of Manitoba, Winnipeg, Manitoba, Canada.)

responsible for increasing survival rates. Continuing research is focused on minimizing the incidence of injury and the mortality/morbidity of spinal cord injury.

Etiology

Spinal cord injury is primarily a problem of the young. Males are three to four times more likely to have suffered a spinal cord injury, and these injuries are most common on the weekends and during the summer months. Motor vehicle crashes contribute the highest number of spinal cord injuries, followed by violence (primarily gunshot wounds), falls, and recreational accidents. Alcohol use and risk-taking behavior are often involved. Other causes of spinal cord injuries include birth injuries, herniated intravertebral disk, or bone spurs related to degenerative changes of aging and osteoporosis. Injuries to the spinal cord are classified by level, degree (complete or incomplete), and mechanism of injury (Box 45.2).

Pathogenesis

Spinal cord injury results from compression (tumor, hematoma, or bony encroachment) and from blunt trauma causing contusion or penetration/transection of neural tissue. The major mechanisms of injury are hyperflexion, hyperextension, and compression (Fig. 45.12). Flexion injury with tearing of the posterior ligaments and dislocation is the most unstable injury and is often associated with severe neurologic deficits. Hyperextension injury is the most common.

BOX 45.2 American Spinal Injury Association Impairment Scale

A = Complete: No motor or sensory function is preserved in the sacral segments S4 to S5.
B = Incomplete: Sensory function (but not motor function) is preserved below the neurologic level and includes the sacral segments S4 to S5.
C = Incomplete: Motor function is preserved below the neurologic level, and more than half of key muscles below the neurologic level have a muscle grade less than 3.
D = Incomplete: Motor function is preserved below the neurologic level, and at least half of key muscles below the neurologic level have a muscle grade of 3 or more.
E = Normal: Motor and sensory function are normal.

Standards for Neurologic Classification of SCI Worksheet, 2013. From American Spinal Injury Association. Available at http://www.asia-spinalinjury.org/elearning/ASIA_ISCOS_high.pdf

Spinal cord injury can be described as occurring in three phases/stages. The acute phase begins with the mechanical trauma to the cord. There is axonal disruption and neuronal death. Blood flow is disrupted, causing ischemia. Cytotoxic edema, invasion of granulocytes, disruption of ionic balance, and neurotransmitter release with resultant excitotoxicity

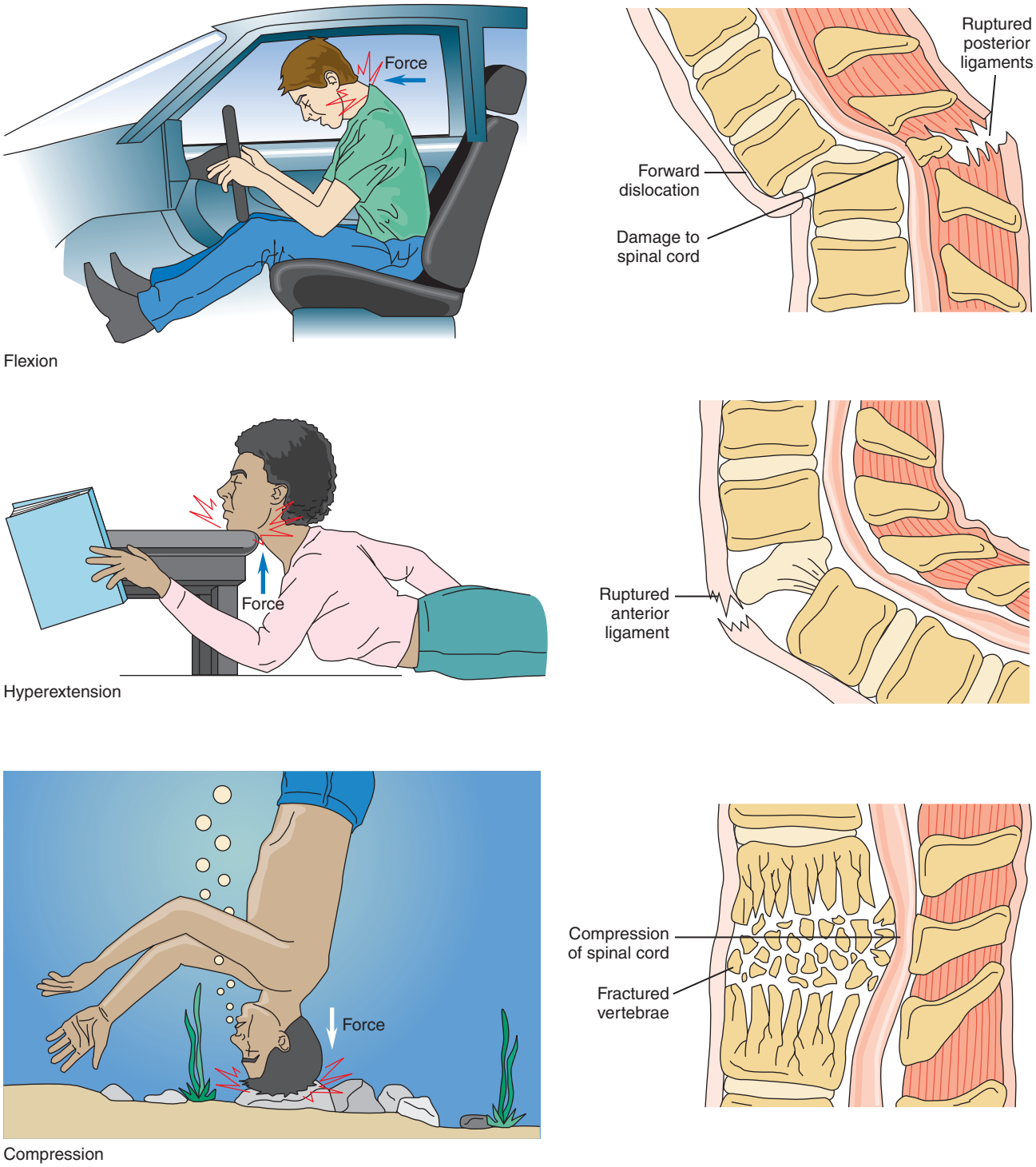


FIG 45.12 Mechanisms of spinal cord injury. Many situations may produce these consequences. This figure shows examples only. **A,** Flexion injury of the cervical spine ruptures the posterior ligaments. **B,** Hyperextension injury of the cervical spine ruptures the anterior ligaments. **C,** Compression fractures crush the vertebrae and force bony fragments into the spinal canal.

occur along with the formation of free radicals. All of these result in additional damage to the spinal cord.

The subacute or intermediate phase starts about 7 days after the initial injury and is characterized by additional oxidative stress, lipid peroxidation, and free radial production. There is macrophage and lymphocyte infiltration with the subsequent increase in cytokines and increased inflammation. This creates a progressive form of neurodegeneration that exacerbates neuronal apoptosis beyond the site of the primary mechanical trauma.

The chronic phase begins after a few weeks to months and can persist for years. During this stage, there is apoptosis of oligodendrocytes and consequent demyelination, cavity, and astroglial scar formation. These events are detrimental to axonal regrowth and recovery.

The systemic hemodynamic changes that occur after spinal cord injury are a major factor in the resulting damage to the spinal cord. Because of the injury to the spinal cord, autoregulation is lost, resulting in a profound drop in systemic blood pressure. This adds to the ischemia of the tissue. In addition, spinal cord injuries are often accompanied by trauma to other organ tissue causing hypoxia, hypotension, hyperthermia/hypothermia, and hypoglycemia/hyperglycemia.

Clinical Manifestations

Immediately after injury to the spinal cord, there is complete loss of function below the level of injury. (This may occur even in incomplete injuries to the spinal cord, causing the injury to appear more severe than it actually is.) This phenomenon, known as *spinal shock,* can last from a few hours to a few weeks. Symptoms below the level of injury include flaccid paralysis of all skeletal muscles; loss of all spinal reflexes; loss of pain, proprioception, and other sensations; bowel and bladder dysfunction with paralytic ileus; and loss of thermoregulation. A return of spinal reflexes indicates the end of spinal shock. As reflex function returns, spastic paraplegia or quadriplegia develops with hyperreflexia and extensor plantar responses, but a flaccid atrophic (lower motor neuron) paralysis may be found depending on the segments of the cord affected. The bladder and bowel may regain some reflex function.

In patients with cervical or upper thoracic cord injury, neurogenic shock is a life-threatening complication. Neurogenic shock is a form of distributive shock caused by the loss of brainstem and higher center control of the sympathetic nervous system. The loss of sympathetic outflow results in hypotension caused by peripheral vasodilation. Bradycardia occurs (secondary to the overriding parasympathetic influence), and there is a loss of the cardiac accelerator reflex. The loss of impulses from the temperature regulatory center in the brain prevents the ability to sweat below the level of injury.

A chronic, ongoing complication of spinal cord injuries occurring at or above the T6 vertebra is autonomic dysreflexia. This is a potentially life-threatening complication that may occur any time after spinal shock has resolved. It is characterized by a sudden episode of hypertension, headache, bradycardia, upper body flushing, and lower body vasoconstriction. Piloerection (goose bumps) and sweating also occur. The usual stimulus initiating autonomic dysreflexia is activation of visceral or cutaneous pain receptors below the level of injury. A full bladder or constipation is a common cause.

Stimulation of afferent pain receptors causes activation of sympathetic efferents in the cord and reflex vasoconstriction. Sustained activation of sympathetic neurons below the level of cord injury increases blood pressure significantly. The hypertension initiates the baroreceptor response. Baroreceptors mediate inhibition of heart rate and vasodilation of vessels above the level of injury. This is responsible for the upper body flushing. Descending signals from the brain cannot pass the cord injury, so inhibition of sympathetic neurons below the level of injury does not occur. Blood pressure may be dangerously high and may require aggressive treatment.

Treatment

Management of spinal cord injuries includes appropriate stabilization of spinal vertebra components to prevent further trauma to the spinal cord. This may be accomplished surgically with internal fixation or with external fixation and bracing. If there is cord compression from bony elements, this should be addressed surgically within 24 hours. During neurogenic shock, patients require intensive care to maintain oxygenation and blood pressure. The use of high-dose methylprednisolone initiated within the first 8 hours after injury may preserve some function by decreasing the secondary injury to the spinal cord. The benefit of this medication is modest at best, and its use is controversial, but it offers some hope of improving function. Ongoing assessment is critical. Methodical neurologic evaluations are important in determining improvement or deterioration in function. Treatment for autonomic dysreflexia includes removing or alleviating the painful stimulus, and in certain situations the use of adrenergic receptor–blocking medications to manage the hypertensive crisis.

Individuals suffering from spinal cord injuries have chronic and ongoing problems with spasticity and contracture related to upper motor neuron damage. They are also at high risk for respiratory and urinary tract infections, skin pressure sores, septicemia, and fecal impaction. Much of the care of patients with spinal cord injuries is aimed at preventing these complications and maximizing function. The rehabilitation phase for these patients is lengthy with emphasis on independence and self-care. Ongoing care of patients with spinal cord injuries is multidisciplinary and should also address the psychosocial impact of this life-changing event. Levels of injury and expected functional ability are summarized in Table 45.3.

KEY POINTS

- Spinal cord injury is usually traumatic, a result of motor vehicle accidents, falls, penetrating wounds, or sports injuries. The cord may be compressed, transected, or contused. Further injury may result from hemorrhage, swelling, and ischemia after injury.
- Spinal shock occurs immediately after injury and is characterized by temporary loss of reflexes below the level of injury. Muscles are flaccid, and skeletal and autonomic reflexes are lost. The end of spinal shock is noted when reflexes return and flaccidity is replaced by spasticity.
- Neurogenic shock may occur after spinal cord injury due to peripheral vasodilation. Hypotension and circulatory collapse may occur. High spinal cord injuries may also affect respiratory muscles, leading to ventilatory failure.
- Autonomic dysreflexia is an acute reflexive response to sympathetic activation below the level of injury. Visceral stimulation (full bladder or bowel) and activation of pain receptors below the injury are common initiating stimuli. Manifestations include hypertension, bradycardia, flushing above the level of injury, and clammy skin below the level of injury. Prompt removal of the offending stimulus is indicated.

Guillain-Barré Syndrome

Etiology and Pathogenesis

Guillain-Barré syndrome, also known as *acute idiopathic polyneuropathy* or *polyradiculoneuropathy,* is an inflammatory demyelinating disease of the *peripheral* nervous system or a lower motor neuron disorder. Between 1 and 4 cases per 100,000 individuals occur annually, with an increasing incidence in the aging population. It is one of the most common causes of nontraumatic paralysis in the Western world. There is a slight male preponderance.

The cause of Guillain-Barré syndrome is not well understood, but it sometimes follows an infection, inoculation, or surgical procedure 1 to 8 weeks before the onset of signs and symptoms. *Campylobacter jejuni* enteritis has been associated with the syndrome. The basis for Guillain-Barré syndrome is immunologic, but the exact mechanism is unknown. Both humoral and cellular immune responses have been implicated in Guillain-Barré syndrome. Autoantibodies and activated lymphocytes are thought to be involved in the segmental demyelination that occurs. Motor neurons are primarily affected, but sensory nerves may also be involved.

Clinical Manifestations and Treatment

Patients with Guillain-Barré syndrome have progressive ascending weakness or paralysis. It usually begins in the legs, spreading often to

TABLE 45.3 Levels of Injury and Expected Functional Ability for Patients With Spinal Cord Injury

Level	Normal Activity	Functional Expectation
C4	Head control Mouth control Shoulder/scapular movement Diaphragm movement	Can use adaptive devices (i.e., mouth stick) for phone, reading, computer Total dependence for transfers/ADLs Pulmonary hygiene concerns; skin-care issues
C5	Shoulder flexion Elbow flexion Increased scapular motion	Can use adaptive devices for self-feeding; can move wheelchair short distances (electric wheelchair preferred); can perform ADLs and bed mobility with assistance; needs pulmonary hygiene assistance
C6	Good elbow flexion Wrist extension Shoulder rotation and abduction	Independent with grooming/feeding with adaptive devices; weak hand grasp; can roll over in bed; can drive with car adaptations; can transfer with assistance; can self-propel wheelchair
C7	Elbow extension Strong wrist extension Good shoulder movement	Can transfer to wheelchair independently; can perform most ADLs independently; excellent bed mobility
T1	Normal hand strength Normal upper body strength	Bed and wheelchair independent; can perform self-catheterization
T10	Normal strength/motion above umbilicus	May stand for exercise with braces; still wheelchair dependent for ambulation
L2–L5	Some leg and thigh movement	Can ambulate indoors with braces/canes
Sacral segments	Mild weakness in lower extremities	Can ambulate with braces/canes; still significant bowel/bladder dysfunction

ADLs, Activities of daily living.

the arms and face. The respiratory muscles may also be affected. The severity and extent of neurologic deficit may vary greatly among patients. Most patients reach the peak of disability in 10 to 14 days. Sensory nerves are affected to a lesser extent than motor neurons. Patients may experience paresthesia or dysesthesia; neuropathic pain may also be present. During this time, patients may demonstrate loss of autonomic regulation, with consequent changes in blood pressure and heart rhythm, and may require intensive care for ventilatory and circulatory support.

A diagnosis of Guillain-Barré syndrome is made through patient history, physical examination, and nerve conduction studies. The CSF characteristically contains high protein concentrations. Other laboratory studies and imaging are used to rule out other causes for neurologic dysfunction.

The majority of patients experience spontaneous recovery; however, approximately 10% of patients may be left with varying levels of disability. Gradually, neurologic function returns, often in a descending pattern, with upper extremities recovering earlier than lower extremities. Treatment within 14 days of onset of symptoms with plasmapheresis, especially in those with severe or rapidly progressing symptoms, has been shown to have some value. Intravenous immunoglobulin is also helpful. However, nearly 5% to 10% of patients may experience one or more exacerbations of symptoms. Nursing care of these patients is aimed at preventing complications of immobility.

KEY POINTS

- Guillain-Barré syndrome is characterized by muscle weakness that begins in the lower extremities and spreads to the proximal spinal neurons.
- The cause is unknown; however, a postinfectious immunologic mechanism is suspected.
- Treatment is supportive, and spontaneous recovery usually occurs.

Bell Palsy

Etiology and Pathogenesis

Bell palsy is an acute idiopathic paresis or paralysis of the facial nerve involving an inflammatory reaction at or near the stylomastoid foramen or in the bony facial canal with probable occurrence of compression, ischemia, and demyelination. The incidence peaks between the ages of 15 and 50 years. Risk factors include diabetes, hypothyroidism, and pregnancy. There is evidence that Bell palsy is caused by a viral infection. Antibodies to the herpes simplex and herpes zoster viruses have been found in patients with Bell palsy.

Clinical Manifestations and Treatment

Symptoms of Bell palsy develop rapidly over 24 to 48 hours. Physical examination shows unilateral facial weakness with facial droop and diminished eye blink, hyperacusis, and decreased lacrimation (Fig. 45.13). Patients may complain of a heavy sensation in their face as well as a decreased sense of taste, but sensation of the face is generally intact. Posterior auricular pain may be present. In the diagnosis of Bell palsy, other causes of facial paralysis, such as bacterial infection (otitis media), tumor, trauma, and cerebrovascular accident (stroke), must be ruled out. MRI, CT, and EMG can be helpful in certain situations. Laboratory testing is of limited value.

Management of Bell palsy is controversial. Most patients recover facial nerve function spontaneously within approximately 3 weeks. However, approximately 30% of patients are left with some level of residual disability. Prevention of corneal damage resulting from the inability of the eye to close is vital. Lubricating drops, ointments, and nighttime eye patching may be necessary. The use of corticosteroids has been shown to improve the likelihood of complete recovery. Because of the association of viruses with Bell palsy, the use of antiviral medications such as acyclovir or valacyclovir has previously been recommended. However, recent studies do not support their use. Surgical decompression of the nerve has not been shown to confer great benefit.

KEY POINTS

- Bell palsy, or neuropathy of the facial nerve, results in paralysis of the muscles on one side of the face. Often a self-limiting condition with unknown cause, Bell palsy may last only a few days or weeks.
- Treatment is supportive, and spontaneous recovery usually occurs.

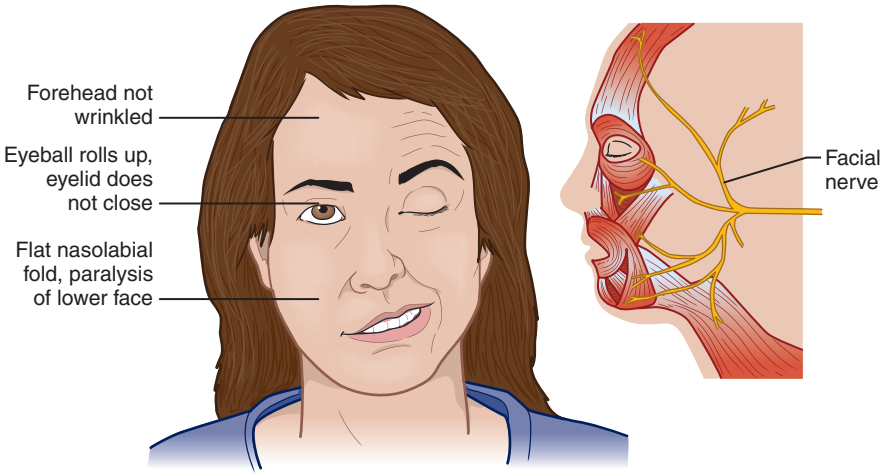


FIG 45.13 Bell palsy. Locations of the branches of the facial nerve (cranial nerve VII) correspond to the areas of peripheral facial paralysis. (From Black JM, Hawks JH: *Medical-surgical nursing: clinical management for positive outcomes,* ed 8, Philadelphia, 2008, Saunders, p 1886.)

SUMMARY

A traumatic event, such as a spinal cord injury, or a chronic neurologic disease, such as dementia, can transform an individual from a relatively healthy state to one of almost complete dependence. At best, some of the neurologic states described in this chapter may resolve spontaneously or require only minor lifestyle adjustment, but more commonly, chronic neurologic conditions require lifetime rehabilitation.

The process of life care planning includes taking stock of current health status, future health care concerns, appropriate resources, and associated costs to address lifelong disability and illness management.

In general, the goal of rehabilitation is to increase self-care and promote a meaningful lifestyle that incorporates the neurologic disability. The primary goal of such tertiary prevention is to help the affected individual maintain the highest possible level of wellness.

RESOURCES

Seizures and Epilepsy

Jobst B, Thomas G: Critical review of the responsive neurostimulator system for epilepsy. *MDER* 8:405–411, 2015. doi:10.2147/mder.s62853.

Lowenstein DH: Seizures and epilepsy. In Kasper D, Fauci A, Hauser S, et al, editors: *Harrison's principles of internal medicine*, ed 19, New York, NY, 2015, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=1130&Sectionid=79755120. (Accessed 26 December 2015).

Méndez-Armenta M, Nava-Ruíz C, Juárez-Rebollar D, et al: Oxidative stress associated with neuronal apoptosis in experimental models of epilepsy. *Oxid Med Cell Longev* 2014:1–12, 2014. doi:10.1155/2014/293689.

Ropper AH, Samuels MA, Klein JP: Chapter 16. Epilepsy and other seizure disorders. In Ropper AH, Samuels MA, Klein JP, editors: *Adams & Victor's principles of neurology*, ed 10, New York, NY, 2014, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=690&Sectionid=49251504. (Accessed 26 December 2015).

Who.int: WHO | Epilepsy, 2015. Available at: http://www.who.int/mediacentre/factsheets/fs999/en/. (Accessed 21 December 2015).

Dementia and Alzheimers Disease

Alzheimer's Association: 2015 Alzheimer's disease facts and figures. *Alzheimers Dement* 11(3):332+, 2015.

Aminoff MJ, Greenberg DA, Simon RP: Dementia & amnestic disorders. In Aminoff MJ, Greenberg DA, Simon RP, editors: *Clinical neurology*, ed 9, New York, NY, 2015, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=1194&Sectionid=78426881. (Accessed 27 December 2015).

Amor S, Peferoen L, Vogel D, et al: Inflammation in neurodegenerative diseases - an update. *Immunology* 142(2):151–166, 2014. doi:10.1111/imm.12233.

Anstey K, Ranmalee E, Hosking D, et al: Bridging the translation gap: from dementia risk assessment to advice on risk reduction. *J Prev Alzheimers Dis* 2(3):189–198, 2015. doi:10.14283/jpad.2015.75.

Bonda D, Wang X, Lee H, et al: Neuronal failure in Alzheimer's disease: a view through the oxidative stress looking-glass. *Neurosci Bull* 30(2):243–252, 2014. doi:10.1007/s12264-013-1424-x.

Levine D, Langa K: Vascular cognitive impairment: disease mechanisms and therapeutic implications. *Neurother* 8(3):361–373, 2011. doi:10.1007/s13311-011-0047-z.

Madhusoodanan S: Pharmacological management of behavioral symptoms associated with dementia. *World J Psychiatry* 4(4):72, 2014. doi:10.5498/wjp.v4.i4.72.

Martorana A, Koch G: Is dopamine involved in Alzheimer's disease? *Front Aging Neurosci* 6:252, 2014. doi:10.3389/fnagi.2014.00252.

McCaulley M, Grush K: Alzheimer's disease: exploring the role of inflammation and implications for treatment. *Int J Alzheimers Dis* 2015:1–10, 2015. doi:10.1155/2015/515248.

Ropper AH, Samuels MA, Klein JP: Chapter 21. Dementia, the amnesic syndrome, and the neurology of intelligence and memory. In Ropper AH, Samuels MA, Klein JP, editors: *Adams & Victor's principles of neurology*, ed 10, New York, NY, 2014, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=690&Sectionid=50910869. (Accessed 27 December 2015).

Ropper AH, Samuels MA, Klein JP: Chapter 39. Degenerative diseases of the nervous system. In Ropper AH, Samuels MA, Klein JP, editors: *Adams & Victor's principles of neurology*, ed 10, New York, NY, 2014, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=690&Sectionid=50910890. (Accessed 27 December 2015).

Seeley WW, Miller BL: Alzheimer's disease and other dementias. In Kasper D, Fauci A, Hauser S, et al, editors: *Harrison's principles of internal medicine*, ed 19, New York, NY, 2015, McGraw-Hill. http://accessmedicine.

mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=1130&Sectionid=79755539. (Accessed 27 December 2015).

Parkinson Disease

Agim Z, Cannon J: Dietary factors in the etiology of Parkinson's disease. *Biomed Res Int* 2015:1–16, 2015. doi:10.1155/2015/672838.

Martin I, Kim J, Dawson V, Dawson T: LRRK2 pathobiology in Parkinson's disease. *J Neurochem* 131(5):554–565, 2014. doi:10.1111/jnc.12949.

Olanow C, Schapira AV, Obeso JA: Parkinson's disease and other movement disorders. In Kasper D, Fauci A, Hauser S, et al, editors: *Harrison's principles of internal medicine*, ed 19, New York, NY, 2015, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=1130&Sectionid=79755616. (Accessed 27 December 2015).

Pdf.org: Statistics on Parkinson's - Parkinson's Disease Foundation (PDF), 2015. Available at: http://www.pdf.org/en/parkinson_statistics. (Accessed 27 December 2015).

Trevor AJ, Katzung BG, Kruidering-Hall M: Drugs used in parkinsonism & other movement disorders. In Trevor AJ, Katzung BG, Kruidering-Hall M, editors: *Katzung & Trevor's pharmacology: examination & board review*, ed 11, New York, NY, 2015, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=1568&Sectionid=95702682. (Accessed 27 December 2015).

Cerebral Palsy and Hydrocephalus

Hall JE, editor: *Guyton and Hall textbook of medical physiology*, ed 13, Philadelphia, 2016, Elsevier.

Kedia S, Knupp K, Schreiner T, et al: Neurologic & muscular disorders. In Hay WW, Jr, Levin MJ, Deterding RR, Abzug MJ, editors: *Current diagnosis & treatment: pediatrics*, ed 22, New York, NY, 2013, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=1016&Sectionid=61601643. (Accessed 1 January 2016).

Kiefer M, Unterbert A: The differential diagnosis and treatment of normal-pressure hydrocephalus. *Dtsch Arztebl Int* 109(1-2):15–26, 2015.

Kinsman SL, Johnston MV: Congenital abnormalities of the CNS. In Kliegman RM, Stanton BF, St. Geme JW, et al, editors: *Nelson's textbook of pediatrics*, ed 19, Philadelphia, 2011, Saunders.

Merck Manuals Professional Edition: Cerebral Palsy (CP) Syndromes - Pediatrics, 2016. Available at: http://www.merckmanuals.com/professional/pediatrics/neurologic-disorders-in-children/cerebral-palsy-%28cp%29-syndromes. (Accessed 1 January 2016).

Rab GT: Chapter 10. Pediatric orthopedic surgery. In Skinner HB, McMahon PJ, editors: *Current diagnosis & treatment in orthopedics*, ed 5, New York, NY, 2014, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=675&Sectionid=45451716. (Accessed 1 January 2016).

Ropper AH, Samuels MA, Klein JP: Chapter 5. Ataxia and disorders of cerebellar function. In Ropper AH, Samuels MA, Klein JP, editors: *Adams & Victor's principles of neurology*, ed 10, New York, NY, 2014, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=690&Sectionid=45424414. (Accessed 1 January 2016).

Multiple Sclerosis and ALS

Aminoff MJ, Kerchner GA: Nervous system disorders. In Papadakis MA, McPhee SJ, Rabow MW, editors: *Current medical diagnosis & treatment 2016*, New York, NY, 2016, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=1585&Sectionid=97193207. (Accessed 1 January 2016).

González H, Pacheco R: T-cell-mediated regulation of neuroinflammation involved in neurodegenerative diseases. *J Neuroinflammation* 11(1):201, 2014. doi:10.1186/s12974-014-0201-8.

Hauser SL, Goodin DS: Multiple sclerosis and other demyelinating diseases. In Kasper D, Fauci A, Hauser S, et al, editors: *Harrison's principles of internal medicine*, ed 19, New York, NY, 2015, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=1130&Sectionid=79756278. (Accessed 1 January 2016).

http://www.nationalmssociety.org/what-is-MS/MS-FAQ-s#question-How-many-people-have-MS. (Accessed 1 January 2016).

Løken-Amsrud K, Lossius A, Torkildsen Ø, Holmøy T: Impact of the environment on multiple sclerosis. *Tidsskr Nor Laegeforen* 135(9):856–860, 2015. doi:10.4045/tidsskr.14.0751.

Ofengeim D, Ito Y, Najafov A, et al: Activation of necroptosis in multiple sclerosis. *Cell Rep* 10(11):1836–1849, 2015. doi:10.1016/j.celrep.2015.02.051.

Ropper AH, Samuels MA, Klein JP: Chapter 36. Multiple sclerosis and other inflammatory demyelinating diseases. In Ropper AH, Samuels MA, Klein JP, editors: *Adams & Victor's principles of neurology*, ed 10, New York, NY, 2014, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=690&Sectionid=50910887. (Accessed 1 January 2016).

Spinal Cord and Peripheral Nervous System

Adzick NS, Thom EA, Spong CY, et al: A randomized trial of prenatal versus postnatal repair of myelomenigocele. *N Engl J Med* 364(11):993–1004, 2011.

Berkley EM, Abuhamad AZ: Chapter 31 Obstetric ultrasound. In Evans AT, Defranco E, editors: *Manual of Obstetrics*, ed 8, Philadelphia, 2014, Wolters Kluwer Health.

Dean S, Lassi Z, Imam A, Bhutta Z: Preconception care: nutritional risks and interventions. *Reprod Health* 11(Suppl 3):S3, 2014. doi:10.1186/1742-4755-11-s3-s3.

Greene N, Copp A: Neural tube defects. *Annu Rev Neurosci* 37(1):221–242, 2014. doi:10.1146/annurev-neuro-062012-170354.

Hauser SL, Amato AA: Guillain-Barré syndrome and other immune-mediated neuropathies. In Kasper D, Fauci A, Hauser S, et al, editors: *Harrison's principles of internal medicine*, ed 19, New York, NY, 2015, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=1130&Sectionid=79756668. (Accessed 3 January 2016).

Laferriere F, Polymenidou M: Advances and challenges in understanding the multifaceted pathogenesis of amyotrophic lateral sclerosis. *Swiss Med Wkly* 145:w14054, 2015. doi:10.4414/smw.2015.14054.

Li Y, Walker C, Zhang Y, et al: Surgical compression in acute spinal cord injury: a review of clinical evidence, animal model studies and potential future directions of investigations. *Front Biol (Beijing)* 9(2):127–136, 2014. doi:10.1007/s11515-014-1297-z.

Neirinckx V, Coste C, Franzen R, et al: Neutrophil contribution to spinal cord injury and repair. *J Neuroinflammation* 11(1):150, 2014. doi:10.1186/s12974-014-0150-2.

Nyati K, Prasad K: Role of Cytokines and toll-like receptors in the immunopathogenesis of Guillain-Barré syndrome. *Mediators Inflamm* 2014:1–10, 2014. doi:10.1155/2014/758639.

Poppe L, Rué L, Robberecht W, Van Den Bosch L: Translating biological findings into new treatment strategies for amyotrophic lateral sclerosis (ALS). *Exp Neurol* 262:138–151, 2014. doi:10.1016/j.expneurol.2014.07.001.

Ropper AH, Samuels MA, Klein JP: Chapter 38. Developmental diseases of the nervous system. In Ropper AH, Samuels MA, Klein JP, editors: *Adams & Victor's principles of neurology*, ed 10, New York, NY, 2014, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=690&Sectionid=50910889. (Accessed 1 January 2016).

Ropper AH, Samuels MA, Klein JP: Chapter 46. Diseases of the peripheral nerves. In Ropper AH, Samuels MA, Klein JP, editors: *Adams & Victor's principles of neurology*, ed 10, New York, NY, 2014, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=690&Sectionid=50910898. (Accessed 3 January 2016).

Schwab J, Zhang Y, Kopp M, et al: The paradox of chronic neuroinflammation, systemic immune suppression, autoimmunity after traumatic chronic spinal cord injury. *Exp Neurol* 258:121–129, 2014. doi:10.1016/j.expneurol.2014.04.023.

Spinal Cord Injury (SCI) Facts and Figures at a Glance. 2015. Available at: https://www.nscic.uab.edu/Public/Facts 2015.pdf. (Accessed 3 January 2016).

Turner M, Swash M: The expanding syndrome of amyotrophic lateral sclerosis: a clinical and molecular odyssey. *J Neurol Neurosurg Psychiatry* 86(6):667–673, 2015. doi:10.1136/jnnp-2014-308946.

Wilson J, Forgione N, Fehlings M: Emerging therapies for acute traumatic spinal cord injury. *CMAJ* 185(6):485–492, 2016.

Www.cdc.gov. CDC - Amyotrophic Lateral Sclerosis: ALS. 2016. Available at: https://wwwn.cdc.gov/als/WhatisALS.aspx. (Accessed 3 January 2016).

Zandian A, Osiro S, Hudson R, et al: The neurologist's dilemma: a comprehensive clinical review of Bell's palsy, with emphasis on current management trends. *Med Sci Monit* 20:83–90, 2014. doi:10.12659/msm.889876.

46

Alterations in Special Sensory Function

Joni D. Marsh

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- What are the general functions and structures of the ear and eye?
- What are the general manifestations of hearing impairment?
- How do conductive and sensorineural mechanisms of hearing loss differ in etiology and treatment?
- What are the predisposing factors, clinical manifestations, and management of otitis media?
- What are the general manifestations of visual impairment?
- What are the causes, clinical manifestations, and management of common visual disorders, including errors of refraction, strabismus, cataract, and retinopathies?
- How do open-angle and acute angle-closure glaucoma differ?
- How do the two forms of macular degeneration differ?
- What are the causes, clinical manifestations, and management of smell and taste disorders?

CHAPTER OUTLINE

***HEARING AND BALANCE,* 937**
Structure and Function of the Ear, 937
External Ear, 937
Middle Ear, 937
Inner Ear, 937
Balance, 938
Vertigo, 938
General Manifestations of Hearing Impairment, 938
Hearing Impairment Disorders, 939
Conductive Hearing Impairment, 939
Loss Caused by Cerumen Impaction and Foreign Body Occlusion, 939
Otosclerosis, 939
Sensorineural Hearing Impairment, 939
Loss Caused by Ototoxic Medications, 939
Loss Caused by Trauma, 939
Presbycusis, 940
Meniere Disease, 940
Otitis Media, 941
Acute Otitis Media, 941
Chronic Otitis Media, 942
Interventions for Individuals with Hearing Impairment, 942
***VISION,* 942**
Structure of the Eye, 942
Visual Pathways, 943
General Manifestations of Visual Impairment, 944
Disorders of the Eye, 944
Errors of Refraction, 944
Myopia, Hyperopia, Presbyopia, and Astigmatism, 944
Age-Related Disorders, 945
Strabismus, 945
Amblyopia, 945
Cataracts, 947
Retinopathy, 947
Retinal Detachment, 947
Diabetic Retinopathy, 947
Age-Related Macular Degeneration, 948
Glaucoma, 948
Chronic Open-Angle Glaucoma, 948
Acute Angle-Closure Glaucoma, 950
Visual Field Deficits, 950
Visual Field Loss, 950
Interventions for Individuals with Vision Impairment, 950
***SMELL AND TASTE,* 952**
Disorders of Smell and Taste, 952

The human body has countless sense organs that fall into two main categories: general sense organs and special sense organs. By far the most numerous are the general sense organs or receptors. The receptors function to produce the general or somatic senses. Examples of these senses are touch, temperature, and pain and the receptors that initiate various reflexes necessary for maintaining homeostasis (see Chapter 43). The largest general sense organ in the body is the skin.

Special sense organs, by comparison, function to produce the unique sensations of hearing, balance, vision, smell, and taste. These senses allow humans to interact with their environment in a meaningful way.

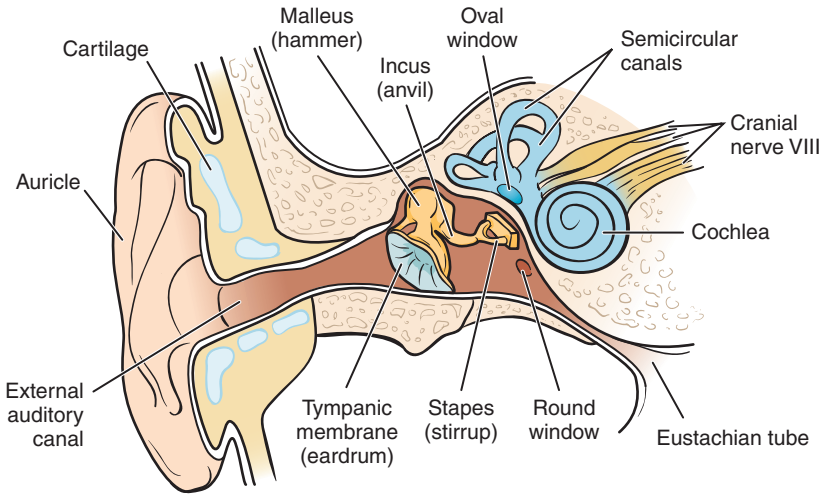


FIG 46.1 Anatomic structures of the ear.

Alterations in sensory function may be acute/short-term, chronic/long-term, or progressive in nature. They may result from such factors as genetics, disease, infection, trauma, and normal aging. Alterations in special sensory function require prompt assessment, evaluation, and treatment from appropriate health professionals. Equally important is an assessment of how the sensory impairment affects the individual's activities of daily living. This chapter discusses special sensory function with regard to physiologic processes, sensory impairment, and the diagnosis and management of these impairments.

HEARING AND BALANCE

STRUCTURE AND FUNCTION OF THE EAR

External Ear

Hearing results from normal functioning of several complex structures both external and internal to the body. Sound consists of waves of vibrations in the air produced in the environment. These vibrations travel much like ripples in a pool of water. Externally, these vibrations are caught and funneled into the ear canal by the auricles (Fig. 46.1). Even though the auricles are in a fixed position and lie close to the head, their shape serves to concentrate sound waves, especially high-frequency waves. The auricles also have an important role in sound localization.

The ear canal has a somewhat S shape from its opening to its termination at the tympanic membrane. This configuration affords both protection from airborne foreign objects and access to sound. The outer portion of the ear canal contains hair to filter out unwanted substances. Along the ear canal are also glands that secrete cerumen. This brown, waxlike substance coats the hairs in the canal to help prevent the entrance of foreign bodies into the ear canal. After entering the ear canal, sound waves strike the tympanic membrane (eardrum) and cause it to vibrate. The tympanic membrane is a thin, elastic membrane that is highly sensitive to changes in pressure.

Middle Ear

The middle ear is a bony, air-containing space that functions primarily as a structure by which sound energy is transmitted from the air to the fluids of the inner ear. The tympanic membrane is connected to the first of the ossicles, the malleus (hammer), followed by the incus (anvil) and stapes (stirrup). The ossicles further amplify the sound waves and then transfer airborne sound waves to the fluid-filled inner ear at the oval window.

The eustachian tube is also part of the middle ear, and although it does not contribute directly to the transmission of sound through the ear, absence of proper function can greatly affect hearing. This tube has a mucosal lining and extends from the middle ear cavity to the nasopharynx. It makes equalization of pressure against the inner and outer surfaces of the tympanic membrane possible, thus improving mobility of the membrane for sound transduction. A patent eustachian tube prevents membrane rupture and discomfort that marked pressure differences can produce.

Inner Ear

The inner ear is composed of the oval window, the cochlea, and the semicircular canals. Within the cochlea are three parallel tubes: the scala vestibuli, the scala media, and the scala tympani. Movement of the perilymph, a fluid much like cerebrospinal fluid, in the scala vestibuli and the scala tympani is eventually dissipated by movement of the round window (Fig. 46.2). The scala tympani and the scala vestibuli are continuous with one another at the apex of the cochlea through an opening called the *helicotrema*.

Transmission of the sound stimulus from the scala vestibuli to the vestibular membrane results in displacement of the endolymph, the fluid contained in the membranous labyrinth of the scala media and the basilar membrane. The organ of Corti, which contains the receptors for hearing, lies on the basilar membrane. Perilymph and endolymph transmit the mechanical vibrations from the footplate of the stapes to the organ of Corti. Endolymph also transports nutrients to the organ of Corti. No direct communication between endolymph and perilymph is normally present.

The organ of Corti consists of a series of sensory hair cells and supporting cells. These cells are innervated by the sensory fibers from the vestibulocochlear nerve (cranial nerve VIII). Overhanging the organ of Corti is a flexible flap of tissue called the *tectorial membrane* (Fig. 46.3). Hairs of the sensory cells of the organs of Corti are in contact with the tectorial membrane. The wave of perilymph induces movement of the basilar membrane, which causes a pull or shearing force on the hair cells in contact with the tectorial membrane. This action transforms the mechanical energy of sound into electrical impulses stimulating the vestibulocochlear nerve. Impulses are relayed through nuclei in the medulla, pons, midbrain, and thalamus before reaching the auditory area of the temporal lobe.

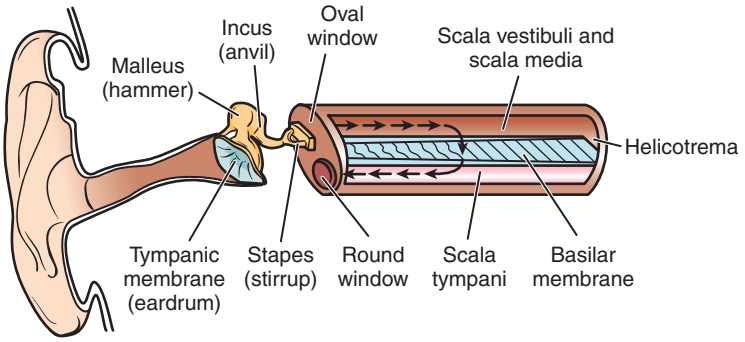


FIG 46.2 Movement of fluid in the cochlea after forward thrust of the stapes.

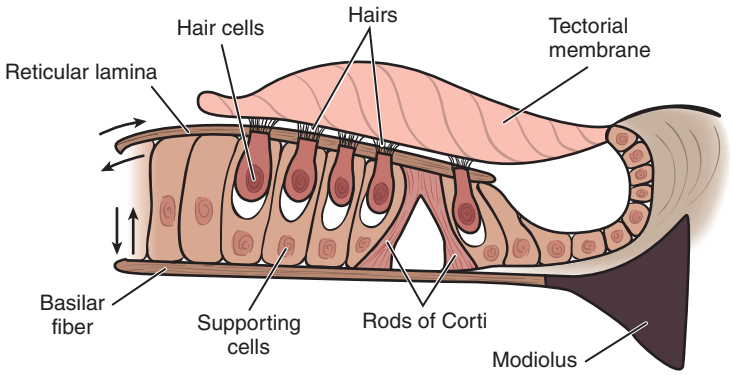


FIG 46.3 Stimulation of hair cells by the back-and-forth movement of the hair cells in the tectorial membrane.

Balance

The ear has dual sensory functions. In addition to its role in hearing, it functions as the sense organ of equilibrium. The stimulation or "trigger" responsible for balance involves activation of receptor hair cells contained in the semicircular canals. Movement of the head causes movement of the endolymph contained in the semicircular canals. The receptor hair cells in turn create a nerve impulse in the vestibular portion of the vestibulocochlear nerve (cranial nerve VIII), where the stimulus is transmitted to the brain. Signals from the inner ear are involved not only in keeping individuals upright, but also in controlling the eye muscles so that the eyes can remain fixed on the same point despite changes in the position of the head.

Vertigo

Vertigo is a common symptom of vestibular disorders rather than a well-defined disease. It is either a sensation of motion without any actual motion or an exaggerated sense of motion; it is not simply a sensation of "spinning." Accompanying symptoms may include nausea, vomiting, pallor, and sweating. Nystagmus often is also noted. It is not associated with a loss of consciousness or a feeling of impending loss of consciousness more common to syncope. Vertigo can have a peripheral vestibular cause (common) or a central nervous system cause (uncommon). In the case of Meniere disease (discussed later in this chapter), a disorder in which vertigo is common, the cause is unknown.

Disorders of the brainstem or cerebellum that may also cause vertigo include tissue ischemia secondary to atherosclerosis; tumors; or conditions such as psychiatric disorders, migraine headaches, or multiple sclerosis. However, in these cases, additional neurologic signs and symptoms are typically present. Drugs may also cause vertigo (alcohol, anticonvulsants, sedatives). Management of vertigo is usually aimed at the cause if known. Medications such as antihistamines and anticholinergics can be helpful.

GENERAL MANIFESTATIONS OF HEARING IMPAIRMENT

Hearing impairment is a common disorder and a leading chronic health problem in the United States. Approximately 15% (26 million) of Americans age 20 to 69 report some degree of hearing loss. Out of 1000 U.S. children, 2 to 3 are born deaf or hard of hearing.

Hearing may be impaired in many ways, and impairments can occur across the age spectrum. Disorders may affect the outer ear, such as impacted cerumen and foreign bodies in the ear canal. The middle ear may be affected by fluid effusion, infection (otitis media), tumors, or diseases such as otosclerosis. Hearing loss may also be caused by repeated exposure to loud sounds or ototoxic medications such as aminoglycoside antibiotics, chemotherapeutic agents, and high-dose loop diuretics. Some of these medications can cause hearing loss even when administered at therapeutic doses. Other causes of hearing loss, especially in children, include infection (measles, meningitis), environmental teratogens (radiation), intrauterine infections (cytomegalovirus, herpes simplex virus, human immunodeficiency virus, and *Toxoplasma*), maternal metabolic disorders (diabetes, hypothyroidism), and exposure to industrial chemicals (solvents or pesticides). In adults, poor nutrition and smoking also contribute to hearing loss. By whatever mechanism hearing impairment occurs, the signs and symptoms are similar.

Symptoms of hearing impairment may be manifested in behavior such as inattentiveness, speaking out of turn in conversations, withdrawal from social situations, increased volume of voice when speaking, increased volume of radio or television, confusion, loss of reaction to loud sounds, and emotional outbursts. Children with hearing impairment may

demonstrate inattentiveness and difficulty with articulation and the development of speech. Alterations in hearing function can generally be classified into two categories, conductive and sensorineural, depending on the cause of the impairment. Some hearing impairments have a component of both. The following alterations in hearing function are categorized according to the primary cause of dysfunction.

HEARING IMPAIRMENT DISORDERS

Conductive Hearing Impairment

Conductive hearing loss occurs when sound cannot reach the cochlea. Individuals with conductive hearing impairment have a decreased sensitivity to sound. This type of hearing impairment is caused by dysfunction in the external or middle ear. Four mechanisms, each resulting in impairment of the passage of sound vibrations to the inner ear, lead to conductive hearing impairment: (1) obstruction (cerumen impaction), (2) mass loading (middle ear effusion), (3) stiffness effect (otosclerosis), and (4) discontinuity (ossicular disruption). Conductive hearing loss is generally correctable with medical or surgical therapy—or in some cases both.

Loss Caused by Cerumen Impaction and Foreign Body Occlusion

Etiology. Cerumen impaction is a common and frequently overlooked cause of conductive hearing loss, especially in the elderly. In most cases, cerumen impaction is self-induced through attempts at cleaning the ear with objects such as cotton swabs. Foreign bodies in the ear canal occur most frequently in children. Objects such as small stones, pieces of wood, peas, beans, and paper are fairly common.

Clinical manifestations and treatment. Sometimes no symptoms are present and the foreign body is discovered on routine examination. If the foreign body is an insect, beating of its wings and movement may cause distress. When symptomatic, however, foreign bodies can cause pain or drainage of pus from the ear. The external ear canal is highly sensitive to touch and bleeds easily, which increases the risk for subepithelial hematomas from minor trauma. Therefore removal of solid foreign bodies carries a risk of additional trauma to the ear canal, as well as tympanic membrane rupture, if the individual is not completely cooperative or removal is difficult. Light anesthesia may be necessary. Firm materials may be removed from the canal with loop or hook instruments, taking care not to push the object farther into the canal. Irrigation should not be performed on organic foreign bodies (beans, peas) because water may cause them to swell. Living insects may be immobilized with lidocaine before removal. Excess cerumen may be removed with gentle irrigation.

Otosclerosis

Etiology. Otosclerosis is a localized, inflammatory disease of the metabolism of endochondral bone of the otic capsule. It is characterized by abnormal removal of mature dense otic capsule bone by osteoclasts, which are replaced by bone of greater cellularity and thickness by osteoblasts. There is increased vascularity and mucosal thickening in the affected area as well. These changes produce a progressive conductive, sensorineural, or mixed hearing impairment. Resorption of bone is followed by the formation of new spongelike bony lesions usually occurring on and around the ossicles of the middle ear. Lesions involving the footplate of the stapes cause decreased transmission of sound waves to the oval window. However, when otosclerotic lesions impinge on the cochlea, permanent sensorineural hearing loss can occur. The basic initiating factors are unknown, but evidence supports an autosomal-dominant tendency with incomplete penetrance and variable expression. The disease is most common in Caucasian middle-aged women. Pregnancy may accelerate the otosclerotic process. Viral infections and autoimmune processes are also thought to be possible causes. The age of onset is variable due to the insidious progression of the disorder, but the most common ages are between 15 and 45. There may be periods of symptom worsening, followed by times of little apparent change.

Diagnosis and treatment. The diagnosis of otosclerosis is made through careful history taking, audiogram/tuning fork testing, and radiologic studies, such as high-resolution computed tomography. Generally, hearing loss begins in one ear, but 80% to 90% of affected individuals will develop bilateral impairment. Although hearing loss may be severe, speech discrimination is preserved except in the instance of cochlear involvement. The individual may report being able to hear better in a noisy environment than in a quiet one. Tinnitus is often present. Hearing tests reveal a conductive loss of varying severity.

Otosclerosis is generally manged surgically in an effort to prevent the conductive hearing loss. The limitation on treatment options for otosclerosis is related largely to the lack of exact knowledge regarding the cause and pathogenesis of the disease. The universally accepted operation for otosclerosis is stapedectomy, or removal of the focus of the disease by removing the stapes and inserting a prosthesis. In the case of otosclerosis involving the cochlea, treatment with oral sodium fluoride and bisphosphonates is associated with some decrease in development of the sensorineural hearing loss. Amplification with hearing aids is another approach.

Sensorineural Hearing Impairment

In sensorineural hearing impairment, the hearing mechanism is disturbed in the inner ear in the cochlea or the vestibulocochlear nerve to the brain. Long-term exposure to loud sounds, ototoxic medication, trauma, metabolic causes, aging, and certain disease states cause sensorineural hearing impairment. Sensorineural hearing loss is usually irreversible. Progress, however, is being made in identifying specific genetic targets that may protect or aid in the regeneration of the hair and supporting cells of the inner ear.

Loss Caused by Ototoxic Medications

Drug toxicity is an increasingly important cause of sensorineural hearing loss. The drugs most well known for this effect are the aminoglycoside antibiotics, loop diuretics, salicylates, quinine and related antimalarials, and cytotoxic antineoplastic drugs. Most ototoxic drugs affect the hair cells of the cochlea. Unfortunately, these ototoxic effects may not become apparent during drug administration but may occur days to weeks after the therapy has been terminated. Ototoxicity may also be unilateral. Aspirin can produce a temporary hearing loss and tinnitus in individuals receiving high doses. In most cases, however, both of these symptoms disappear after aspirin use is terminated.

Loss Caused by Trauma

Etiology. Acquired sensorineural hearing loss caused by chronic, repeated exposure to loud sounds is common in the U.S. population. Four million people work each day in an environment with damaging levels of noise, and 10 million people in the United States have noise-induced hearing loss. Noise-induced hearing loss can be associated with the use of firearms, personal stereo systems, and power tools and with occupations such as firefighting, construction, agriculture, mining, and manufacturing. Transportation and the military are other occupations that may affect hearing.

The loudness of sound/noise is measured in the logarithmic units of decibels (dB). A normal whisper is measured at approximately 30 dB, a conversation at 3 feet at 50 to 60 dB. In contrast, ambulance sirens have been measured at 120 dB, lawnmowers and motorcycles at 90 dB.

Sounds exceeding 85 dB are considered potentially injurious, and chronic noise exposure is the most damaging. If exposure is severe enough, most structures of the inner ear can be damaged, including the organ of Corti. Sensory hair cells and supporting cells are lost because of overexposure. Noise-induced hearing loss typically is bilateral and affects higher (speech) frequencies first.

Noise exposure has two phases: the first is a temporary *threshold* shift. When the ear is exposed to a loud sound, it will show a loss of sensitivity (a rise in the threshold for sound). If the hearing returns to normal after the sound has been removed, the shift was temporary and no permanent damage has occurred. If hearing does not return to normal, damage has occurred and the hearing impairment is permanent. Such a permanent threshold shift is the second phase of the damage. The ears of some individuals are more easily affected by noise, and considerable damage may occur before individuals are aware of the hearing loss.

Clinical manifestations. Individuals with hearing loss caused by noise trauma report that they are unable to discriminate words, particularly in noisy environments. Complaints about tinnitus are expressed more often than complaints about hearing loss. A diagnosis of noise-induced hearing loss is made through careful history and audiometric testing. Because noise-induced hearing loss is irreversible, no medical therapy can help once the problem has been established. Prevention is presently the only treatment for this type of hearing impairment.

Sensorineural hearing impairment can also occur with head trauma and subsequent damage to the structures of the inner ear. If blood is coming from the ear or the temporal bone is fractured, damage should be suspected. As a rule, hearing loss from trauma or head injury is permanent if the cochlea is damaged.

Presbycusis

Presbycusis is a multifactorial sensorineural hearing loss and the most common form of hearing loss in older adults. Approximately one-third of people ages 65 to 70 and 80% of those older than 85 suffer from age-related hearing loss. Typically, the hearing impairment is of gradual onset, is bilateral, and results in difficulty hearing high-pitched tones and conversational speech. Presbycusis can progress to involve the middle and lower tones. Frequently, individuals complain that people are mumbling to them but deny any other type of hearing loss.

Etiology. In the elderly, hearing loss represents the convergence of many risk factors, including noise and ototoxic medication exposure, as well as genetic and lifestyle factors. However, four causes of presbycusis have been theorized: (1) sensory, characterized by atrophy and degeneration of the sensory and supporting cells; (2) neural, typified by loss of neurons in the cochlea and central nervous system, including changes in the brain; (3) metabolic, characterized by atrophy of the wall of the cochlea affecting central auditory processing; and (4) mechanical, in which the middle ear undergoes changes in properties with a resulting conductive hearing loss. Some of these age-related changes are shown in the Geriatric Considerations: Changes in Hearing box. Given the multitude of factors involved, uncertainty remains regarding the exact cause of presbycusis.

Diagnosis. Assessment of an individual with suspected presbycusis should begin with exclusion of all other causes of hearing impairment. Diseases such as diabetes, stroke, and heart disease may produce effects similar to those seen with hearing loss and must be ruled out. The diagnosis is made by obtaining a thorough history and performing audiometric studies. Individuals with presbycusis respond well to hearing aids that amplify sound. Many simple lifestyle adjustments that will be mentioned at the end of this section can dramatically improve the quality of life for an individual experiencing presbycusis. It is important for these patients to avoid excessive noise exposure and ototoxic drugs, which may cause further deterioration of hearing loss.

Meniere Disease

Etiology and pathogenesis. Meniere disease is an idiopathic disorder in which there is an impairment of the ability of the inner ear to regulate normal homeostatic systems, including the production, maintenance, and recycling of endolymph and perilymph. It has been theorized that increases in endolymph cause distention and rupture of the scala media and that over time this repeated trauma leads to degeneration of the neural end organs of the cochlea. Newer theories suggest ischemic/vascular causes, and a close association has been made between Meniere disease and migraine headache.

Many conditions, including autoimmunity, allergies, viral and bacterial infections (such as syphilis), head trauma, metabolic derangements, and chronic stress, have been suggested as causative agents, but the precise cause cannot be established in most cases. Men and women are equally affected by this disorder, and the onset of symptoms is typically in the fifth decade of life.

Clinical manifestations. Clinical manifestations of Meniere disease include episodic tinnitus, fluctuating sensorineural hearing loss, vertigo, and sensations of ear fullness. In the early stages, hearing loss fluctuates with a return to normal. Remissions and exacerbations are typical. The hearing loss is usually in the low tones, and symptoms are usually unilateral. As the disease progresses, hearing loss becomes permanent. Episodes of vertigo may be immediately preceded by the sensation of pressure in the ear, increased hearing loss, increased tinnitus, or an alteration in the quality of these symptoms. The onset of vertigo is usually sudden, reaches maximal intensity within a few minutes, usually lasts for an hour or more, and either subsides completely or continues as a sensation of unsteadiness for several hours or days. The tinnitus is typically a low buzzing or blowing sound and is frequently louder before the attack of vertigo. The attacks are not precipitated by positional changes and may be several weeks or months apart. In the initial stages of the disease they may be years apart. If not treated, the episodes may become more frequent and severe. Nystagmus, which occurs only during acute attacks, may be directed to the side opposite the involved ear. The hearing loss progresses in a stepwise fashion, and as hearing loss increases, vertigo tends become less severe.

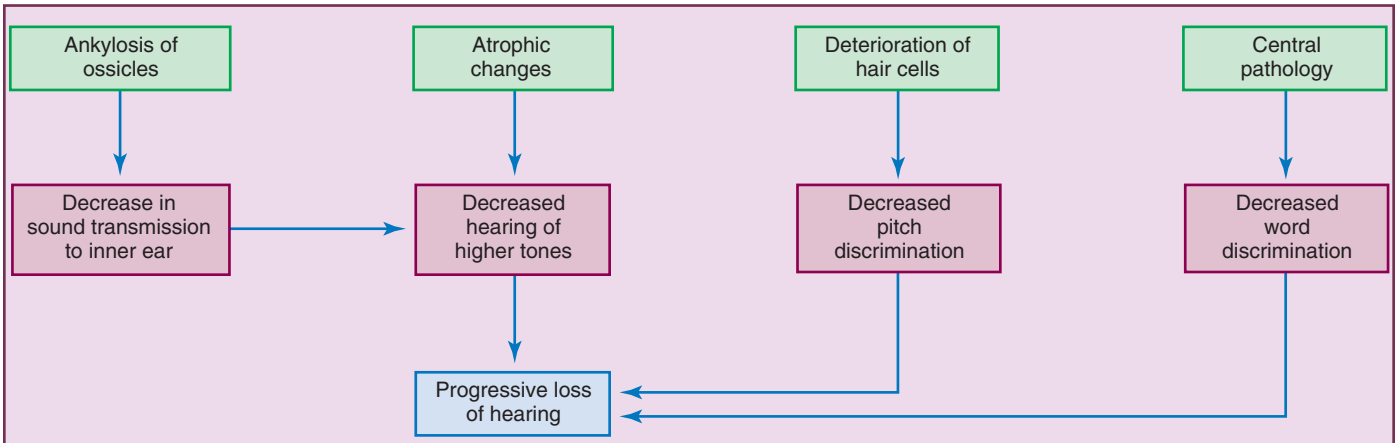
Diagnosis and treatment. Physical examination, including neurologic and otolaryngologic examination, is generally normal in those with Meniere disease. Radiologic studies are often used to rule out other causes of the symptoms of Meniere disease such as acoustic neuroma. Electrophysiologic studies, such as auditory brainstem response testing and electrocochleography, and audiometric tests, as well as glycerol dehydration testing, can lead to a diagnosis of Meniere disease.

Treatment for Meniere disease consists of providing symptomatic relief during acute episodes with antiemetics and anticholinergics such as meclizine. Between acute attacks, eating a low-sodium diet and using diuretics may help reduce the volume of endolymph. Cessation of smoking, management of stress, and elimination of caffeine from the diet are also suggested. Newer interventions aimed at migraine and/or vascular etiology include migraine prophylactic medications and management of obesity, blood pressure, and hyperlipidemia.

Several surgical interventions are used to manage Meniere disease. Shunts can be placed to drain excess endolymph, and ablation of portions of the eighth cranial nerve and destruction of the labyrinth are options. In refractory cases, patients may undergo intratympanic corticosteroid or gentamycin injections. These interventions have different indications, risks, and benefits associated with them. Almost all patients who choose surgical intervention have failed to respond to medical treatment.

GERIATRIC CONSIDERATIONS

Changes in Hearing



Presbycusis, or age-related hearing problems, occurs after age 50 and is thought to be caused by structural changes in the organs of hearing. Ankylosis of the ossicles can lead to a functional decrease in transmission of sound to the inner ear.

In the inner ear or cochlea, degeneration of hair cells, changes in the basilar membrane, or atrophic changes can lead to decreased hearing of higher tones. With these changes is also noted a decline in pitch discrimination. As hearing is progressively lost, even lower-pitch tones will be more difficult to hear. Degenerative changes in the central auditory pathways along with increased information processing time also contribute to the inability to discriminate words.

TABLE 46.1 Comparison of Otitis Media Types

Type	Onset/Duration	Symptoms	Treatment Options
Acute otitis media	Sudden onset, associated with upper respiratory tract infections	Reddened tympanic membrane with poor mobility, may be bulging or ruptured, ear pain	Antibiotics, analgesics, antipyretics, or "watch and wait"
Recurrent acute otitis media	Three or more episodes in 6 months/4 or more episodes in 12 months with complete resolution between episodes	Same as above	Lifestyle modification Daily doses of prophylactic antibiotics in some cases, ventilation tube placement
Chronic otitis media	Duration of more than 12 wk, may develop as consequence of acute otitis media	Thick immobile tympanic membrane, purulent drainage from ear, may have conductive hearing loss, pain is rare	Removal of debris from middle ear, ventilation tube placement
Otitis media with effusion	May precede or follow any type of otitis media; presence of effusion for 12 weeks or more	Ear popping; feeling of pressure in middle ear; hearing loss; retraction of tympanic membrane, fluid line, or bubbles	Treat acute otitis media or "watch and wait"

OTITIS MEDIA

Otitis media is an inflammation of the middle ear. It is almost always due to poor functioning of the eustachian tube and is often diagnosed by the presence of effusion. It is the most common reason for a child to require medical attention.

Otitis media is more common in the winter months when viral and bacterial infections are most prevalent. Upper respiratory tract infections can cause eustachian tubes to become blocked and predispose individuals to middle ear inflammation. Children are especially susceptible because of shorter, more flexible, and horizontally positioned eustachian tubes. The dysfunction of the eustachian tube prevents middle ear secretions from draining and creates negative pressure in the middle ear space. Negative pressure leads to the introduction of infected nasopharyngeal secretions into the middle ear. Risk factors for otitis media include use of pacifiers, secondhand cigarette smoke exposure, gastroesophageal reflux, poor socioeconomic conditions, daycare attendance, and propped bottles. Males, Native Americans, Eskimo children, children with craniofacial abnormalities such as cleft palate, and individuals with Down syndrome have a higher incidence of otitis media.

Much confusion surrounds the use of terminology in categorizing otitis media. This confusion relates to the presence of effusion and the length of illness (Table 46.1).

Acute Otitis Media

Acute otitis media is characterized by the sudden onset of ear pain in association with symptoms of upper respiratory tract infection. Although older children and adults complain of pain, younger children may demonstrate irritability, difficulty eating and sleeping, or tugging at the affected ear, as well as fever. Physical examination reveals a reddened tympanic membrane that has poor mobility. Bulging or rupture of the tympanic membrane causing otorrhea may also be present. In children

older than 2 years, spontaneous resolution of symptoms often occurs within 4 days of onset.

In 2013 the American Academy of Pediatrics updated published guidelines for the diagnosis and management of acute otitis media. Antibiotic therapy was recommended for all children younger than 6 months of age, for children ages 6 to 24 months with otorrhea, with unilateral/bilateral involvement with severe symptoms, and for all children older than 2 years with otorrhea, with bilateral/unilateral severe infection. In children older than 2 years with milder symptoms, a "wait and see" approach may be used. Adequate pain relief and fever control are important, as well as repeated evaluation. High-dose amoxicillin is an appropriate choice for initial therapy, and amoxicillin/clavulanic acid is recommended for those who have previously been treated for acute otitis media. Surgical placement of ventilation tubes in the tympanic membrane is also done in cases of recurrent otitis media. Complications of unresolved otitis media include hearing loss, mastoiditis, meningitis, osteomyelitis of the skull bones, and facial paralysis.

Chronic Otitis Media

Chronic otitis media is inflammation in the middle ear lasting longer than 12 weeks. Irreversible damage has occurred to structures in the middle ear. This damage manifests in many forms, including atrophy or perforation of the tympanic membrane or adhesions in the middle ear causing tympanic membrane retraction. Calcification of the ossicles may occur, as well as the formation of cholesteatomas (benign, slowly growing collections of skin tissue) within the middle ear space. The hallmark clinical sign of chronic otitis media is purulent drainage from the ear. Pain is an uncommon finding, and conductive hearing loss may occur. Chronic otitis media generally develops as a consequence of acute otitis media, but it may follow other diseases or trauma. Management of chronic otitis media generally includes surgical removal of debris in the middle ear, placement of ventilation tubes in the tympanic membrane, and adenoidectomy to assist with eustachian tube function.

INTERVENTIONS FOR INDIVIDUALS WITH HEARING IMPAIRMENT

In general, interventions for individuals with hearing loss are aimed at maximizing their residual hearing ability and allowing for compensation with other senses. Early identification and intervention are vital, and there are several recommendations for screening for hearing loss across the life span.

To improve communication, adequate visual contact should be made. Lighting and positioning should be such that the individual can see the speaker's lips. Reductions should be made in background noise. Speech should be at a normal rate and rhythm and at normal volume. Shouting can distort sounds and actually make them more difficult to hear. The speaker should use shorter sentences and gestures such as pointing when appropriate.

Devices that amplify sound or transform sounds into tactile or visual signals may be helpful. Amplifiers for the telephone, television, or radio; closed-captioned television; and teletypewriters are examples of these devices. Others include doorbells and telephones that glow as well as ring and flashing smoke detectors and alarm clocks.

Implanted hearing devices such as cochlear implants may be helpful for patients over the age of 2 years with profound hearing loss. Surgically implanted cochlear electrodes work together with an external processor that converts sound waves to electrical signals that can be recognized by the brain. These devices can improve communication and provide psychosocial benefits.

KEY POINTS

- Perception of sound requires that sound waves be transmitted through the outer ear canal, across the tympanic membrane, and through the ossicles to the oval window. Movement of the oval window initiates movement of perilymph, which causes movement of endolymph through the vestibular membrane. This fluid's motion stimulates the neurosensory organs of hearing—the hair cells. Bending of the hair cells induces action potentials in the cochlear nerve, which projects to the brainstem. Neural projections to the auditory area in the temporal lobe result in sound perception.
- Balance is controlled by hair cells contained in the semicircular canals. Stimulation of these cells by head movement causes nerve impulses to be transmitted to the brain to keep individuals upright and control eye movement.
- Vertigo, the sensation of motion or aggravation of motion, is a cardinal symptom of disorders of the vestibular system. Vertigo is often associated with nystagmus and nausea.
- Hearing loss may result from interruptions in any part of the sound transmission pathway. Disorders of the outer and middle ear are generally termed *conductive* because sound waves are not reliably conducted to sensory organs of hearing. Accumulation of wax in the outer ear, ossification of bones, and middle ear infections and edema may result in conductive hearing loss. Conductive hearing loss is amenable to treatment.
- Sensorineural hearing loss is due to dysfunction of the hair cells or neural pathways to the brain. Chronic exposure to loud noise, ototoxic drugs, head trauma, and aging changes may lead to sensorineural hearing loss. Sensorineural hearing loss is not as amenable to treatment as is conductive loss.
- Otosclerosis is a disorder characterized by resorption of healthy bone and deposition of weak, spongelike bone in the ossicles of the middle ear, most frequently the stapes. These bony lesions lead to progressive conductive hearing loss.
- Presbycusis is a gradual sensorineural hearing loss common in older adults. Its cause is unclear and difficult to distinguish from other types of hearing loss, especially noise trauma.
- Meniere disease is a chronic inner ear disease of unknown cause characterized by vertigo and progressive unilateral sensorineural hearing loss.
- Otitis media, or inflammation of the middle ear, is most frequently seen in children and commonly results from eustachian tube dysfunction after upper respiratory tract infections. Otitis media can be both acute and chronic.

VISION

Healthy vision requires three basic processes to function appropriately: (1) formation of an image on the retina, (2) stimulation of rods and cones, and (3) conduction of nerve impulses to the brain. Malfunction of any of these processes can disrupt normal vision.

STRUCTURE OF THE EYE

The eye is a spherical structure contained in the bony cavity of the eye socket composed of three basic layers: the sclera, the choroid, and the retina (Fig. 46.4). The sclera is white and opaque and is made up of dense connective tissue. It aids in protecting the inner structures of the eye and helps maintain the shape of the eye. The sclera merges with the coverings of the optic nerve on the posterior of the eye. The clear front portion of the sclera is the cornea. The cornea is also composed of dense connective tissue and has a greater curvature than the sclera that causes it to protrude from the sclera. No blood vessels are located in the cornea. Deep within the anterior portion of the sclera, at its conjunction with the cornea, lies a ring-shaped venous sinus: the canal

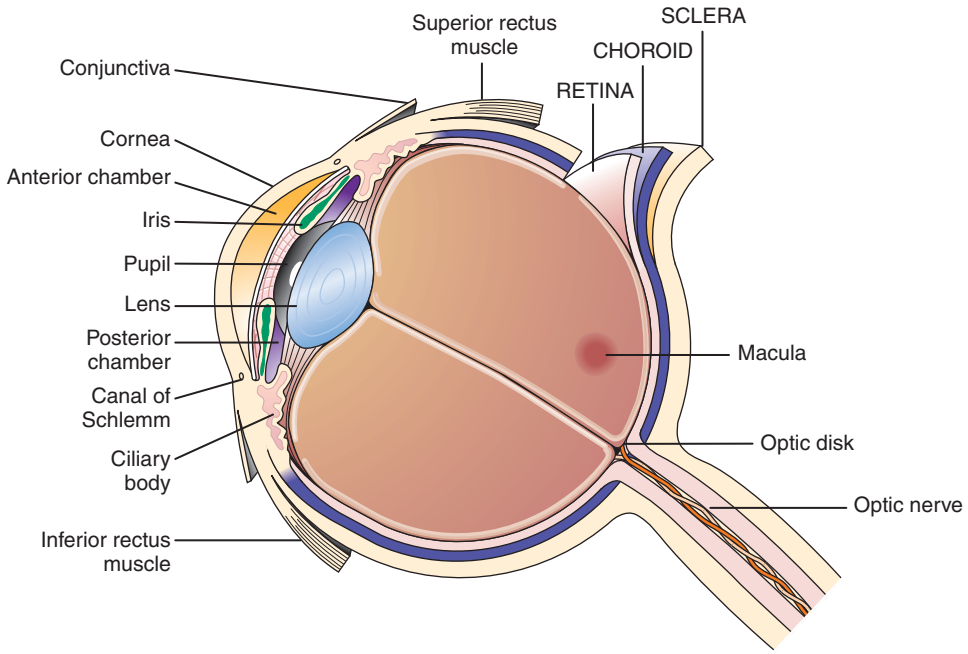


FIG 46.4 Anatomic structures of the eye.

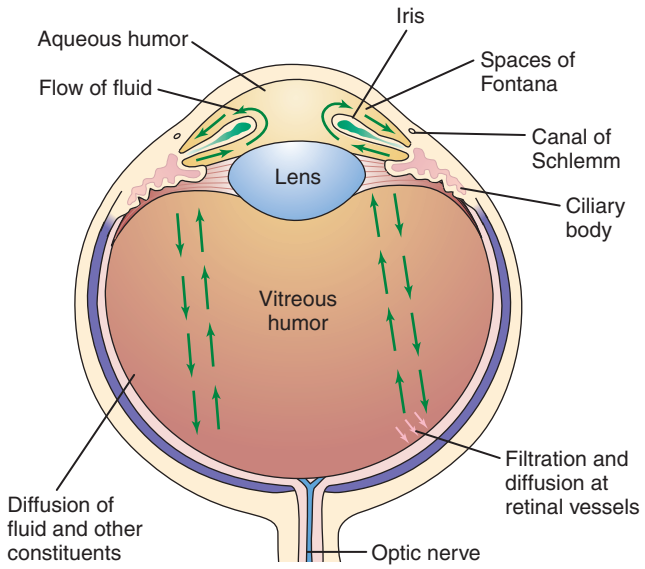


FIG 46.5 Circulation of the aqueous and vitreous humor of the eye.

of Schlemm. Affixed to the sclera are the extraocular muscles that control eye movement.

The choroid layer of the eye is highly vascularized and darkly pigmented. Attached to this layer is the iris. The iris is a muscular diaphragm whose pigments are responsible for eye color. The iris controls the size of the pupil, the opening through which light stimuli enter the posterior portion of the eye. Behind the pupil is a clear lens. The lens is a transparent, avascular elastic membrane. This elasticity assists in focusing light stimuli on the retina.

The eye is composed of anterior and posterior chambers separated by the lens and iris. The anterior chamber is filled with aqueous humor, the transparent protein-free liquid that is formed in the ciliary body and drained through the canal of Schlemm (Fig. 46.5). Aqueous humor provides oxygen and nutrients to the lens and cornea and is continually being formed and reabsorbed. The balance between formation and reabsorption of aqueous humor regulates the total volume and pressure of the intraocular fluid. The posterior chamber is the portion of the eye behind the lens that contains a thicker fluid, vitreous humor.

VISUAL PATHWAYS

The innermost layer of the eye is the retinal layer. It is here that light waves are transformed into nerve impulses. The retina is composed of several layers. It contains two types of photoreceptors (rods and cones) and four types of neurons (bipolar cells, ganglion cells, horizontal cells, and amacrine cells) (Fig. 46.6). The photoreceptor cells synapse with the bipolar cells, which in turn synapse with ganglion cells. Ganglion cell axons converge to leave the eye within the optic nerve. Rods are important for nighttime and peripheral vision and outnumber cones by nearly 20 to 1. Cones are stimulated by relatively high-intensity light and are responsible for color and visual acuity. The greatest concentration of cones occurs in the macula. This area is devoid of retinal vessels and is responsible for the most detailed vision.

Transmission from photoreceptors to bipolar cells and then to retinal ganglion cells is modified by the horizontal and amacrine cells. These cells function to "sharpen" and enhance the responses of the ganglion cells.

The pigmented layer of the retina, or the retinal pigment epithelium, is one cell thick. It functions to protect and nourish the retina. The retinal pigment epithelium also removes metabolic cellular debris from the photoreceptor cells, prevents new blood vessel growth into the retina, and absorbs light to diminish scattering and thereby enhance vision.

Action potentials from the rods and cones are communicated throughout the other layers of the retina. All axons of the ganglion neuron extend back to a small circular area in the posterior of the eye, the optic disk, where the optic nerve passes through the sclera. When an image is focused on the retina, it is projected upside down and reversed left to right. An object in the upper temporal visual field of the right eye reflects its image on the lower nasal area of the retina. The optic disk is a blind spot because it does not contain any rods or cones. The optic nerve maintains the spatial arrangement of upside-down

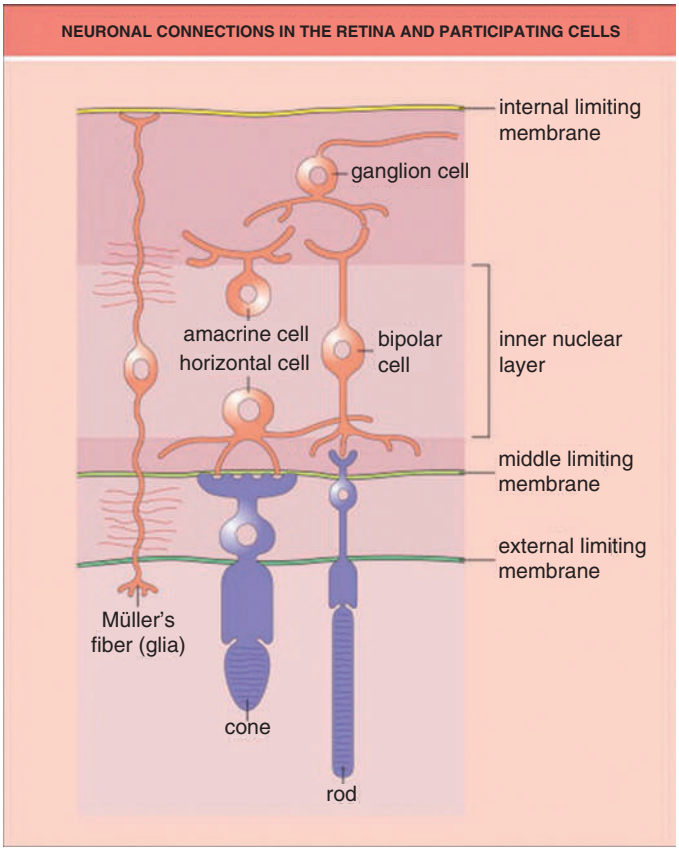


FIG 46.6 Neuronal connections in the retina and participating cells. The inner nuclear layer contains the nuclei of the bipolar cells (second neuron) and Müllerian glia. The amacrine cells are found on the inside and the horizontal cells on the outside of this layer, next to their respective plexiform connections. (From Schubert HD: *Ophthalmology*, ed 4, Philadelphia, 2014 Saunders.)

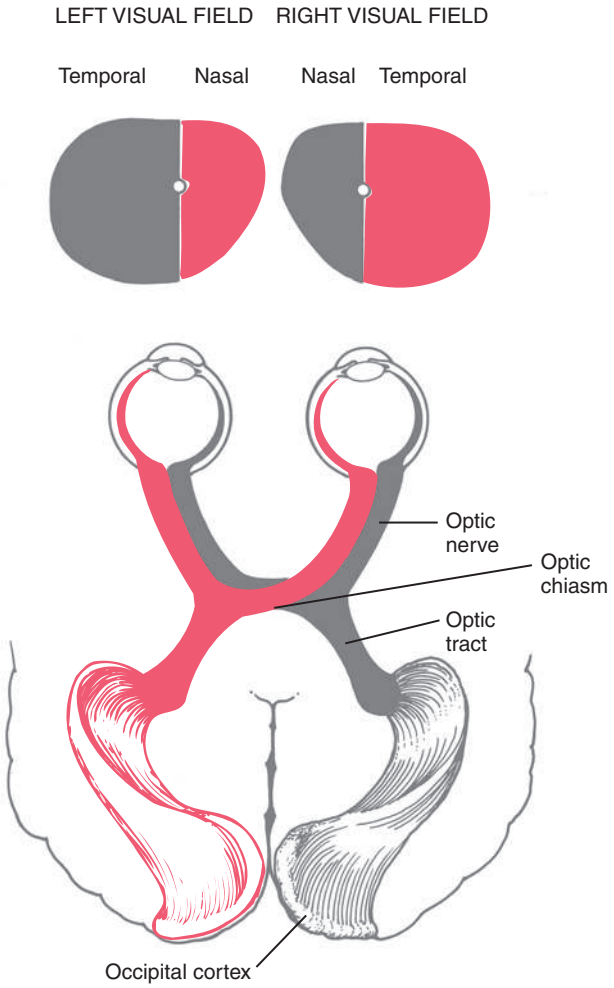


FIG 46.7 Visual pathways. (From Jarvis C: *Physical examination and health assessment*, ed 7, Philadelphia, 2016, Saunders, p 285.)

and reversed images, and at the optic chiasm just anterior to the pituitary gland, half of the nerve fibers cross over to the other side of the brain (Fig. 46.7). The left optic tract contains fibers from the left half of each retina, and the right optic tract contains fibers from the right half of each retina. The nerve impulse travels through the optic nerves to connections in the thalamus and finally connects with the neurons of the occipital cortex.

GENERAL MANIFESTATIONS OF VISUAL IMPAIRMENT

Visual impairment may occur or become evident at any time during the life span. If these impairments occur during infancy or early childhood and are not immediately detected and managed, vision may not develop normally (see Pediatric Considerations: Development of Newborn Vision). In older children, academic performance may suffer. In adults and elderly individuals, poor eyesight affects activities of daily living and can limit an individual's ability to function normally and meaningfully in the environment. The various effects of aging on the eye can be seen in Geriatric Considerations: Changes in the Eyes.

Clues that may indicate a visual impairment include squinting, closing one eye, tilting the head, having redness of the eye or excessive tearing, and eye rubbing. These signs are especially helpful in identifying children who may be unable to verbalize visual difficulties. In older children and adults, complaints of blurred vision, halos, "floaters" in the visual fields, headaches, and eye pain may indicate a visual impairment. A thorough health history and physical examination, along with *visual acuity testing* and *ophthalmoscopic examination,* will provide health care providers with the necessary information to appropriately treat or refer individuals with visual impairment.

DISORDERS OF THE EYE

Errors of Refraction

Focusing a clear image on the retina is essential for good vision. In a normal eye, light rays enter the eye and are focused into a clear, upside-down image on the retina. The brain can easily "right" the upside-down image in conscious perception, but cannot correct an image that is not sharply focused (Fig. 46.8).

Myopia, Hyperopia, Presbyopia, and Astigmatism

If the eye is elongated, the image focuses in front of the retina rather than on it. The retina receives only a fuzzy image. This condition, called *myopia* or nearsightedness, can be corrected with concave contact lenses or glasses. Frequently seen in late childhood or early adolescence, individuals with myopia are unable to see distant objects clearly.

If the eye is shorter than normal, the image focuses behind the retina, also producing a fuzzy image. This condition, called *hyperopia* or farsightedness, can be corrected with convex lenses. *Presbyopia* is the

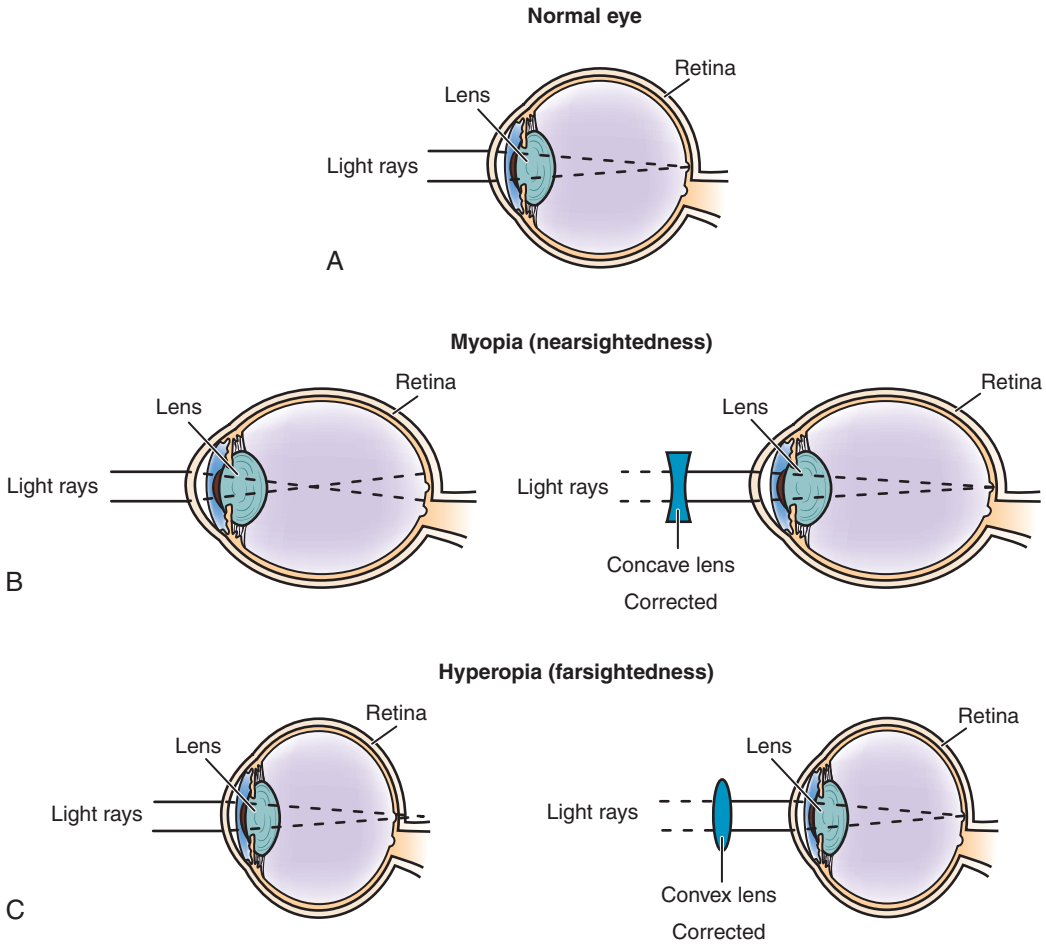


FIG 46.8 A, Light rays are focused to produce a clear visual image (emmetropia). **B,** In myopia, light rays are focused in front of the retina. A concave lens moves the focus back onto the retina and results in a clear image. **C,** In hyperopia, light rays are focused behind the retina. A convex lens moves the focus forward so that the light rays fall directly on the retina.

loss of accommodative capacity. This inability to see near objects clearly occurs most commonly in middle age and is frequently corrected with reading glasses. An irregularity in curvature of the cornea or lens, termed *astigmatism,* produces a distorted image and is corrected with glasses or contact lenses that are formed with the opposite curvature.

Age-Related Disorders

Strabismus

To make visual perceptions meaningful, the visual images in the two eyes normally fuse with each other on corresponding points of the two retinas. Strabismus, also called *squint* or *cross-eyedness,* is a condition of ocular misalignment. It results from an abnormality of the neuromuscular control of the eyes. The eyes appear misaligned on examination. Symptoms include squinting and frowning when reading, closing one eye to see, having trouble picking up objects, being dizzy, and having headaches. Strabismus is often caused by an abnormal "set" of the fusion mechanism of the visual system and is most commonly found in children, affecting about 4% of the population. Studies have shown that advanced maternal age, cigarette smoking during pregnancy, and low birth weight contribute to the incidence of strabismus. There also may be a genetic link predisposing children to certain types of strabismus.

In the early efforts of the child to fixate the two eyes on the same object, one of the eyes fixates satisfactorily but the other fails to fixate, or both eyes fixate satisfactorily but never simultaneously. Soon the patterns of conjugate movements of the eyes become abnormally set so that the eyes never fuse.

Treatment of strabismus includes occlusion therapy, or patching of the good eye to force use of the weak eye; use of corrective lenses; surgery on the eye muscles; use of prisms; and exercises for the eye. If management of strabismus is begun before 24 months of age, amblyopia may be prevented.

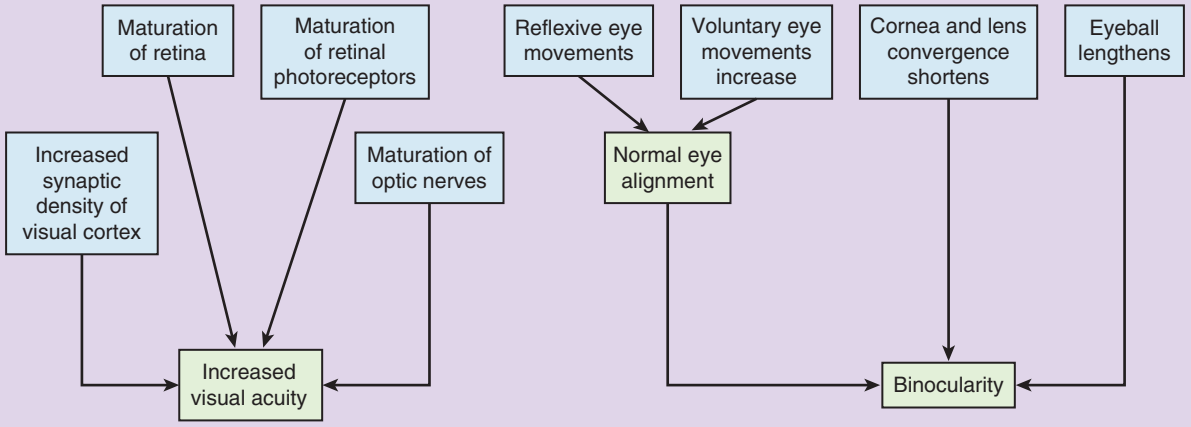
Amblyopia

Amblyopia is poor vision, even with the proper optical correction, in one or both eyes. It is the most common cause of decreased vision in the pediatric population, affecting 1% to 4% of children. It results from altered visual development despite normal-appearing retinal and optic nerve pathways. Amblyopia occurs when the normal course of visual development is interrupted, such as when visual images do not fuse, as in the case of untreated strabismus, or when severe refractive errors are present. Conditions such as congenital cataracts, uncorrected astigmatism, and other errors of refraction may also interfere with visual development.

The diagnosis of amblyopia is confirmed when a complete ophthalmologic examination reveals a decrease in visual acuity that cannot be explained by organic causes. Although screening for it is much easier in older children, treatment is more difficult at this point. Therefore screening must take place at an early age. Successful management of

PEDIATRIC CONSIDERATIONS

Development of Newborn Vision

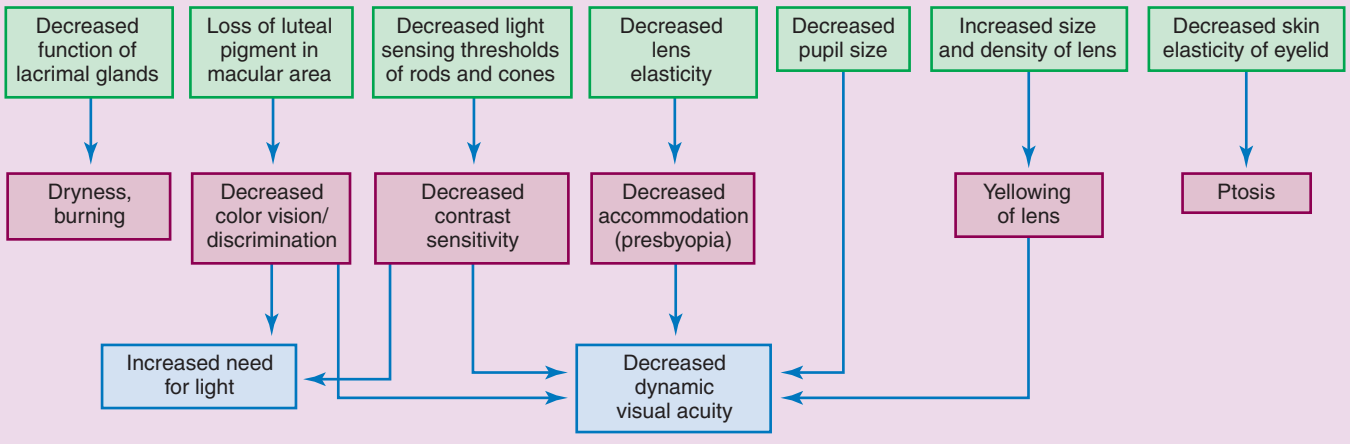


At birth, the newborn's visual acuity is 20/400. By 6 months of age, the infant's visual acuity is 20/30 because of increased synaptic density of the visual cortex and maturation of the retina, retinal photoreceptors, and optic nerves. The newborn has several reflexive eye movements such as the blink reflex, corneal reflex, and pupillary reactions to light. The newborn is unable to coordinate the head and eyes at birth, so the doll's eyes reflex is present. As the infant begins to explore the environment visually, the voluntary eye movements increase. The reflexive and voluntary movements begin to coordinate to create normal eye alignment.

Eyes must be close to alignment for vision. If strabismus (eyes not properly aligned) occurs, it can lead to blindness. The shapes of the eyeball, cornea, and lens change as the infant grows. The eyeball is less spherical than the adult eye and is too short for the lens. As the eyeball lengthens, convergence created by the cornea and the lens decreases, which helps create binocularity. Binocularity is perceived vision from both eyes simultaneously; it begins to develop at age 6 weeks and is established at 4 months. Stereopsis, or depth perception, develops by 7 months of age.

GERIATRIC CONSIDERATIONS

Changes in the Eyes



Aging affects all parts of the eye. Minor changes include decreased skin elasticity, alterations in lacrimal gland function, and shrinking of the vitreous body. The changes in the lens and retina of the eye are more significant. These changes cause decreased color vision and discrimination, reduced contrast sensitivity, and diminished accommodation. As a result, the elderly need brighter light to see and do not differentiate color well. The elderly also have less dynamic visual acuity.

The retina is affected by a loss of the luteal pigment in the macular areas, as well as reduced light-sensing thresholds of the rods and cones. These changes lead directly to a slowing of dark adaptation and a decrease in the ability to discern brightness and colors, particularly shorter light wavelengths such as blues and greens.

A primary change in the aging eye is the development of presbyopia. Presbyopia is caused by a decrease in the elasticity of the lens and a decrease in the effectiveness of the ciliary muscle, which lead to an inability to focus on near objects.

amblyopia depends on several factors. The most important is the age of onset and the length of time between onset and the commencement of treatment. Management of amblyopia includes the use of atropine to blur vision or patching of the "stronger" eye. This forces the brain and weaker eye to work together to stimulate vision.

Cataracts

Cataracts are a clouding or opacity of the lens that leads to gradual, painless blurring of vision and eventual loss of sight. By age 80, 50% to 70% of Americans have cataracts. They are more common in Caucasians than in African Americans or Hispanics. Cataracts result from the process of aging (senile), trauma (causing lens rupture and swelling), congenital factors (Down syndrome, intrauterine rubella infection), metabolic disease (diabetes mellitus, hypoparathyroidism), and certain medications (systemic or inhaled corticosteroids). Cigarette smoking and heavy alcohol consumption may also increase the risk of cataract formation.

Ocular tissues are continuously subject to a variety of physical, chemical, and environmental insults that generate reactive oxygen species/oxidative stress and create a damaging environment. These tissues are also relatively avascular in nature, limiting the effectiveness of the body's normal antioxidative mechanisms. Both eyes may be affected, but at different rates. Patients with cataracts may experience increased glare at night, blurred vision, and altered color perception. Persons with opacity in the central portion of the lens can generally see better in dim light when the pupil is dilated. The degree of visual loss corresponds to the density of the cataract.

A diagnosis of cataracts can be made through examination of the eye with an ophthalmoscope or slit lamp. As the cataract worsens or "matures," visualization of the retina becomes increasingly difficult until finally the pupil appears white and the retina cannot be visualized at all. Treatment for cataracts involves surgical removal and replacement of the lens. This procedure is completed on an outpatient basis, with individuals returning home immediately after surgery.

Retinopathy

Retinopathy is any disorder of the retina. Damage to the retina impairs vision because even a well-focused image cannot be perceived if some or all of the light receptors do not function properly. Retinopathies can result from a variety of causes, the most common being trauma and vascular disease, especially in individuals with diabetes mellitus and hypertension.

Retinal Detachment

Detachment of the retina is usually spontaneous but may be secondary to trauma such as sudden blows to the head. Eye tumors, myopia, and cataract extraction are other common predisposing factors; however, detachments may also occur spontaneously. Retinal detachments are classified into three categories. Exudative (or serous) detachments result from accumulation of serous or hemorrhagic fluid in the subretinal space, generally due to hydrostatic factors (e.g., severe, sudden hypertension), inflammation (sarcoidosis), or neoplastic effusions. The second type, tractional retinal detachment, occurs when mechanical forces on the retina caused by fibrosis and scarring pull it away from the underlying epithelium (injury or surgery to the eye). This type is most commonly caused by diabetic retinopathy. The third type of retinal detachment is spontaneous or rhegmatogenous. As individuals age, the vitreous humor shrinks and traction develops, causing separation.

Tearing of the retina allows vitreous fluid to flow behind the retina and cause traction and progressive detachment (Fig. 46.9). The area of detachment increases rapidly, and visual loss is progressive. Common manifestations of retinal detachment include the sudden appearance of floating spots that may decrease over a period of weeks and odd flashes of light that appear when the eye moves. Other symptoms include blurring of vision in a single eye that appears as though "a curtain is being pulled down over the eye." If untreated, the retina may detach entirely and result in total blindness in the affected eye. However, if diagnosed and treated early, permanent vision loss can be prevented. Retinal detachments may also cause vitreous hemorrhage.

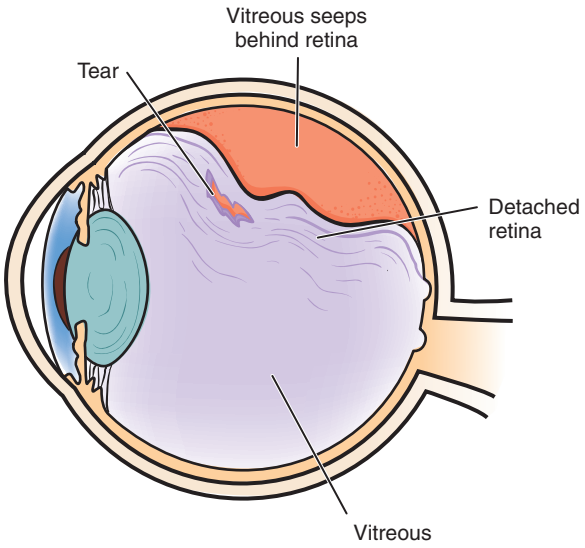


FIG 46.9 Retinal detachment.

Retinal detachments are diagnosed through ophthalmoscopic examination. The retina appears to hang in the vitreous humor like a gray cloud. One or more retinal tears, generally crescent-shaped, are usually present. Management of retinal detachment is aimed at closing tears in the retina and positioning the fragments of the retina so that reattachment can occur.

Diabetic Retinopathy

Etiology and pathogenesis. Diabetic retinopathy is one of the most common complications of diabetes, afflicting about 20% of adults with the disease. Symptoms increase in prevalence and severity with increasing duration of illness and poorer control of blood glucose levels. Diabetic retinopathy is a disease of the vasculature of the retina. In diabetes, the retinal capillary becomes diseased; it loses the ability to transport red blood cells and thus oxygen and nourishment to the retina, with consequent tissue hypoxia and ischemia. Chronic hyperglycemia and ischemia create altered energy and metabolic homeostasis. This and other detrimental factors such as oxidative stress promote inflammation, apoptosis, and increased production of vascular endothelial growth factors. The end results are vascular and neuronal degeneration, vascular leakage, diabetic macular edema, and retinal neovascularization.

Diabetic retinopathy can be divided into two categories: nonproliferative and proliferative. In nonproliferative diabetic retinopathy, retinal veins become dilated and microaneurysms develop. This effect is a result of damaged vascular epithelium. Small retinal hemorrhages and cotton-wool spots (infarctions in the nerve fibers) occur. Early in the process, visual changes may be minimal or resolve after a few days. As the disease progresses, retinal edema occurs. If the edema involves the macular area, visual acuity is noticeably affected.

Proliferative diabetic retinopathy is characterized by the development of new but abnormal blood vessels (neovascularization) caused by the loss of retinal blood flow and ischemia. These new vessels affect vision in two ways: first, because they are abnormal, they are prone to leakage

of blood into the vitreous cavity and may thus result in vitreous hemorrhage. Second, the vessels firmly attach themselves to the retina and grow out into the vitreous humor. The subsequent traction on the retina increases the risk for retinal detachment.

Clinical manifestations. Diabetic retinopathy is associated with complaints of blurred, darkened, and distorted vision. Visual changes may fluctuate in severity. Some individuals complain of being unable to read or have vague changes in vision.

Diagnosis and treatment. The diagnosis of diabetic retinopathy is made through careful history taking, visual acuity testing, and ophthalmologic examination and retinal angiography. The most important factor in the management of diabetic retinopathy is prevention. Intensive blood glucose level control and blood pressure management have been shown to slow the progression or reduce the risk of developing diabetic retinopathy. Laser treatments are also used to prevent further vision loss. Because the retina is nervous system tissue, it does not regenerate efficiently. Therefore treatment may prevent any further injury to eye tissue, but it cannot restore vision.

Management of proliferative diabetic retinopathy must be instituted as soon as possible to prevent blindness. Surgical intervention and laser procedures are used in conjunction with the measures used for nonproliferative retinopathy. Because of the risk of diabetic retinopathy, it is recommended that individuals with diabetes mellitus have annual ophthalmologic examinations.

Age-Related Macular Degeneration

Etiology and pathogenesis. Age-related macular degeneration (AMD) is the leading cause of blindness among older adults in the developed world. The exact causes of macular degeneration are unknown, but the outcome is bilateral progressive macular deterioration with central vision loss. Risk factors for developing AMD include age, female gender, history of cigarette smoking, family history of AMD, increased serum cholesterol level, cardiovascular disease, hypertension, obesity, and previous cataract surgery. Degeneration of the retinal pigment epithelium linked to oxidative stress and inflammatory chemicals appear to be key factors in development of the disorder. There is a reduced ability for oxygen to diffuse through to the retinal pigment epithelium and photoreceptors, resulting in release of growth factors and cytokines, which stimulate growth of vessels into the subretinal space. These new vessels leak, causing distortion and reduction of the clarity of central vision. Visual loss may also result from cell death and atrophy of the retinal pigment epithelium. AMD includes a wide spectrum of findings that can be divided into two subgroups. Manifestations, diagnosis, and management of each subgroup differ.

"Dry" or nonneovascular geographic atrophic AMD is the most common form, causing visual loss attributable to degeneration of the outer retina, the pigmented layer, and the choroidal layer. There are subretinal accumulations of cellular debris known as *drusen,* along with metabolic dysfunction of the retina. Hard drusen may be seen during ophthalmologic examination and appear as discrete yellow deposits on the retina. Atrophic AMD often affects just one eye initially but later develops in the unaffected eye. It causes a gradual decline in vision.

In "wet" neovascular or exudative AMD, visual loss is usually more rapid in onset and causes more severe visual disruption. Impairment of barrier function allows for subretinal fluid collections, which may cause retinal detachments and/or neovascularizations. These fluid buildups may be visualized on retinal examination.

Clinical manifestations. AMD is generally painless. In the atrophic form, the initial symptom is slightly blurred vision and decreased ability to see fine detail. Often patients need more light for completing fine tasks such as reading and needlework. As the disorder progresses, the area of central vision loss becomes larger and darker.

Exudative AMD may also be manifested by a progressive blurring of vision. A hallmark of this form of AMD is the wavy appearance of straight lines. This occurs because of distortion of the retina from fluid accumulations behind it. Vision may be lost rapidly or occur suddenly in previously undiagnosed patients attributable to retinal detachment or hemorrhage.

Diagnosis and treatment. AMD is diagnosed with thorough history and physical examination to rule out other causes of visual loss, visual acuity testing, dilated retinal examination, and use of the Amsler grid (Fig. 46.10). If AMD is suspected, fluorescein angiography may be completed by the ophthalmologist. In this examination, fluorescein dye is injected into the patient. Photos of the retina show characteristic changes of the choroidal vascular layer.

Management of AMD depends on the type and severity of the disease. Antioxidant/zinc vitamin supplementation has been shown to slow/delay the progression of AMD. Other treatment modalities are aimed at correcting the vascular changes and include laser photocoagulation and photodynamic therapy. Research is currently focusing on a variety of antiinflammatory therapies and agents that block neovascularization such as anti–vascular endothelial growth factor monoclonal antibodies. As with all progressive diseases, patients with AMD should have regular comprehensive eye examinations and daily self-evaluation using the Amsler grid.

Glaucoma

Glaucoma is characterized by progressive loss of vision and usually an increased intraocular pressure. As fluid pressure inside the eye and against the retina increases, blood flow through the retina slows. Reduced blood flow causes degeneration of the retina and thus loss of vision. Glaucoma can be categorized into two main types: chronic open-angle and acute closed-angle (narrow-angle) (Fig. 46.11).

Glaucoma is more common in the elderly, African Americans, those with a family history, those with myopia, and individuals with diabetes. It may also occur as a result of trauma, inflammation, or exposure to corticosteroids.

Chronic Open-Angle Glaucoma

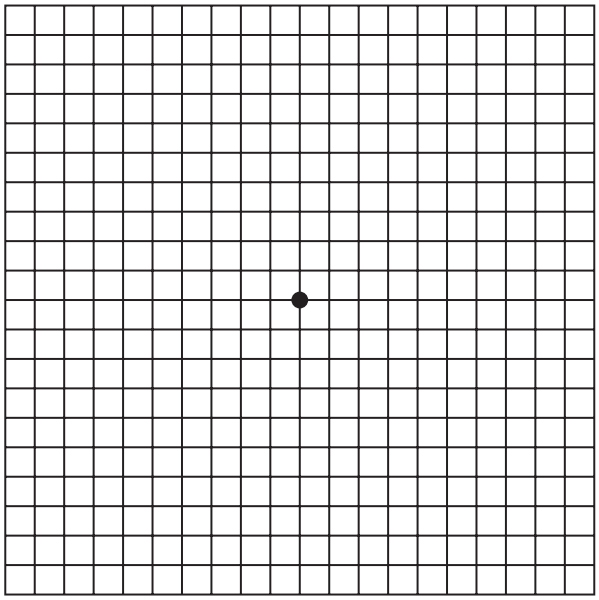
Etiology and pathogenesis. The cause of open-angle glaucoma is not clear. The drainage channels for aqueous humor appear normal. The disease is often bilateral and has a genetic component. Open-angle glaucoma accounts for the majority of all cases of glaucoma.

Open-angle glaucoma has an insidious onset with no symptoms in the early stages. However, the intraocular pressure is consistently elevated, and, over a period of months or years, symptoms including gradual loss of vision in the periphery resulting in tunnel vision appear. The mechanism is retinal ganglion cell apoptosis leading to axonal loss in the optic nerve. The optic disc becomes atrophic. Affected individuals may have complaints of vague but persistent dull eye pain or an inability to distinguish colors. Halos may appear around lights if the intraocular pressure is markedly elevated.

Diagnosis and treatment. The diagnosis of open-angle glaucoma is made through intraocular pressure measurement, ophthalmoscopic examination of the optic disc, and central visual field testing. Because of the insidious nature of the disorder, it is recommended that all individuals older than 35 years have an intraocular pressure measurement and ophthalmoscopic examination every 3 to 5 years. If a family history of glaucoma is present, more frequent examination is recommended.

Management of open-angle glaucoma is aimed at increasing drainage of aqueous humor and decreasing intraocular pressure. Prostaglandin analog eye drops (latanoprost 0.005%, bimatoprost 0.03%) are first-line agents. β-Adrenergic–blocking eye drops, such as timolol, are used to

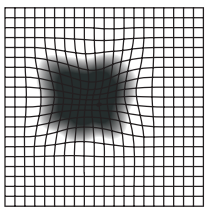
AMSLER GRID



The Amsler Grid is used to test for and monitor symptoms of Macular Degeneration (MD). It should not be depended upon for diagnosis.

Instructions:

1. Do not remove glasses or contact lenses normally used for reading.
2. Hold grid at eye level approximately 33cm away in a well lit room
3. Cover one eye and focus on the centre dot with the uncovered eye (make sure the eye is fully covered)
4. Repeat with the other eye



If the grid lines are wavy, broken or distorted, or if there are blurred or missing patches, this may be a symptom of MD. **Please see reverse side for what to do.**

Macular Disease Foundation AUSTRALIA

Our focus is your vision

1800 111 709

www.mdfoundation.com.au

(Reverse side)

Macular Degeneration (MD) is the leading cause of blindness and severe vision loss in Australia. It is primarily age related and referred to as Age-related Macular Degeneration (AMD).

Risk Factors include age (over 50 years), a direct family history and smoking.

The Amsler Grid should be used daily as per instructions to test for symptoms of MD which can include one or more of the following:

- Distortion where straight lines appear wavy or bent
- Dark patches or empty spaces appearing in the centre of your vision

In addition, key symptoms may also include:

- Distinguishing faces becomes a problem
- Difficulty with reading or any other activity that requires fine vision
- Increased need for illumination, sensitivity to glare, decreased night vision or poor colour sensitivity

What to Do

If you experience any of the above symptoms or any changes in your vision, immediate referral to an optometrist or ophthalmologist is critical. Never dismiss any changes in vision as part of getting older. Early detection can save your sight.

For information, guidance, understanding and support for you and your family contact the Macular Disease Foundation Australia.

1800 111 709

www.mdfoundation.com.au

Free information kit available.

FIG 46.10 *Upper,* What an Amsler grid looks like to an individual with normal vision. *Lower,* What an Amsler grid might look like to an individual with macular degeneration. (From Macular Degeneration Foundation. Available at: www.mdfoundation.com/au/resources/1/Amsler_Grid.pdf.)

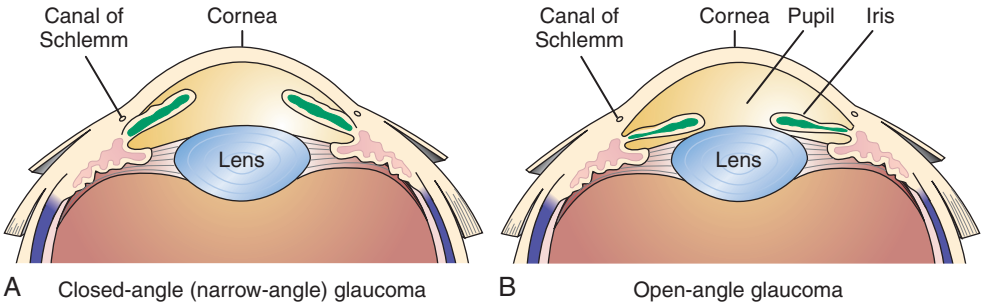


FIG 46.11 Closed-angle (narrow-angle) glaucoma compared with open-angle glaucoma. **A,** In closed-angle glaucoma, the outflow of aqueous humor is obstructed by the iris root of the dilated pupil. **B,** In open-angle glaucoma, the obstruction to outflow of aqueous humor is in the drainage canals.

help decrease intraocular pressure by decreasing aqueous humor production. Miotics such as pilocarpine are also useful in that they constrict the pupil and thereby stimulate the ciliary muscles to pull on the trabecular meshwork surrounding the canal of Schlemm to increase the flow of aqueous humor. If the intraocular pressure elevation persists or the optic nerve damage progresses despite treatment, laser surgery aimed at the trabecular meshwork may be done to lower intraocular pressure.

Acute Angle-Closure Glaucoma

Etiology and pathogenesis. Primary acute angle-closure glaucoma is caused by abnormality of the angle between the pupil and lateral cornea. This angle is narrow and blocks outflow of aqueous humor when the pupil is dilated. Angle-closure glaucoma is much less common than open-angle glaucoma but is more prevalent in the elderly, those with hyperopia, and Asian populations. This form of glaucoma has a rapid onset and is treated as an emergency. Angle-closure glaucoma is associated with pupillary dilation and thus might occur when an individual is sitting in a darkened room or during times of stress. Forward displacement of the iris toward the cornea with dilation narrows or closes the chamber angle, obstructing the outflow of aqueous humor (see Fig. 46.11). The rapid increase in intraocular pressure results in ischemic damage.

Manifestations of angle-closure glaucoma include severe eye pain, nausea and vomiting, blurred vision with halos around lights, redness of the eye, a steamy cornea, and a dilated pupil that is nonreactive to light.

Diagnosis and treatment. Diagnosis of angle-closure glaucoma involves the same tests as used for diagnosis of open-angle glaucoma. Treatment again in this case is aimed at decreasing intraocular pressure. Acutely, carbonic anhydrase inhibitors such as acetazolamide and miotics may be used to decrease aqueous humor production and blockage. Laser iridectomy usually results in a permanent cure.

Visual Field Deficits

Visual Field Loss

Etiology and pathogenesis. Visual field loss can be caused by changes in the eye itself, as is the case in cataracts, or result from tumors, vascular lesions, and demyelinating lesions near or in the neural pathways of the retina, optic nerve, or visual cortex of the brain.

Damage to the visual pathway does not always result in a total loss of vision. Depending on where the damage occurs, only part of the visual field may be affected (Fig. 46.12). Monocular field loss indicates disease of the retina or optic nerve. For example, a certain form of neuritis often associated with multiple sclerosis can cause loss of only the center of the visual field, called a *scotoma.*

Damage or lesions may also cause bilateral visual field losses or loss of half of the visual field, called *hemianopsia*. Lesions of the optic chiasm, usually caused by pituitary tumors, characteristically produce a bitemporal hemianopsia. Lesions occurring behind the optic chiasm cause a homonymous hemianopsia, which is a visual field loss involving the same side in both eyes. The more posterior the lesion in the visual pathway, the more congruous (similar size, shape, location) are the defects in the two eyes. Cerebrovascular accidents (strokes) and tumors are responsible for most of these lesions.

Diagnosis and treatment. Visual field deficits are easily and rapidly assessed through confrontation (i.e., comparison of the patient's vision to the examiner's vision). Visual field deficits should be suspected if patients demonstrate one-sided neglect of their environment or eye deviation toward the side of the lesion. Treatment for visual field loss includes managing the underlying cause (tumor removal), adapting the patient's environment, and teaching compensatory techniques.

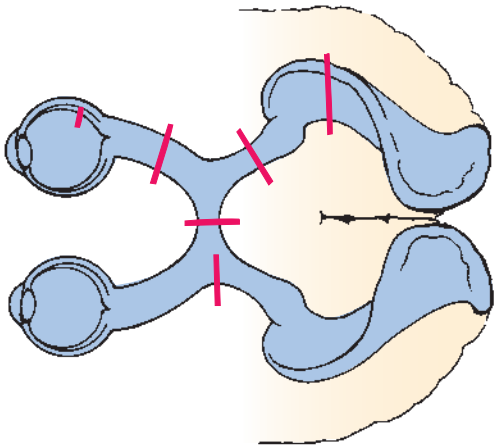
INTERVENTIONS FOR INDIVIDUALS WITH VISION IMPAIRMENT

Once the visual impairment of an individual has been thoroughly investigated, specific interventions may be prescribed. Interventions may be classified into three general categories: assistive devices, environmental adaptations, and behavioral techniques. Proper care and cleaning of contact lenses and eyeglasses directly influences the effectiveness of the prosthesis. Tinted lenses are generally available and may be effective in reducing glare for some individuals. Pocket magnifiers are frequently useful for persons with an acuity impairment. Large print is now available on many household items (e.g., watches, playing cards, telephones, books), and various textures are used in further modifications for the visually impaired.

An unchanging, structured environment where items are kept in fixed locations familiar to the visually impaired person promotes safety and independence. Attempts to structure temporary environments, such as by introducing personal items into a hospital room, might yield positive results if consistently considered by the staff. Attention to adequate lighting, glare reduction, and appropriate use of contrasting colors enhances safety and independent function of those with visual impairment.

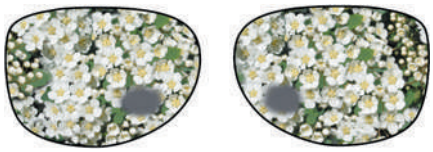
Behavioral techniques for the health care professional and visually impaired individuals can promote client comfort, safety, and independence. Such techniques for the professional include announcing oneself at all interactions and explaining sensory occurrences. Encouraging independence and social interaction often benefits individuals inasmuch as they may experience anger, frustration, or changes in self-concept as a result of their visual deficit.

Visually impaired individuals may be taught to wait several minutes for changes in dark–light adaptation and to avoid abrupt changes in

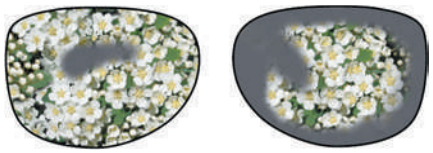


1. Retinal damage
   - Macula—central blind area (e.g., diabetes):

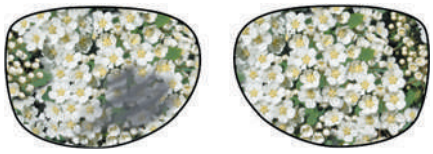
   - Localized damage—blind spot (scotoma) corresponding to particular area:



   - Increasing intraocular pressure—decrease in peripheral vision (e.g., glaucoma). Starts with paracentral scotoma in early stage:



   - Retinal detachment. Person has shadow or diminished vision in one quadrant or one half of visual field:



2. Lesion in globe or optic nerve. Injury here yields one blind eye, or unilateral blindness:

3. Lesion at optic chiasm (e.g., pituitary tumor)—injury to crossing fibers only yields a loss of the nasal part of each retina and a loss of both temporal visual fields. Bitemporal (heteronymous) hemianopsia:

4. Lesion of outer uncrossed fibers at optic chiasm (e.g., aneurysm of left internal carotid artery exerts pressure on uncrossed fibers). Injury yields left nasal hemianopsia:

5. Lesion of right optic tract or right optic radiation. Visual field loss in right nasal and left temporal fields. Loss of same half of visual field in both eyes is homonymous hemianopsia:

FIG 46.12 Visual field losses. (From Jarvis C: *Physical examination and health assessment,* ed 7, Philadelphia, 2016, Saunders p. 318.)

lighting. Individuals should be discouraged from looking directly into bright lights to reduce glare. Assessment of the visually impaired patient's ability to summon help in the health care and home settings is advised.

KEY POINTS

- Visual acuity depends on the formation of discrete patterns of light on the retina. Errors of refraction such as myopia and hyperopia cause light to focus in front of or behind the retina, respectively. Irregular curvature of the cornea results in astigmatism. These disorders are correctable with lenses to refract the light to the appropriate retinal location.
- Strabismus occurs when both eyes do not focus together to form a single image. If this is not corrected, the brain may ignore the image from one eye in an attempt to avoid double imaging; eventually, amblyopia can result.
- Cataracts are due to opacification of the lens that blocks and scatters light. Cataracts may be congenital, traumatic, or associated with aging.
- Retinopathy is any disorder affecting the retina. Trauma and systemic disorders such as diabetes mellitus, hypertension, and vascular disease are the most common causes of retinopathy.
- Retinal detachment is characterized by tearing of the retina away from the choroid layer of the eye, with seepage of vitreous humor behind the retina causing further detachment.
- Diabetic retinopathy is a disorder of the retinal vessels characterized by the formation of microaneurysms and hemorrhage (nonproliferative) or neovascularization and subsequent leakage and retinal detachment (proliferative).
- Macular degeneration is an age-related, progressive loss of central vision attributable to atrophic or exudative changes to the macula of the retina.
- Glaucoma occurs when intraocular pressure is increased by a decrease in the outflow of aqueous humor from the anterior chamber of the eye. Open-angle glaucoma has an unclear cause because no clear obstruction impedes the outflow of aqueous humor. Angle-closure glaucoma occurs when the angle between the pupil and lateral aspect of the cornea is narrow and blocks outflow when the pupil is dilated.
- Visual field losses are caused by lesions anywhere along the visual pathways. The location of the lesion determines monocular or binocular involvement and the portion of vision lost.

SMELL AND TASTE

The senses of smell and taste allow separation of noxious or even lethal agents from those that are desirable. The sense of smell has a protective function in signaling danger: animals use smell to recognize the proximity of other animals, and humans use smell to sense harmful substances, such as smoke or spoiled food items, in the environment. The sense of taste allows a person to select food in accordance with desire and perhaps also in accordance with tissue needs. Both senses are strongly tied to primitive emotional and behavioral functions of the nervous system. These chemical senses are interrelated and will be discussed together.

Nerve fibers of the olfactory system have their cell bodies in the mucous membrane of the upper and posterior parts of the nasal cavity. The sense of smell begins with chemical stimulation of these cells. Axons of these receptor cells pass through the cribriform plate and travel to the olfactory area of the cortex through the first cranial nerve. These nerves lie under the frontal lobes of the brain. It has been long thought that olfactory impulses reach the cerebral cortex without relay through the thalamus, making olfaction unique among the sensory systems. However, newer pathways have been identified that pass through the thalamus and are thought to aid in the conscious analysis of odor.

Like stimuli for smell, stimuli for taste are chemical. Food particles dissolved in fluid stimulate sensory receptors (taste buds) located on the surface of the tongue and in lesser density on the palate, tonsillar pillars, the epiglottis, and even proximal esophagus. Stimulation from the sensory receptors is conducted through the cranial nerves of taste (VII, IX, X) to connections in the brainstem and thalamus with eventual termination in the gustatory cortex in the parietal lobe. The gustatory sensory receptors have a heightened sensitivity for one of the primary taste sensations (sweet, salty, sour, or bitter); however, they can respond to a variety of stimuli. The number of sensory receptors for taste diminishes with age, often affecting nutritional status.

Disorders of Smell and Taste

Etiology and pathogenesis. Olfactory disorders range from loss or reduction in the sense of smell to distortions and olfactory hallucinations. Commonly the sense of smell is diminished in those who smoke and in individuals with conditions involving congestion and swelling of the nasal mucosa, such as allergies and sinusitis. Head trauma often results in the loss of smell because of actual shearing of the neuronal fibers as they traverse the cribriform plate. Tumors and large cerebral aneurysms of the anterior cerebral and anterior communicating arteries are lesions capable of diminishing olfactory sense. Epilepsy and psychiatric disorders may be associated with olfactory hallucinations. Most recently, decreased sense of smell has been identified as an early indicator of neurodegenerative disorders such as Parkinson disease and Alzheimer dementia.

A decreased gustatory sense can also result from heavy smoking, as well as extreme dryness of the tongue and mucous membranes. A variety of medications are known to alter the sense of taste, including certain antidepressant, antithyroid, antirheumatic, and anticancer medications. In addition, influenza-like illnesses and lesions on the thalamus and parietal lobe may impair taste sensation.

Clinical manifestations. Individuals with smell dysfunction frequently complain of a diminished ability to taste. They may experience a decreased appetite and use excessive amounts of salt, sugar, or other seasonings on their foods. These individuals may stop reacting to strong smells and not notice their own body odor. Smell dysfunction increases the risk of accidents in that these individuals may not detect signs of imminent danger such as gas or smoke. In addition, spoiled food may be ingested, and excessive use of salt is associated with health risks.

Diagnosis and treatment. Assessment of the sense of smell is done by asking the individual to smell different known odors while keeping the eyes closed. Irritating substances such as ammonia should be avoided because they stimulate the trigeminal nerve. Assessment of gustatory sense should include the primary taste sensations in appropriate areas of the tongue, with the surface of the tongue wiped clean between substances. Questions regarding weight loss and appetite add valuable information to the assessment data.

Interventions for those with smell and taste dysfunction focus on augmenting the stimulus, teaching the individual to rely on other senses, and changing the environment. Because their senses of smell and taste are unreliable in identifying spoiled foods, people with these dysfunctions are encouraged to adhere to a strict schedule for discarding leftovers and be aware of expiration dates on food products. Because significant nutritional problems may occur, it is important to educate and monitor the individual's diet. The creative use of seasonings and spices along with variations in the texture and presentation of food may enhance appetite. Individuals with taste impairments are encouraged to avoid blended foods and to practice frequent oral hygiene. Smoke detectors should be installed in all rooms where smell-impaired individuals sleep, and fire safety should be emphasized.

KEY POINTS

- The senses of smell and taste result from chemical stimulation of specialized nerve fibers located in the nose and the tongue and are closely related to each other. Nerve impulses travel through the cranial nerves to separate areas of the brain.
- Changes in smell and taste most commonly result from smoking and inflammation caused by colds, sinusitis, or allergies. A change in smell or taste sensation in the absence of an obvious cause may indicate a brain tumor and should prompt a thorough neurologic evaluation.

SUMMARY

Humans interact with their environment by means of the special senses of hearing, vision, smell, and taste. Through a variety of stimuli, including chemicals, light, and sound, individuals are able to enjoy everything from a symphony performance to a hot fudge sundae. The special senses also protect individuals from harm by allowing the perception of smoke or alarms. Only when these special senses are impaired does their importance become apparent.

Loss of these senses may result from congenital conditions, trauma, tumors, illness, or unknown causes. Loss may also be a consequence of aging. The mechanism of impairment may be a disruption of the mechanical aspect of the special sense, as in the obstruction of sound waves from cerumen impaction in the ears, or may be a neurologic event, as in the occurrence of homonymous hemianopsia after a stroke. Regardless of the cause of loss of the sense of hearing, vision, smell, or taste, prompt intervention and treatment can make tremendous differences in outcome. The loss may be totally corrected, or its progression may be slowed. Treatment may be aimed at the underlying cause of the sensory impairment or at altering the individual's behavior or environment to maximize the remaining function. In working with individuals who have alterations in special sensory function, health care professionals have the opportunity to make a great difference in that person's quality of life.

RESOURCES

Hearing

American Academy of Otolaryngology-Head and Neck Surgery. *What you should know about otosclerosis*. Available at: www.entnet.org/content/what-you-should-know-about-otosclerosis. (Accessed 21 November 2015).

Brown KD, Banachi V, Selesnick SH: Diseases of the external ear. In Lalwani AK, editor: *Current diagnosis & treatment in otolaryngology—head & neck surgery*, ed 3, New York, 2012, McGraw-Hill, (Chapter 47). http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=386&Sectionid=39944089. (Accessed 21 November 2015).

Centers for Disease Control. National Institute for Occupational Safety and Health. *Noise and hearing loss prevention*. Available at: www.cdc.gov/niosh/topics/noise/stats.html. (Accessed 21 November 2015).

Chiarella G, Petrolo C, Cassandra E: The genetics of Meniere's disease. *Appl Clin Genet* 9, 2015. doi:10.2147/tacq.s59024.

Driscoll CW, Carlson ML: Otosclerosis. In Lalwani AK, editor: *Current diagnosis & treatment in otolaryngology—head & neck surgery*, ed 3, New York, 2012, McGraw-Hill, (Chapter 51). http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=386&Sectionid=39944093. (Accessed 30 November 2015).

Foster C: Optimal management of Meniere's disease. *Ther Clin Risk Manag* 301:2015. doi:10.2147/tcrm.s59023.

Foster CA, Breeze RE: The Meniere attack: An ischemia/reprofusion disorder of the inner ear sensory tissues. *Med Hypotheses* 81:1108–1115, 2013.

Friedman NR, Scholes MA, Yoon PJ: Ear, nose, & throat. In Hay WW, Jr, Levin MJ, Deterding RR, Abzug MJ, editors: *Current diagnosis & treatment: pediatrics*, ed 22, New York, 2013, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=1016&Sectionid=61597764. (Accessed 21 November 2015).

Johnson J, Lalwani AK: Vestibular disorders. In Lalwani AK, editor: *Current diagnosis & treatment in otolaryngology—head & neck surgery*, ed 3, New York, 2012, McGraw-Hill, (Chapter 56). http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=386&Sectionid=39944099. (Accessed 21 November 2015).

Lalwani AK: The aging inner ear. In Lalwani AK, editor: *Current diagnosis & treatment in otolaryngology—head & neck surgery*, ed 3, New York, 2012, McGraw-Hill, (Chapter 53). http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=386&Sectionid=39944096. (Accessed 21 November 2015).

Lustig LR, Schindler JS: Ear, nose & throat disorders. In McPhee SJ, Papadakis MA, Rabow MW, editors: *Current medical diagnosis & treatment 2015*, New York, 2011, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=1019&Sectionid=5766800. (Accessed 21 November 2015).

Milligan S, McCrery S: Should children with acute otitis media routinely be treated with antibiotics? No: most children greater than two years do not require antibiotics. *Am Fam Physician* 88(7):2013. online.

Minovi A, Dazert S: Diseases of the middle ear in childhood. *GMS Curr Top Otorhinolaryngol Head Neck Surg* 13:2014. doi:10.3205/cto000114. Doc11.

Nidcd.nih.gov. *It's Important to Have Your Baby's Hearing Screened [NIDCD]*. 2015. Available at: http://www.nidcd.nih.gov/health/hearing/pages/screened.aspx. (Accessed 21 November 2015).

Nidcd.nih.gov. *Noise-Induced Hearing Loss*. 2015. Available at: http://www.nidcd.nih.gov/health/hearing/pages/noise.aspx. (Accessed 21 November 2015).

Parker A, Cross S, Jackson I, et al: The goya mutation identifies distinct novel roles for MAP3K1 in cochlear sensory hair cell development and survival. *Dis Model Mech* 2015. doi:10.1242/dmm.023176.

Quesnel A, Seton M, Merchant S, et al: Third-Generation Bisphosphonates for Treatment of Sensorineural Hearing Loss in Otosclerosis. *Otol Neurotol* 33(8):1308–1314, 2012. doi:10.1097/mao.0b013e318268d1b3.

Rettig E, Tunkel D: Contemporary concepts in management of acute otitis media in children. *Otolaryngol Clin North Am* 47(5):651–672, 2014. doi:10.1016/j.otc.2014.06.006.

Ropper AH, Samuels MA, Klein JP: Deafness, dizziness, and disorders of equilibrium. In Ropper AH, Samuels MA, Klein JP, editors: *Adams & Victor's principles of neurology*, ed 10, New York, 2014, McGraw-Hill, (Chapter 15). http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=690&Sectionid=49251502. (Accessed 21 November 2015).

Saccomano S: Dizziness, vertigo, and presyncope. *Nurse Pract* 37(12):46–52, 2012. doi:10.1097/01.npr.0000422206.92550.5b.

The Mayo Clinic. *Hearing loss: Risk factors*. Available at: www.mayoclinic.org/diseases-conditions/hearling-loss/basics/risk-factors/con-20027684. (Accessed 21 November 2015).

Walling AD, Dickson GM: Hearing loss in older adults. *Am Fam Physician* 85(12):1150–1156, 2012.

Yong J, Wang D: Impact of noise on hearing in the military. *Mil Med Res* 2(1):6, 2015. doi:10.1186/s40779-015-0034-5.

Vision

Ah-kee EY, Egong E, Shafi A, et al: A review of drug-induced acute angle closure glaucoma for non-ophthalmologists. *Qatar Med J* 6, 2015. http://dx.doi.org/10.5339/qmj.2015.6.

CDC. *National Diabetes Statistics Report: Estimates of Diabetes and its Burden in the United States*. 2014 Available at: www.cdc.gov/diabetes/pdfs/data/2014-report-estimates-of-diabetes-and-its-burden-in-the-united-states.pdf. (Accessed 28 November 2015).

Chou R, Dana T, Bougatsos C: Screening for visual impairment in children ages 1-5 years: Update for the USPSTF. *Pediatrics* 127(2):e442, 2011. doi:10.1542/peds.2010-0442.

Cotter S, Varma R, Tarczy-Hornoch K, et al: Risk factors associated with childhood strabismus. *Ophthalmology* 118(11):2251–2261, 2011. doi:10.1016/j.ophtha.2011.06.032.

Feltgen N, Walter P: Rhegmatogenous retinal detachment- an ophthalmologic emergency. *Dtsch Arztebl Int* 111(1-2):12–22, 2014.

Fletcher EC, Chong N, Augsburger JJ, Corrêa ZM: Retina. In Riordan-Eva P, Cunningham ET, Jr, editors: *Vaughan & Asbury's general ophthalmology*, ed 18, New York, 2011, McGraw-Hill (Chapter 10). http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=387&Sectionid=40229327. (Accessed 21 November 2015).

Gupta V, Rajagopala M, Ravishankar B: Etiopathogenesis of cataract: An appraisal. *Indian J Ophthalmol* 62(2):103, 2014. doi:10.4103/0301-4738.121141.

Hall J, Guyton A: *Guyton and Hall textbook of medical physiology*, Philadelphia, 2016, Elsevier.

Kang J, Loomis S, Rosner B, et al: Author response: Comparison of risk factor profiles for primary open-angle glaucoma subtypes defined by pattern of visual field loss: True risk factors or arbituary definition? *Invest Ophthalmol Vis Sci* 56(11):6532, 2015. doi:10.1167/iovs.15-17944.

Ma J, Wang J, Zhang S: The unfolded protein response and diabetic retinopathy. *J Diabetes Res* 2014:1–14, 2014. doi:10.1155/2014/160140.

Mahmood S, Roberts S, Aslam T, et al: Routine versus as-needed bevacizumab with 12-weekly assessment intervals for neovascular age-related macular degeneration. *Ophthalmology* 122(7):1348–1355, 2015. doi:10.1016/j.ophtha.2015.03.017.

Motley W, Asbury T: Strabismus. In Riordan-Eva P, Cunningham ET, Jr, editors: *Vaughan & Asbury's general ophthalmology*, ed 18, New York, 2011, McGraw-Hill (Chapter 12).

National Eye Institute. *Cataracts defined*. Available at: https://nei.nih.gov/eyedata/cataract#1. (Accessed 21 November 2015). http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=387&Sectionid=40229329. (Accessed 21 November 2015).

Powers AC: Diabetes mellitus: complications. In Kasper D, Fauci A, Hauser S, et al, editors: *Harrison's principles of internal medicine*, ed 19, New York, 2015, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=1130&Sectionid=79753119. (Accessed 29 November 2015).

Salmon JF: Glaucoma. In Riordan-Eva P, Cunningham ET, Jr, editors: *Vaughan & Asbury's general ophthalmology*, ed 18, New York, NY, 2011, McGraw-Hill (Chapter 11). http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=387&Sectionid=40229328. (Accessed 30 November 2015).

Terluk M, Kapphahn R, Soukup L, et al: Investigating mitochondria as a target for treating age-related macular degeneration. *J Neurosci* 35(18):7304–7311, 2015. doi:10.1523/jneurosci.0190-15.2015.

Umapathy A, Donaldson P, Lim J: Antioxidant delivery pathways in the anterior eye. *Biomed Res Int* 2013:1–10, 2013. doi:10.1155/2013/207250.

Waxman SG: The visual system. In Waxman SG, editor: *Clinical neuroanatomy*, ed 27, New York, 2013, McGraw-Hill (Chapter 15). http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=673&Sectionid=45395980. (Accessed 28 November 2015).

Smell and Taste

Doty RL, Bromley SM: Disorders of smell and taste. In Kasper D, Fauci A, Hauser S, et al, editors: *Harrison's principles of internal medicine*, ed 19, New York, 2015, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=1130&Sectionid=79725514. (Accessed 30 November 2015).

47

Pain

Joni D. Marsh

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- How do the processes of transduction, transmission, perception, and modulation relate to the phenomenon of nociception?
- How is neurotransmission of pain signals modulated at the receptor, spinal cord, and brain?
- How do acute pain and chronic pain differ with regard to cause and clinical manifestations?
- Why are some painful sensations perceived at a distance from the site of injury (referred)?
- Why is it important to adequately manage pain?

CHAPTER OUTLINE

Physiology of Pain, 955
Transduction, 956
Transmission, 956
Perception, 958
Modulation, 958
TYPES OF PAIN, 960
Acute Pain, 961
Headache, 962
Etiology and Pathogenesis, 962
Clinical Manifestations, 962
Diagnosis and Treatment, 962
Chronic Pain, 962
Fibromyalgia Syndrome, 963
Etiology and Pathogenesis, 963
Clinical Manifestations, 963
Diagnosis and Treatment, 963
Cancer-Related Pain, 964
Neuropathic Pain, 964
Trigeminal Neuralgia, 965
Etiology and Pathogenesis, 965
Clinical Manifestations, 965
Diagnosis and Treatment, 965
Diabetic Neuropathy, 965
Etiology and Pathogenesis, 965
Clinical Manifestations, 965
Diagnosis and Treatment, 966
Postherpetic Neuralgia, 966
Etiology and Pathogenesis, 966
Clinical Manifestations, 966
Diagnosis and Treatment, 966
Ischemic Pain, 966
Referred Pain, 966
Physiologic Responses to Pain, 967
Pain in the Young and the Elderly, 967
TREATMENT MODALITIES, 967
Pharmacologic and Nonpharmacologic Pain Management, 967
Interrupting Peripheral Transmission of Pain, 968
Modulating Pain Transmission at the Spinal Cord, 968
Altering the Perception and Integration of Pain, 968

Pain is a complex physiologic and perceptual phenomenon. Because pain is very much a subjective experience, defining and assessing it are difficult. The International Association for the Study of Pain defines it as "an unpleasant sensory and emotional experience associated with actual or potential tissue damage or described in terms of such damage." McCaffery offered a clinically useful definition: "Pain is whatever the experiencing person says it is, existing whenever the experiencing person says it does." Accurate assessment and optimal management of pain are extremely important, not only because relief of pain and suffering is ethically desirable, but also because unrelieved pain is physiologically harmful. Studies have documented the benefits of adequate pain control on the rate of recovery, health care costs, and postoperative morbidity. In fact, The Joint Commission has developed standards of care regarding the assessment and management of pain that must be followed by all accredited agencies. Pain has been referred to as the *fifth vital sign*.

PHYSIOLOGY OF PAIN

The physiologic mechanisms involved in the pain phenomenon are termed *nociception*. Nociception can be divided into four stages:

transduction, transmission, perception, and modulation. Transduction is the process of converting painful stimuli to neuronal action potentials at the sensory receptor. Transmission refers to the movement of action potentials along neurons that make their way from the peripheral receptor to the spinal cord and then centrally to the brain. Perception occurs when the brain receives pain signals and interprets them as painful. The complex mechanism whereby synaptic transmission of pain signals is altered is called *modulation.* It is clinically useful to conceptualize pain physiology according to these four processes because each stage provides an opportunity for intervention in the pain experience (Fig. 47.1).

Transduction

Most pain begins in the periphery when free nerve endings called *nociceptors* are stimulated. Nociceptors transduce noxious stimuli into neuronal action potentials that progress centrally to the spinal cord and then the brain. Nociceptors are found in skin; muscle; connective tissue; the circulatory system; and the abdominal, pelvic, and thoracic viscera. Stimulation can be the result of direct damage to nerve endings, or it can result from release of chemicals at the site of injury.

Numerous substances are involved in the initiation of nociceptive impulses. Some of these substances are released as a direct result of tissue injury, whereas others may be produced as part of the inflammatory response to the injury. Some of the important chemical mediators of pain include K^+, H^+, lactate, histamine, serotonin, bradykinins, acetylcholine, substance P, and prostaglandins. These chemicals alter the membrane potential of the pain receptor, and if depolarization is sufficient, action potentials are generated. When these impulses are conducted centrally, the second step (transmission) is initiated.

Prostaglandin involvement in the process of nociceptor stimulation is of particular interest because prostaglandin inhibitors such as aspirin and other nonsteroidal antiinflammatory drugs (NSAIDs) are commonly used to manage pain. Prostaglandins are formed when cells are damaged and an enzyme, phospholipase A, breaks down phospholipids in the cell membrane and converts them to arachidonic acid (Fig. 47.2). Arachidonic acid undergoes further breakdown by the enzyme cyclooxygenase to form prostaglandins. Sensitization by prostaglandins lowers the threshold of nociceptive fibers so that stimuli that would not cause pain under normal circumstances are now pain producing. NSAIDs prevent prostaglandin production by inhibiting the action of cyclooxygenase.

Transmission

Stimulated nociceptors transmit impulses to the central nervous system (CNS) by means of specialized sensory fibers. The primary sensory fibers involved in the transmission of nociceptive impulses are the Aδ and C fibers. The characteristics and functions of these fibers are summarized in Table 47.1. In general, the larger, myelinated Aδ fibers transmit the nociceptive impulses very quickly as an initial response to tissue injury. The nature of the pain carried by the fast-traveling Aδ fibers is characterized as sharp, stinging, and highly localized. In contrast, unmyelinated C fibers transmit pain more slowly. Pain transmitted by C fibers is poorly localized and has a dull or aching quality that lingers long after the initial sharp pain abates. The majority of pain sensations travel via C fibers and project to areas of the brain that evoke emotional responses such as displeasure and anxiety.

Most sensory afferent pain fibers enter the spinal cord by way of the posterior nerve roots (Fig. 47.3). The cell bodies of pain neurons are located in the dorsal root ganglion. As the afferent neurons enter the dorsal horn, collateral branches spread up and down the spinal cord for two to three segments by way of the tract of Lissauer. These

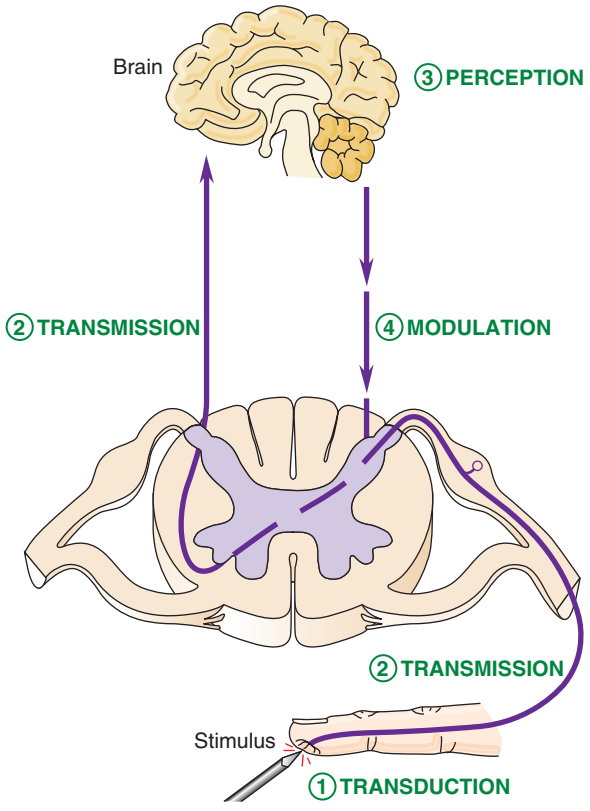


FIG 47.1 Four processes of pain signaling: transduction, transmission, perception, and modulation.

TABLE 47.1 Afferent Sensory Pain Fibers

Feature	Aδ Fibers	C Fibers
Structure	Myelinated	Unmyelinated
Amount	10%	90%
Source	Thermal, mechanical stimuli	Polymodal stimuli (mechanical, thermal, chemical)
Speed	Fast traveling, 5–10 m/sec	Slower traveling, 0.6–2 m/sec
Sensory quality of pain mediated	Sharp, stinging, cutting, pinching	Dull, burning, aching

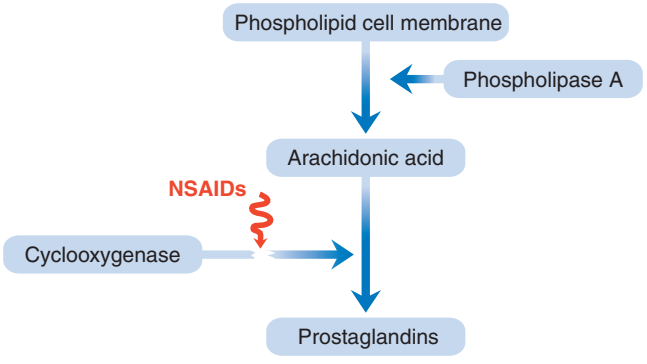


FIG 47.2 Tissue injury results in the release of prostaglandins from the breakdown of phospholipids in cell membranes. Nonsteroidal antiinflammatory drugs (NSAIDs) inhibit the cyclooxygenase enzyme and block the production of prostaglandins.

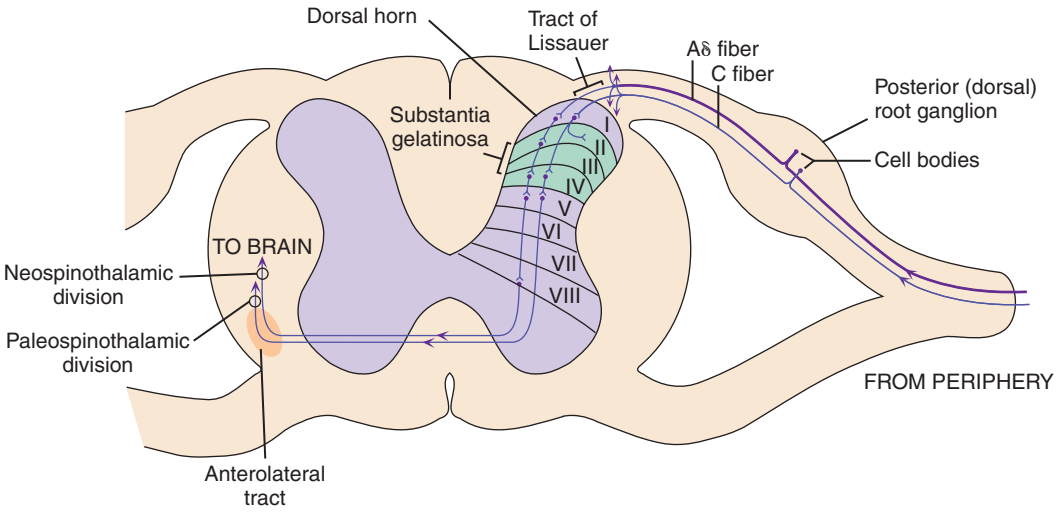


FIG 47.3 Spinal cord segment showing primary afferent pain fibers, Aδ and C fibers, entering the dorsal horn, synapsing on interneurons, crossing to the opposite side, and traveling to the brain in the anterolateral tract.

spinal connections are important for reflex postural adjustments when a painful body part is suddenly withdrawn from the painful stimulus.

Sensory afferent neurons synapse with interneurons, anterior motor neurons, and sympathetic preganglionic neurons in specific regions of the spinal cord (see Fig. 47.3). Aδ fibers and C fibers carry excitatory impulses from cutaneous pain receptors in small, localized areas of the skin to interneurons in lamina I. Many of the neurons originating in lamina I cross the spinal cord to activate neurons in the anterolateral tract.

Laminae II and III represent a key anatomic region of the cord involved in pain transmission known as the *substantia gelatinosa*. The substantia gelatinosa is characterized by multiple synaptic connections among primary sensory afferent neurons, interneurons, and anterolateral ascending fibers. There is an opportunity at this point for pain signal transmission to be modulated by other sensory input or from CNS activity. Pain signals can be either enhanced or blocked at these synapses.

Another key synaptic area involved in nociception is lamina V. Numerous Aδ and C fibers deliver somatic input from mechanical, thermal, and chemical receptors in the periphery to lamina V. Sensory afferent neurons from visceral receptors also terminate in lamina V. The convergence of both somatic and visceral fibers in lamina V may help explain the phenomenon of referred pain, in which pain from a visceral organ is perceived at the body surface. The remaining, deeper laminae VI to VIII receive sensory input from muscles, joints, and visceral afferent fibers.

A number of neurotransmitters and neuropeptides are involved in synaptic transmission in the spinal cord. Substance P is a well-known example of this. Others include excitatory amino acids (glutamate), γ-aminobutyric acid (GABA), cholecystokinin, and calcitonin gene–related peptide. These neurotransmitters bind to the next neurons in the pathway and thereby initiate action potentials. The pain signal is propelled along its pathway toward the brain. Interruption of these synaptic processes can inhibit pain transmission. The synapses in the spinal cord are extremely important points of pain modulation by both endogenous and exogenous means.

The excitatory neurotransmitter glutamate is involved in carrying the nociceptive message from primary afferent fibers to secondary neurons. Glutamate binding to its *N*-methyl-D-aspartate (NMDA) receptors on the postsynaptic neuron is thought to induce a kind of synaptic memory in the pain pathway. Excessive or repeated stimulation of C fibers sensitizes the spinal cord neurons so that even mild stimulation may be perceived as painful. This phenomenon has been termed *wind-up* and may be an important mechanism in the development of chronic pain syndromes. Drugs that inhibit glutamate production may impede the wind-up response, thereby controlling pain before synaptic memory of the pain develops in the pain pathways.

Pain signals transmitted by the spinal interneurons are then conducted to the brain by ascending spinal pathways (Fig. 47.4). The major pathway for pain signal transmission up the spinal cord is the anterolateral tract, so named because it travels in the anterolateral portion of the white matter of the spinal column. This tract is also called the *spinothalamic tract* in some texts and has two divisions: the neospinothalamic tract and the paleospinothalamic tract. Both divisions cross at the spinal segment and carry pain signals up the contralateral (opposite) side of the cord. Thus nociceptor input from the right side of the body travels in the anterolateral tracts on the left side of the cord, whereas pain signals from the left side of the body travel on the right side of the cord.

The neospinothalamic division has fewer synapses in the cord and projects first to the thalamus and then to the primary somatosensory cortex. Aδ fiber signals are transmitted in this tract and reach the brain quickly to provide specific information about pain location with little emotional connotation. C fiber impulses travel mainly in the paleospinothalamic division, which makes a greater number of synapses and reaches the brain more slowly. The paleospinothalamic tract projects to widespread brain areas and stirs aversive emotional responses. The paleospinothalamic tract travels with the neospinothalamic tract in the anterolateral portion of the spinal cord to the level of the medulla and then sends diffuse projections to the reticular formation, the mesencephalon, and, finally, the thalamus. From the thalamus, further projections to the cerebral cortex, limbic system, and basal ganglia occur. The pain sensation from C fibers is poorly localized, longer lasting, and more distressing than Aδ fiber pain. Pain impulses entering the brainstem, reticular activating formation, thalamus, and other lower brain centers cause conscious perception of pain, but the cortex is important in interpreting pain quality.

The brain can localize a pain sensation to a particular part of the body because nociceptor pathways are kept in specific anatomic order

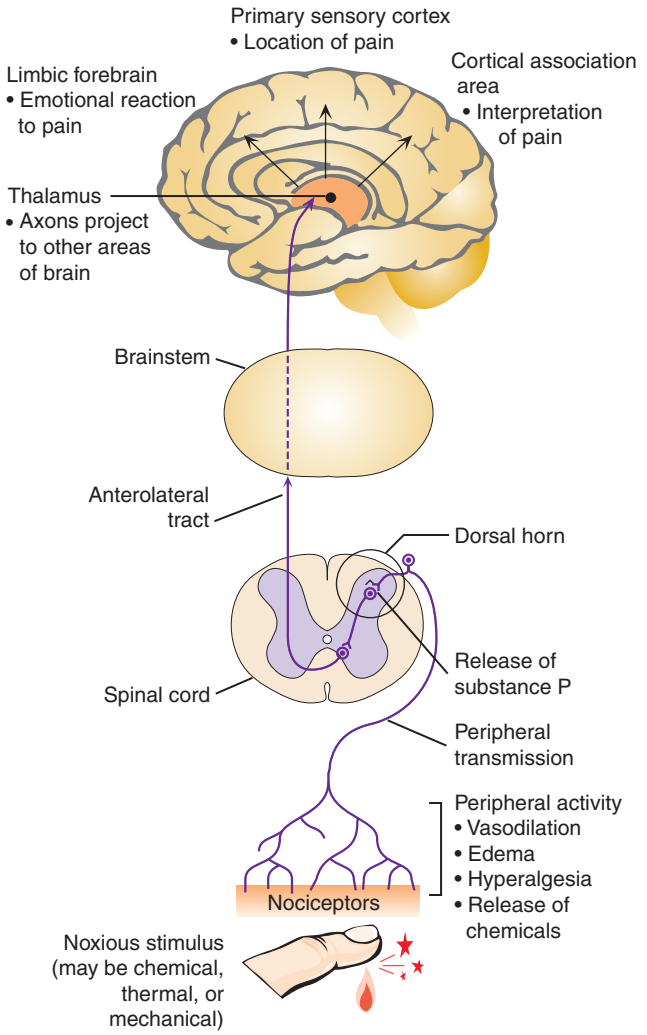


FIG 47.4 Anterolateral nociceptive pathways travel up the spinal cord and project to the thalamus, somatosensory cortex, cortical association areas, and limbic structures.

in the cord and somatosensory cortex. Each spinal nerve contains the nociceptor fibers for a particular area of the body surface, called a *sensory dermatome* (Fig. 47.5). Dermatomal maps are useful for locating a source of neurologic pain. Pain that follows a dermatomal distribution is due to spinal nerve compression or trauma and is called a *radiculopathy.* Vertebral disk disease is a common cause of radiculopathy. Peripheral neuropathies, in contrast, do not follow a dermatomal pattern. Examples of peripheral neuropathies are carpal tunnel syndrome (median nerve) and diabetic neuropathy, which often affects both legs in a stocking-like pattern.

Perception

Perception is the result of neural processing of pain sensations in the brain. Perception includes an awareness and interpretation of the meaning of the sensation. Pain perception is influenced by attention, distraction, anxiety, fear, fatigue, and previous experience and expectations. Pain perception is not localized to a specific brain area. Complete removal of the somatosensory areas of the cerebral cortex does not destroy an animal's ability to perceive pain. Numerous neuronal networks are necessary to localize, process, and interpret painful sensations. The primary somatosensory cortex, association cortex, frontal lobe, and limbic structures all participate in this processing.

Pain perception can be described in terms of pain threshold and pain tolerance. Pain threshold is the level of painful stimulation required to be perceived and is remarkably similar from one individual to another. Pain tolerance is the degree of pain that one is willing to bear before seeking relief. Pain tolerance varies widely among individuals and within the same individual under differing conditions. Age, culture, family upbringing, gender, and previous pain experience influence tolerance to pain. Environmental factors, including noise, bright light, and interrupted sleep, may affect pain tolerance.

Pain expression is the way in which the pain experience is communicated to others. Pacing, writhing, jaw clenching, facial grimacing, muscle guarding, crying, moaning, groaning, and verbal descriptions may be used to express pain. Thus the highly variable nature of pain expression among individuals makes accurate pain assessment difficult.

Modulation

Modulation of pain signals occurs at multiple sites along the pain pathway. A fair amount is known about pain modulation at the spinal cord, where neurons from nociceptors, somatosensory receptors, and descending neurons from the CNS all converge and interact. Modulation also occurs at the peripheral nociceptor ending and within the brain; however, these mechanisms are less well understood.

Attempts to decrease the perception of painful stimuli may be initiated spontaneously by the person experiencing pain. Rubbing, pressing, or shaking the painful area may reduce the intensity of pain. In 1965 Melzack and Wall proposed the gate control theory to explain how stimulation of large "touch" neurons could inhibit the transmission of nociceptor impulses. This theory was very useful in focusing efforts to understand pain signal processing at the spinal cord level.

Central to the gate control theory is the capacity for interneurons in the spinal cord to modify the transmission of nociceptor impulses. The original gate control theory suggested that impulses carried by large myelinated cutaneous fibers (Aβ) could "close the gate" on nociceptor impulses so that pain signals would be blocked in the spinal cord and not allowed to progress centrally to the brain. The physiologic process underlying the closing-the-gate mechanisms has been the subject of much study. Numerous interneurons, neurotransmitters, and neuropeptides have been implicated in this complex gating mechanism.

Descending pathways from the brain to the dorsal horn region of the spinal cord are important modulators of the pain response (Fig. 47.6). These descending pathways originate in a brainstem nucleus called the *raphe magnus* and project to the dorsal horn regions of laminae I, II, and IV. Neurotransmitters released by these neurons can inhibit synaptic transmission of pain signals. One way to inhibit synaptic transmission is through presynaptic inhibition of substance P release from nociceptor neurons (Fig. 47.7). Opioids such as endorphins are thought to be the mediators of presynaptic inhibition. A similar inhibitory effect can be achieved by administering opioid drugs, such as morphine, that bind to opioid receptors and mimic the effect of endorphins.

The raphe magnus receives input from two other brain areas important in the pain response: the periaqueductal gray (PAG) area in the midbrain and the rostral pons in the brainstem. The PAG area has a high concentration of endogenous opioids (endorphins and enkephalins) that are known to produce analgesic effects similar to narcotic drugs. Stimulation of the PAG area causes release of these endogenous opioids and sends nerve impulses to the raphe magnus. Serotonin (5-hydroxytryptamine) is the neurotransmitter that conveys analgesic signals from the PAG area to the raphe magnus. This finding helps explain the pain-relieving action of drugs that enhance serotonin activity in the brain, such as tricyclic antidepressants.

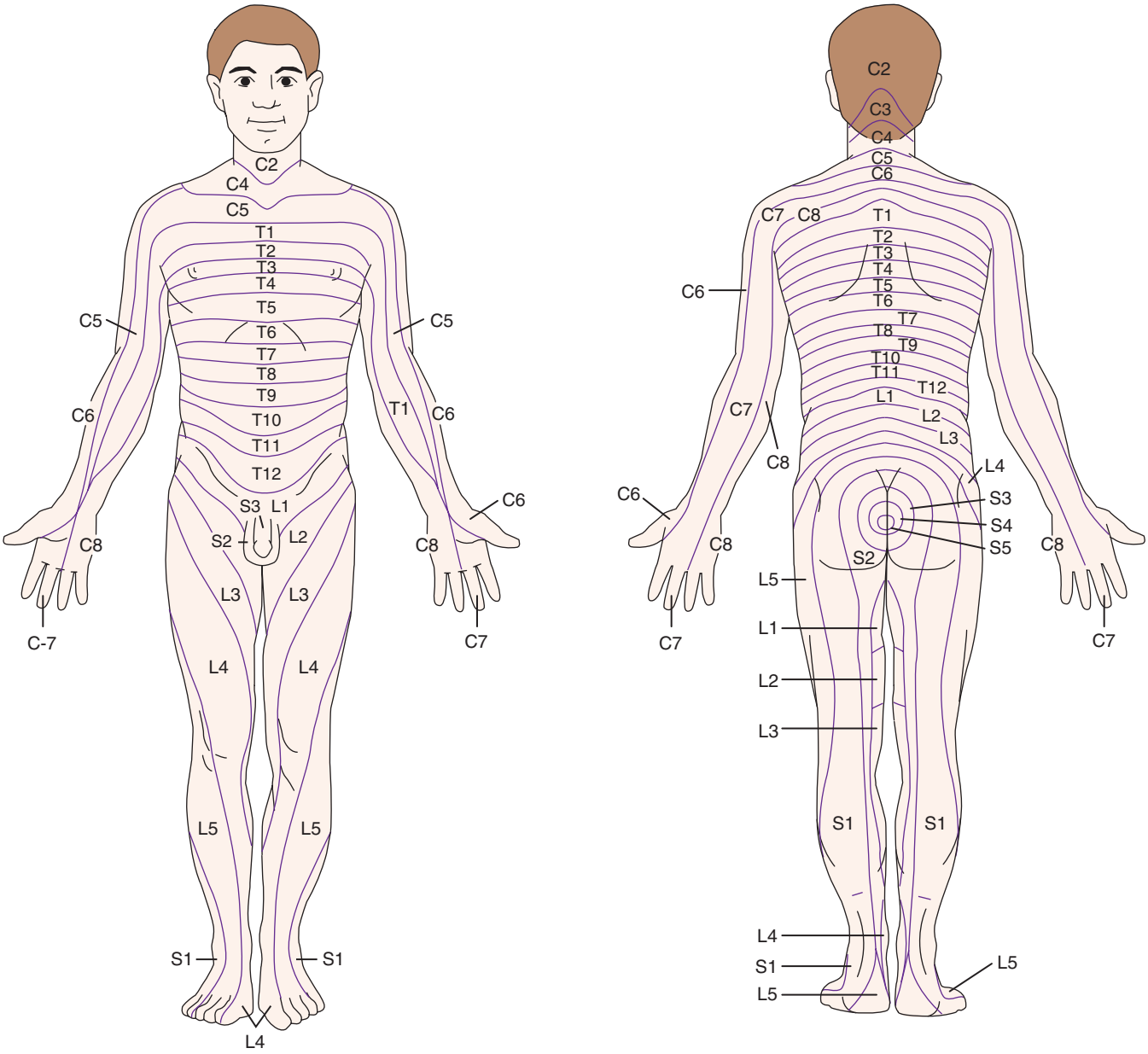


FIG 47.5 Sensory dermatomes. Pain located in the pattern of a dermatome occurs with spinal nerve injury and is referred to as *radiculopathy*.

The neurons projecting to the raphe magnus from the rostral pons secrete norepinephrine as the neurotransmitter. Stimulation of these neurons also produces an analgesic effect. Clonidine, a drug that mimics the effect of norepinephrine in the brain, has been shown to have pain-relieving properties.

The descending pathways from the brain provide an important means for gating the flow of pain impulses from the periphery to the brain. The PAG area is apprised of the flow of pain signals because it receives input by way of the thalamus and limbic structures.

Pain modulation occurs not only at the cord level, but also in the brain itself. Opioids produced in the brain are thought to be important modulators of pain perception. Specific opioid receptors were identified within the brain in the early 1970s. Also discovered around this time were the naturally occurring morphinelike substances termed *endorphins.* The word *endorphin* is a combination of two words: *endogenous* (coming from within the body) and *morphine* (from the Latin word *morpheus,* meaning "sleep inducing"). The term *endorphin* actually refers to two groups of naturally occurring peptides: enkephalins (which are pentapeptides) and three types of endorphin polypeptides (α-, β-, and γ-endorphin). Of these, most is known about β-endorphin.

During times of stress, pain, or emotion, the brain apparently creates its own analgesia through the secretion of endogenous opioids—a process known as *stress-induced analgesia.* Stress-induced analgesia is reversed by naloxone, a drug that blocks opioid receptors, thus supporting the role of endogenous opioids in the process. As previously described, the PAG area produces large quantities of endogenous opioids that are thought to inhibit pain signal perception within the brain. High concentrations of β-endorphin are also found in the pituitary gland, and it is likely that the release of pituitary stress hormones (adrenocorticotropic hormone) is accompanied by the release of endorphins. The adrenal glands also produce endogenous opioids as a sympathetic response to stress. Endorphins released into the bloodstream by the

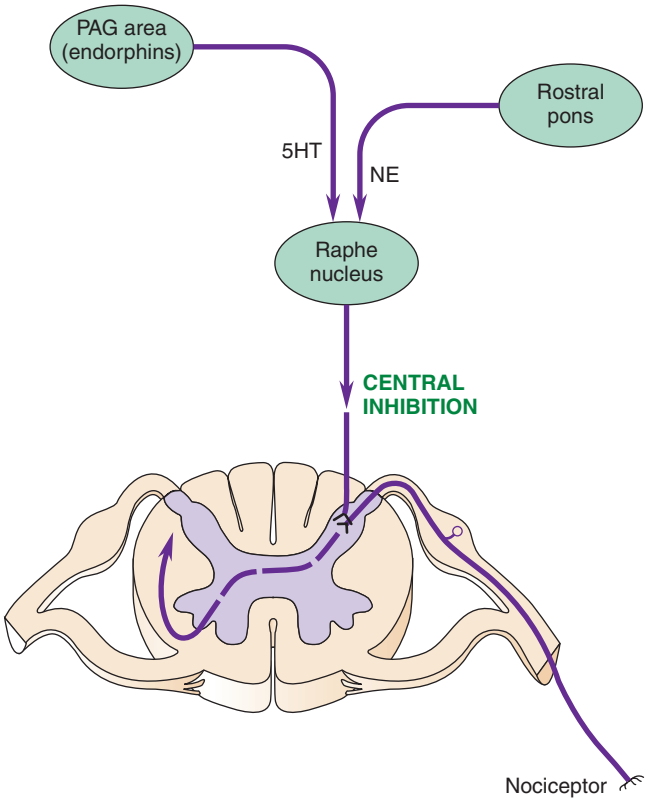


FIG 47.6 Descending pathways from the brain are thought to regulate pain impulse transmission in the dorsal horn. These regulatory neurons originate in the brainstem raphe magnus, which receives input from the periaqueductal gray (PAG) area and the rostral pons. Stimulation of these brain areas induces analgesia. *5HT,* Serotonin; *NE,* norepinephrine.

pituitary gland and adrenal gland have their effects in the periphery because they cannot effectively cross the blood–brain barrier.

Opioids have different effects depending on the types of receptors they activate. Four types of opioid receptors have been identified: mu (μ), kappa (κ), sigma (σ), and delta (δ) (Table 47.2). The distribution of the specific opioid receptors varies throughout the body. The μ and κ receptors have analgesic activities. The μ receptors are found in high concentration in the brain, where they are thought to modulate pain perception. The κ receptors are concentrated primarily in the spinal cord, where they contribute to pain modulation by CNS-descending pathways. Each opioid receptor subtype is associated with a number of undesirable side effects. Depending on the affinity for certain receptors, different drugs may have differing analgesic potency and side effect profiles (Table 47.3).

KEY POINTS

- Nociception can be conceptualized as four interdependent processes: stimulus transduction, signal transmission, pain perception, and pain modulation.
- Nociceptor activity is transmitted to the spinal cord by two types of neurons: larger, myelinated Aδ fibers, which transmit sharp, localized sensations, and small, unmyelinated C fibers, which transmit dull, aching, poorly localized sensations.
- Pain signals are transmitted by afferent fibers that enter the spinal cord through the dorsal horn, synapse on interneurons, and then cross the cord and project centrally in the anterolateral tract.

TABLE 47.2 Opioid Receptor Activity

Opioid Receptor	Activity
Mu (μ)	Analgesia
	Sedation
	Respiratory depression
	Pupil constriction
	Nausea and vomiting
	Constipation
	Urine retention
	Pruritus
Kappa (κ)	Analgesia
	Sedation
	Respiratory depression
	Pupil constriction
	Diuresis
Sigma (σ)	No analgesia
	Vasomotor stimulation
	Tachypnea
	Pupil dilation
	Psychotomimetic effects (hallucinations, paranoia, delirium)
Delta (δ)	No analgesia
	Respiratory depression
	Nausea and vomiting
	Pruritus

- The anterolateral tract has two divisions: the neospinothalamic tract, which carries Aδ fiber input and projects to the thalamus and then the sensory cortex, and the paleospinothalamic tract, which carries C fiber input and projects diffusely to the reticular formation, mesencephalon, and thalamus.
- Perception of painful stimuli involves several brain structures, including the primary somatosensory cortex, association areas, and limbic structures. Pain perception is influenced by culture, environment, and physical status and varies widely among individuals.
- Afferent pain signals can be modulated at several levels. Descending pathways project from the PAG area and rostral pons by way of the raphe magnus to inhibit pain neurons in the dorsal aspect of the spinal cord. Pain is also modulated within the brain/spinal cord by endogenous opioids (enkephalins, endorphins).

TYPES OF PAIN

Pain can be categorized into two major subtypes: physiologic pain and pathologic pain. Physiologic pain occurs when tissue injury has occurred and aids in prevention of further injury or, in some cases, survival. This is the pain felt when a person, for example, has touched a hot stove. This injury will likely heal, and the memory of the experience will hopefully prevent future similar injuries. Pain of acute appendicitis also may be considered physiologic pain because it alerts a person to a serious problem. In contrast, pathologic pain occurs after tissue injury, but long-term changes occur both within the peripheral and the central nervous systems. These changes occur along somatosensory pathways from the periphery to the cortex. The pain sensation can be significantly enhanced (hyperalgesia), or nonnoxious stimuli may cause pain (allodynia). These changes in pain perception and modulation serve no beneficial purpose to learning or survival. These underlying changes

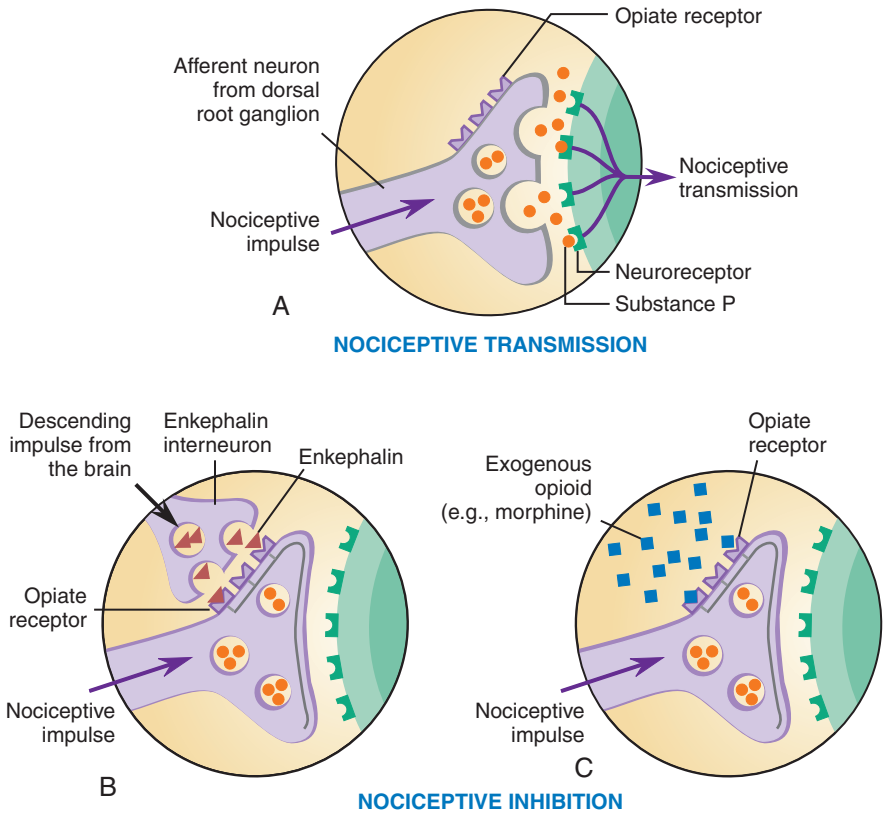


FIG 47.7 Pain transmission and inhibition at the molecular level. **A,** Nociceptive transmission to higher levels of the central nervous system. **B,** Nociception inhibited through binding of endogenous opioids (e.g., enkephalin). The release of substance P is prevented. **C,** Nociception inhibited through binding of exogenous opioid (e.g., morphine). Release of substance P is prevented.

TABLE 47.3 Receptor Affinity of Commonly Used Opioids

		AGONIST		ANTAGONIST	
Drug	**Receptor Affinity**	**Pure**	**Partial**	**Pure**	**Partial**
Morphine, meperidine, hydromorphone, methadone, fentanyl	Mu (μ)	XXX			
	Kappa (κ) *(morphine only)*				
	Delta (δ)				
Buprenorphine	Mu (μ)		X		
	Kappa (κ)			X	
Butorphanol	Mu (δ)		X		
	Kappa (κ)	X			
Nalbuphine	Mu (δ)				X
	Kappa (κ)	X			
Pentazocine	Mu (δ)				X
	Kappa (κ)	X			
Naloxone	Mu (δ)			X	
	Kappa (κ)			X	
	Delta (δ)				X
Naltrexone	Mu (δ)			X	
	Kappa (κ)			X	

are theorized to be the cause of neuropathic pain, fibromyalgia, and other chronic pain syndromes.

Pain most commonly is classified according to duration (acute, chronic), source (cancer, neuropathic, ischemic), or location and referral pattern. Pain is a symptom of an underlying problem rather than a primary disorder; attempts to alleviate pain should be accompanied by efforts to locate and manage the underlying etiology. The character, location, and duration of pain can provide helpful clues to aid the diagnostic process.

ACUTE PAIN

Pain is categorized as being acute or chronic, depending on the duration of symptoms. Acute pain results from tissue injury and resolves when

TABLE 47.4 Physiologic Responses to Acute Pain

Criteria	Response
Signs and symptoms	↑ Heart rate
	↑ Blood pressure
	↑ Respiratory rate
	Dilated pupils
	Pallor and perspiration
	Nausea and vomiting
	Urine retention
Physiologic response	Blood shifts from superficial vessels to striated muscle, heart, lungs, and brain
	Bronchioles dilate to ↑ oxygenation
	↑ Gastric secretions
	↓ Gastrointestinal motility
	↑ Circulating blood glucose
	Hypomotility of bladder and ureters

Adapted from D'Arcy Y, Burns S: Pain and sedation management. In: Burns S, editor, *AACN essentials of progressive care nursing,* ed 3. New York, 2014, McGraw-Hill Professional.

the injury heals, usually in less than 3 months. Acute pain is typically accompanied by clinical signs and symptoms of pain that result from stimulation of the sympathetic nervous system. These signs and symptoms include an elevated heart rate, respiratory rate, and blood pressure, as well as pallor, sweating, and nausea (Table 47.4). Persons experiencing acute pain may express pain behavior such as pacing, grimacing, crying, or moaning. Short-term therapy with nonopioid and opioid agents is often helpful. The risk of becoming dependent on pain medications is minimal in persons experiencing acute pain. Adequate management of pain during an acute episode may help prevent the development of some types of chronic pain syndromes.

Headache

Etiology and Pathogenesis

Headache is one of the most common causes of acute pain. It is estimated that 47% of the adult population have suffered from a headache in the last year. Headaches are classified according to etiologic categories (e.g., tension, migraine, sinus). However, it is very common for headache type to be misdiagnosed, with migraines frequently being labeled as tension or sinus headache. Migraine is one of the most common headache types, affecting 10% to 15% of adults and adolescents. Previous theories of migraine included simply a vascular causation. It was believed that the spasm of cerebral vessels and the ensuing vasodilation caused the throbbing pain typical of migraine. Migraine is now known to be a complex neurologic disorder that affects multiple cortical, subcortical, and brainstem areas that regulate autonomic, affective, cognitive, and sensory functions.

Migraines probably result from dysfunction of the brainstem areas involved with modulation of craniovascular afferent fibers. There is a release of inflammatory chemicals, including local release of calcitonin gene–related peptide, which causes vasodilation. The inflammatory cascade that results in migraine can be initiated by many "migraine triggers." There is a genetic predisposition for sensitivity to these headache triggers. The brainstem becomes activated, causing stimulation of the trigeminal nucleus caudalis. This activation causes pain signals to be sent to the thalamus and cerebral cortex, where pain is perceived and is responsible for many signs/symptoms of migraine. Any of the branches of the trigeminal nerve can refer pain signals along this pathway. If left untreated, continuous activation of the trigeminal nucleus caudalis may result in central sensitization and more refractory pain.

Migraines are associated with several comorbid conditions, including depression and anxiety. Those who have experienced childhood trauma are also at higher risk for migraines. Other conditions include sleep disturbance, epilepsy, patent foramen ovale, ischemic stroke, fibromyalgia, and chronic fatigue syndrome.

Clinical Manifestations

Typical signs of a migraine headache include severe unilateral pounding or throbbing pain that may be accompanied by nausea, vomiting, photophobia, phonophobia, and lacrimation. The pain is increased by routine physical activity. Some migraines may be preceded by an aura such as flashing lights or other visual disturbances and unilateral paresthesias. Associated physical symptoms may include sinus/nasal congestion, neck muscle stiffness and pain, vertigo, and changes in bowel pattern. Affective symptoms such as mood changes and irritability are not uncommon.

Diagnosis and Treatment

Headaches are diagnosed through careful history and physical examination. The presence of brain tumors, infection, hydrocephalus, and increased intracranial pressure must be ruled out. Physical assessment should include the ears, nose and throat, sinuses, temporomandibular joint, neck musculature, cranial nerves, and retinal examination. General cognitive, neurologic, and motor function should also be examined. Headaches caused by trauma, following a worsening pattern, accompanied by other neurologic symptoms, or developing suddenly and described as the "worst headache ever" require neuroimaging. The International Headache Society has established criteria for a migraine diagnosis, and this has been adopted by many health care providers (Box 47.1).

Headaches are managed with a wide variety of therapies and medications, each aimed at a different piece of the pathophysiologic puzzle. Depending on the type and frequency of the headache, prophylactic medications may also be used. One mainstay of nonpharmacologic migraine therapy is the avoidance of headache triggers. Patients are encouraged to eliminate vasoactive substances from their diets, including caffeine, cheese, chocolate, foods containing nitrates and nitrites, and monosodium glutamate. Adherence to regular sleep–wake schedules is helpful, as is the use of stress management techniques in preventing the occurrence of migraines. Other nonpharmacologic therapies for headache include resting in a quiet, darkened room or applying cold packs to the head and back of the neck. Numerous medications are used to control migraines and other types of headaches. A key to successful treatment is the prompt use of these agents at the onset of pain. Prophylactic therapy employing administration of antiepileptics (topiramate), antidepressants (selective serotonin reuptake inhibitors, tricyclic antidepressants), β-blockers (propranolol), α-blockers (clonidine), or calcium channel blockers (verapamil); injections of botulinum toxin type A (Botox) into the scalp; or manipulation of hormone levels in younger female patients may be necessary. Refer to Table 47.5 for a comparison of pharmacologic therapies for the acute management of migraine headache.

CHRONIC PAIN

Pain is considered chronic when it lasts more than several months beyond the expected healing time (usually more than 6 months). When chronic pain is not due to a malignancy, its cause is often difficult to ascertain. In chronic pain, pain is no longer protective and appears to be self-perpetuating. Two concepts have emerged as probable mechanisms for this phenomenon. Peripheral sensitization represents a reduction

BOX 47.1 International Headache Society Diagnostic Criteria for Migraine

Migraine Without Aura	Migraine With Aura
At least five headache attacks lasting 4–72 hours and having at least two of the following: • Unilateral location • Pulsating quality • Moderate or severe intensity • Aggravation by routine physical activity At least one of the following characteristics is present: • Nausea and/or vomiting • Photophobia and phonophobia Headache/symptoms not attributable to another disorder	At least two headache attacks with: One or more of the following reversible aura symptoms: • Visual • Sensory • Speech and/or language • Motor • Brainstem • Retinal At least two of the following four characteristics: • At least one aura symptom spreads gradually over ≥5 minutes and/or two or more symptoms occur in succession • Each individual aura symptom lasts 5–60 minutes • At least one aura symptom is unilateral • The aura is accompanied or followed within 60 minutes by headache Headache/symptoms not attributable to another disorder

Adapted from: The International Classification of Headache Disorders, ed 3 (beta version). *Cephalalgia*. 2013;33(9):629–808. doi:10.1177/0333102413485658.

in threshold and an amplification in the responsiveness of nociceptors that occur when the peripheral terminals of the primary sensory neurons are exposed to inflammatory mediators and damaged tissue. Increased peripheral transduction sensitivity develops. Central sensitization results in changes in the properties of neurons in the CNS. It is an abnormal state of responsiveness or increased gain of the nociceptive inputs. When neurons in the dorsal horn of the spinal cord are affected by central sensitization, they develop increased spontaneous activity and a reduction in the threshold for activation by peripheral stimuli, show increased responsiveness to stimulation, and have an enlargement of receptor field.

Chronic pain is generally not associated with signs and symptoms of sympathetic activity. As the body becomes accustomed to pain, the sympathetic nervous system desensitizes itself to the noxious input; therefore symptoms are more often psychological. Lack of sleep because of pain causes fatigue and irritability. Loss of a job or loss of body image because of pain causes personal and family difficulties. Treatment failures may create a sense of hopelessness or distrust of caregivers.

Depression is a common finding in individuals experiencing chronic pain, and conversely, those with depression are more likely to suffer from chronically painful conditions. The CNS undergoes long-term plastic changes associated with chronic pain and depression. Circuit and molecular mechanisms that underlie this plasticity have begun to emerge. In many cases the cause of the chronic pain cannot be determined, and therefore treatment is difficult. The use of narcotic pain relievers is discouraged because of the necessity of long-term therapy and therefore a risk of dependency. Satisfactory treatment may require numerous coordinated approaches, and the patient may benefit from the services of a pain clinic that specializes in multimodal therapies.

Fibromyalgia Syndrome

Etiology and Pathogenesis

Fibromyalgia syndrome (FMS) is a chronic pain syndrome affecting an estimated 2% of the population. Women are affected more frequently than men. FMS is a collection of symptoms without a clear physiologic cause. It is neither degenerative nor progressive. Patients have a history of chronic widespread pain affecting all four extremities, sleep disturbances, and fatigue. Cognitive dysfunction, anxiety, and depression are also common. The cause of FMS is unknown. However, etiologic studies have identified several risk factors for the development of the syndrome. Individuals with a medical history of excessive stress, trauma (both physical and emotional), sexual abuse, viral infections (parvovirus, hepatitis C, Epstein–Barr), and endocrine disorders (hypothyroidism) are more commonly affected. Disordered pain mechanisms in the CNS are a suspected factor in FMS. Patients with FMS have a lower threshold for pain than those without the disorder. Pain maintenance and modulation mechanisms in the brain and spinal cord are also suspect. Central sensitization is thought to be a key factor.

Clinical Manifestations

Patients complain of pain that waxes and wanes and that does not follow a dermatomal pattern. The pain tends to be exacerbated by physical exertion. Hyperalgesia and allodynia are common. Musculoskeletal examinations are generally normal. Other symptoms commonly seen associated with FMS include sleep disturbance/insomnia with nonrestorative sleep and irritable bowel syndrome. Fatigue is a hallmark of the syndrome. Depression and anxiety are also common, along with cognitive difficulties such as problems with attention and short-term memory.

Diagnosis and Treatment

FMS is a diagnosis of exclusion. Thyroid disorders, myopathies, rheumatoid arthritis, and chronic viral infections (e.g., human immunodeficiency virus) must be excluded. In FMS there is a lack of objective or laboratory findings. However, the American College of Rheumatology has established criteria to assist in the diagnosis of FMS. An individual must complain of widespread pain that has been present for at least 3 months without other reasonable explanation. The presence of pain in 11 of 18 "trigger" or "tender" points when pressure is applied to these areas (Fig. 47.8) is diagnostic. The use of patient-completed diagnostic surveys, including the Widespread Pain Index and Symptom Severity Scale, was recognized in 2011 as an alternative means of diagnosing fibromyalgia.

The management of FMS begins with patient education. Although sometimes disabling, FMS is not a fatal illness and does not affect life span. Treatment includes a variety of medications including antidepressants, such as the selective serotonin reuptake inhibitors and the tricyclic antidepressants. Restoration of sleep patterns seems to be a key factor of successful treatment. NSAIDs and muscle-relaxing agents are also helpful. Opioid medications and corticosteroids are generally avoided because these are not effective long-term therapies. The Food and Drug Administration has approved three therapies for the pain of FMS. Pregabalin (Lyrica) is an analog of the neurotransmitter GABA. Milnacipran (Savella) and duloxetine (Cymbalta) are in the class of medications of serotonin–norepinephrine reuptake inhibitors. New research is focusing on lifestyle factors and the symptoms of fibromyalgia. Gluten-free diets have shown promise in reducing the severity of pain. Other nonpharmacologic therapies include regular physical exercise and psychological counseling.

TABLE 47.5 Acute Migraine Therapies

Drug	Route of Administration	Comments
Serotonin Receptor Agonists (Triptans)		
Sumatriptan (Imitrex) Zolmitriptan (Zomig) Rizatriptan (Maxalt) Naratriptan (Amerge) Eletriptan (Relpax)	PO, nasal spray, subQ injection, dissolvable tablet	Increases serotonin; cannot be given with monoamine oxidase inhibitors; contraindicated in renal/hepatic failure; risk of serotonin syndrome if given with selective serotonin reuptake inhibitors. Caution in cardiovascular disease
Ergot Alkaloids		
Ergotamine with caffeine (Wigraine, Ercaf, Cafergot)	IV, IM, subQ, PO, rectal suppository, inhaler	Rarely used because of development of newer agents with fewer side effects; increases serotonin; causes intracerebral vasoconstriction; best if given early after onset; contraindicated in patients with cardiovascular disease and pregnancy; may cause nausea/vomiting
Dihydroergotamine mesylate (DHE, Migranal)		Rarely used now; used for refractory headaches
Nonsteroidal Antiinflammatories and Nonopiates		
Acetaminophen (Tylenol) Ketorolac (Toradol) Naproxen sodium (Naprosyn, Aleve) Ibuprofen (Motrin, Advil) Aspirin	PO, IV, IM	May cause gastrointestinal upset/bleeding; frequent use may cause "rebound" headache
Barbiturate-Hypnotic Combinations		
Butalbital with aspirin and caffeine (Fioricet, Fioricet with codeine)	PO	High abuse/addictive potential; may cause sedation; rebound headache with frequent use; withdrawal potential
Opiates/Combinations		
Acetaminophen with codeine, oxycodone, or hydrocodone (Tylenol no. 3/4, Percocet, Lorcet) Butorphanol (Stadol)	PO, nasal spray	High abuse/addictive potential; may cause sedation; considered "rescue medications"

Adapted from: Ropper AH, Samuels MA, Klein JP: Chapter 10. Headache and other craniofacial pains. In: Ropper AH, Samuels MA, Klein JP, editors, *Adams & Victor's principles of neurology,* 10e, New York, 2014, McGraw-Hill; Hammond A, Holcomb M: Exploring treatment options for migraine headache. *The Clinical Advisor.* 2015;18(8):35–42.
IM, Intramuscular; *IV,* intravenous; *PO,* oral; *subQ,* subcutaneous.

CANCER-RELATED PAIN

Cancer pain is a subcategory of chronic pain, although it may be associated with acute pain episodes. Malignant pain differs from nonmalignant chronic pain in that it often has an identifiable cause. Pain associated with cancer may result from infiltration of organs or compression of structures by an expanding tumor, or it may occur as a result of treatments that damage tissue such as radiation therapy or chemotherapy. In patients with cancer pain, clinical signs and symptoms are often a mixture of sympathetic nervous system activation and behavioral changes. Unremitting cancer pain requires a multifaceted approach and use of potent medications. Often the quality of life is a larger consideration than the length of life, and adequate pain control is a major factor affecting the quality of life.

NEUROPATHIC PAIN

Neuropathic pain is a complex, often disabling, chronic pain that is caused by lesions or disease of the somatosensory nervous system. This may occur in the peripheral or central nerves. The pain results from the actual damage or dysfunction of the nerves rather than stimulation of the pain receptors. Nerve injury from surgery, tumor growth, metastasis, radiation therapy, chemotherapy, elevated blood glucose level, viral infection, or trauma often causes neuropathic pain. It is characterized by constant aching sensations that may be interrupted by bursts of burning or shocklike pain in the affected area. Allodynia is common. Neuropathic pain may not occur immediately after an injury. Days, weeks, or even months after the tissue-damaging source of pain has resolved, the onset of neuropathic pain can initiate a new and complex pain state. Pain often seems to be out of proportion to the area of tissue damage.

Peripheral processes in neuropathic pain include production of mediators (cytokines, protons, nerve growth factor), alterations in calcium channels, sodium channels, hyperpolarization-activated nucleotide-gated ion channels, and potassium channels. Phenotypic switches and sprouting of nerve endings and involvement of the sympathetic nervous system also occur. These result in increased afferent nociceptor firing. Stimulation of the NMDA receptor; activation of microglia, oligodendrocytes, and astrocytes; and increased production of nerve growth factor and brain-derived neurotrophic factor, together with loss of spinal inhibitory control, are responsible for central neuron hyperexcitability. This leads to the maintenance of neuropathic pain.

Examples of neuropathic pain include postherpetic neuralgia, diabetic neuropathy, trigeminal neuralgia, epidural spinal cord compression, cauda equina compression, plexus injuries, and phantom limb pain. Sympathetically maintained pain is a unique type of neuropathic pain

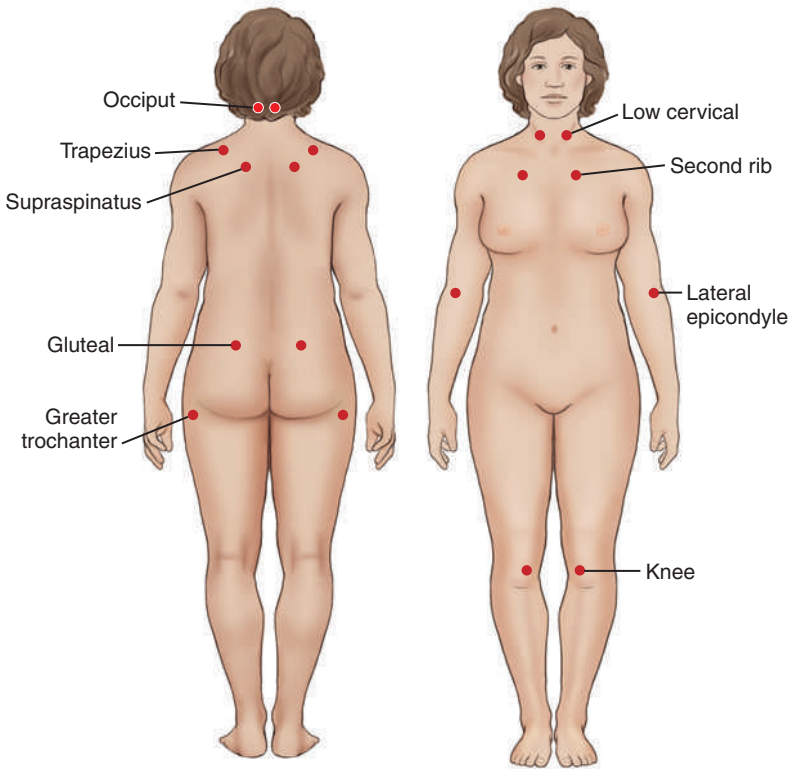


FIG 47.8 Posterior and anterior trigger points in fibromyalgia.

that may occur in the absence of nerve injury. Sympathetically maintained pain is attributed to hyperactivity of the sympathetic nervous system. The release of norepinephrine from sympathetic nerve endings sensitizes nociceptors such that they respond to a lower level of nociceptor stimuli. Not all patients affected by this type of pain exhibit the same symptoms; however, the most prevalent clinical manifestations are allodynia, hyperalgesia, atrophy of the affected extremity, coldness in the affected area, and dystrophic changes, most often manifested as hair loss and a shiny appearance of the skin. Neuropathic pain is difficult to manage. It is frequently unresponsive to opioid or other pharmacologic therapy.

Trigeminal Neuralgia

Etiology and Pathogenesis

Trigeminal neuralgia is a form of neuropathic pain that can be quite disabling for patients. It appears as sudden, momentary, and excruciating pain along the second and third divisions of the trigeminal nerve. Trigeminal neuralgia is more common in women than in men and occurs more frequently in middle-aged or older individuals. If trigeminal neuralgia occurs at an earlier age, multiple sclerosis should be ruled out. Other causes of trigeminal neuralgia include lesions or tumors of the brainstem. Chronic compression of the trigeminal nerve by a vessel is suspected in most cases. This causes demyelination of the trigeminal nerve and interruption and alteration in nerve signaling.

Clinical Manifestations

The pain of trigeminal neuralgia is often described as sharp or shooting; some have compared it to the pain of an electrical shock. Patients may be pain-free between episodes or complain of a dull ache in the affected area. Sometimes patients may only have a few episodes of pain followed by a long remission period. However, others may unfortunately experience an increase in frequency and duration of pain. Anxiety is common, because patients worry about when their next attack may occur.

Diagnosis and Treatment

The diagnosis of trigeminal neuralgia is most frequently based on the clinical history. The results of neurologic evaluation are normal if there is no underlying lesion. Management of trigeminal neuralgia includes use of antiseizure medications such as carbamazepine (Tegretol), oxcarbazepine (Trileptal), or lamotrigine (Lamictal). Surgical nerve decompression has been used successfully for trigeminal neuralgia in patients who do not respond to or cannot tolerate the medications. Gamma-knife radiosurgery may be effective, and new research is looking at the use of botulinum toxin A.

Diabetic Neuropathy

Etiology and Pathogenesis

One of the most common complications of diabetes, diabetic neuropathy affects approximately 90% of all persons with diabetes. Diabetic neuropathy is caused by damage to the peripheral nerves. The exact pathogenetic mechanism is unknown. The combination of numbness, ongoing pain, allodynia, and hyperalgesia suggest changes occurring both in the peripheral and central nervous systems. Proposed peripheral changes include changes in Na^{+} and Ca^{++}channel distribution and expression, sympathetic sprouting, loss of peripheral spinal inhibition, axonal degeneration and regeneration, and damage to small nerve fibers. Activation of glia with the production of inflammatory cytokines is thought to contribute to the peripheral nerve damage. Proposed central mechanisms associated with diabetic peripheral neuropathy include central sensitization, changes in the balance of facilitation/inhibition within descending pathways, and increased thalamic vascularity.

Clinical Manifestations

Although pain is the most common feature, patients also complain of numbness and tingling, mild weakness, and loss of vibratory sense and proprioception. Fine touch and vibratory sensation are decreased. Patients

complain of burning pain in the distal bilateral lower extremities, often with a symmetric distribution. Pain is frequently worse at night.

Diagnosis and Treatment

Diabetic neuropathy is confirmed through careful physical examination. Diabetic patients are encouraged to maintain strict control of their blood glucose levels to prevent neuropathy. Management of this disorder includes the use of a wide variety of topical and systemic pain medications. Systemic therapeutic agents include tricyclic antidepressants, serotonin–norepinephrine reuptake inhibitors, and anticonvulsants. Although opioids can help with pain relief of diabetic neuropathy, their use has been limited because of tolerance and dependence issues. Side effects of all of these medications can limit their usefulness, especially in elderly patients or those with other comorbid conditions. Many patients use a combination of topical and systemic therapies. An important nonpharmacologic treatment for diabetic neuropathy is the prevention of further complications. Diabetic patients are strongly encouraged to perform daily foot examinations, taking precautions against the development of foot sores and ingrown toenails. The combination of numbness and impaired circulation makes diabetic patients at high risk for undetected injuries that do not heal and become easily infected. Amputation is a possible outcome.

Postherpetic Neuralgia

Etiology and Pathogenesis

A common but disabling complication of the varicella virus is herpes zoster. Years after an individual has recovered from the chickenpox virus, herpes zoster (shingles) may occur. This is a reactivation of the latent virus that has lain dormant along the nerve roots.

Postherpetic neuropathy is persistent pain that lasts for more than 8 weeks after the onset of skin lesions. Approximately 15% of patients with herpes zoster develop complications, the most common of which is postherpetic neuralgia. Risk factors for the development of this neuralgia are advanced age and history of immune compromise.

Clinical Manifestations

Herpes zoster is characterized by a burning pain that follows along a dermatomal pathway and is accompanied by a blistering rash. It occurs in individuals who have a history of varicella infection (chickenpox). Frequently the pain is present before the eruption of the blisters.

Diagnosis and Treatment

The diagnosis is most often made clinically; however, cultures can be used to determine the presence of the virus. The early use (within 72 hours of eruption of rash) of antiviral medications such as acyclovir (Zovirax) can decrease the risk of developing postherpetic neuralgia.

Management of the neuralgia includes both topical and systemic therapies. Transdermal lidocaine and capsaicin cream may be helpful in mild cases. Anticonvulsants and tricyclic antidepressants are also useful. NSAIDs and opioids may be mildly helpful as well. There is also a vaccine that adults may receive that will prevent or lessen the severity of a varicella zoster outbreak and decrease the risk of postherpetic neuralgia.

ISCHEMIC PAIN

Pain resulting from a sudden or profound loss of blood flow to the tissues in a particular part of the body may result in ischemic pain. Decreased perfusion leads to tissue hypoxia and injury, with release of inflammatory and pain-producing chemicals. Ischemic pain is described as aching, burning, or prickling (paresthesia). The symptoms of ischemic pain depend on the origin of the ischemia. For example, pain of cardiac origin is visceral and radiates to the arm or jaw. This pain is perceived as being deep, aching, diffuse, and pressing. Ischemia resulting from acute deep venous occlusion is also aching and has a deep quality and gradual onset. Acute arterial occlusion may be felt as either burning or aching, but has a sudden onset.

Chronic ischemic pain can occur in atherosclerotic syndromes. Arteriosclerosis obliterans occurs gradually as plaque develops in the intima of the arteries, most often arteries of the lower extremities. In the early stages, the pain, called *intermittent claudication,* is associated with physical activity, is alleviated with rest, and has a cramping quality. In severe cases, ischemic neuropathy may ensue and cause a more consistent burning, shooting pain in the leg or foot.

Management of ischemic pain is directed at improving blood flow and reducing tissue hypoxia. Acute ischemia is usually associated with a thrombus or embolus and can be managed with drugs to dissolve the clot or surgery to remove it. Chronic ischemia is most often associated with atherosclerosis and may be improved through lifestyle changes, including smoking cessation, weight loss, reduction of lipid levels, and regular exercise. Surgical bypass procedures or placement of intravascular stents are other therapeutic modalities.

REFERRED PAIN

Referred pain is perceived in an area other than the site of the injury. It is often felt at some distance from the point of nociceptor activation. A familiar example is the pain of myocardial infarction that is felt in the jaw or left arm. Other examples of referred pain include shoulder pain after pelvic procedures, diaphragmatic irritation from peritonitis, and cutaneous abdominal pain experienced with visceral irritation or tension. Common patterns of referral are shown in Fig. 47.9. Pain is generally referred to other structures in the same sensory dermatome. Convergence of nociceptors from internal organs with somatic afferents from the body surface occurs in the dorsal horn of the spinal cord. The brain cannot differentiate the two sources of pain signals and tends to attribute the visceral pain to a body surface location. Patterns of referred pain are fairly uniform and can be used to help locate a source of visceral pathologic process.

KEY POINTS

- Acute pain results from tissue injury and generally resolves when the injury resolves. The clinical manifestations result from activation of the sympathetic nervous system (elevated heart rate, blood pressure, and respiratory rate; dilated pupils; perspiration; and pallor).
- Headaches are a common but disabling cause of acute pain. Migraine headaches are caused by an interaction between neurotransmitters and cerebrovascular mechanisms and may be triggered by factors such as stress, foods, and sleep deprivation. There are a variety of treatments for headaches, but all must be initiated early in the course of the headache.
- Chronic pain lasts several months beyond the expected healing time and is often not associated with sympathetic manifestations of pain owing to physiologic adaptation. Instead, changes in personality or lifestyle may occur. Individuals may experience acute and chronic pain simultaneously, as commonly occurs in advanced cancer.
- FMS is a poorly understood cause of chronic pain. It is more common in women and has many associated signs and symptoms. FMS is best treated with multiple approaches, both pharmacologic and nonpharmacologic.
- Neuropathic pain results from injury to peripheral or central nerves as a consequence of surgery, tumor, trauma, or drugs and has a constant, achy, or shocklike quality. The sympathetic nervous system may maintain neuropathic pain by releasing norepinephrine onto nociceptors.

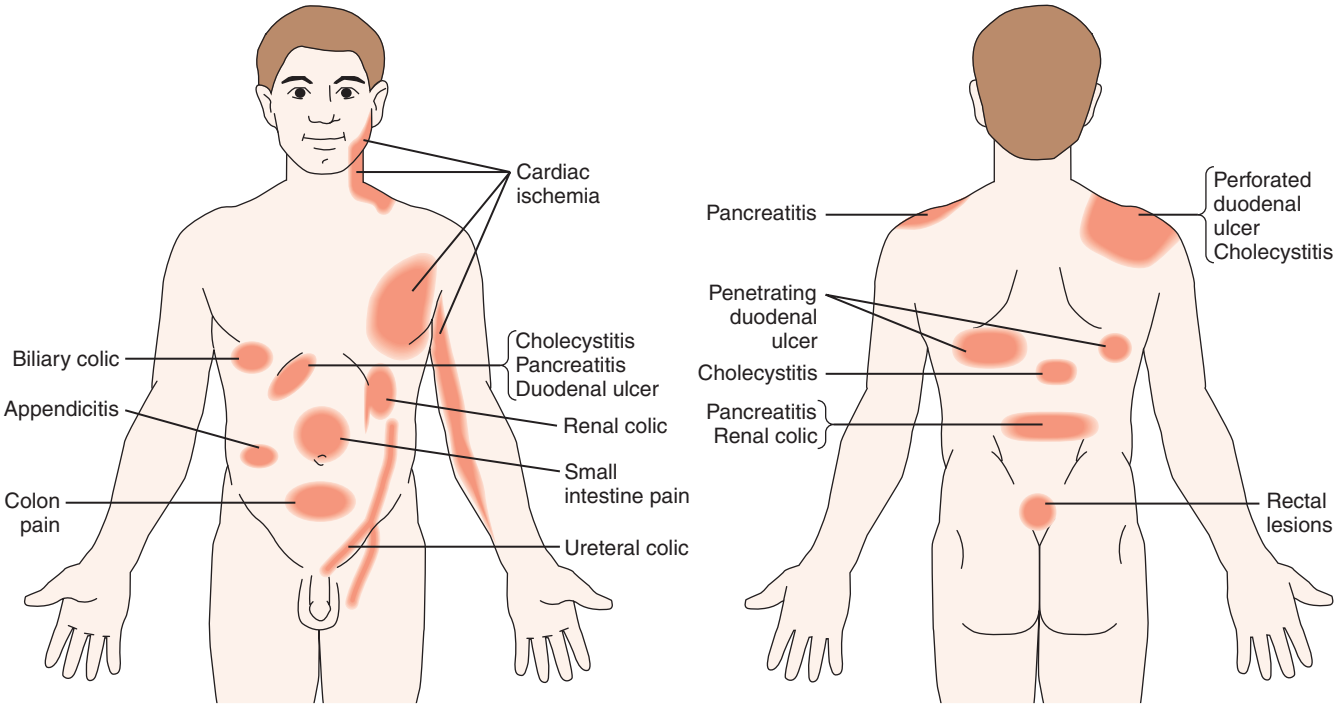


FIG 47.9 Areas of referred pain.

- Ischemic pain occurs when inadequate blood flow to tissues results in cellular injury and release of chemicals that stimulate and/or damage nociceptors.
- Referred pain is a painful sensation perceived at some distance from an injury but generally within the same dermatome. Referred pain is thought to occur because of the convergence of visceral nociceptor activity with primary somatic afferents in the posterior horn of the cord.

PHYSIOLOGIC RESPONSES TO PAIN

The autonomic nervous system, which is responsible for much of the physiologic response to pain, includes both the sympathetic and parasympathetic divisions. Activation of the sympathetic nervous system results in a predictable cluster of physical signs and symptoms, including an elevated heart rate, blood pressure, and respiratory rate, as well as dilated pupils, perspiration, and pallor (see Table 47.4). Sympathetic stimulation results in constriction of superficial vessels to divert blood to striated muscle, heart, and lungs; bronchodilation; increased cardiac contractility; and increased levels of circulating blood glucose. In addition, although gastrointestinal motility and secretion decrease, sphincter tone increases. Nausea, vomiting, and even paralytic ileus may develop. Hypomotility of the bladder and ureters can also result from sympathetic activation and lead to urine retention. Pain stimulates the release of numerous stress hormones, including antidiuretic hormone, aldosterone, and cortisol. These hormones help the kidneys conserve fluid and stimulate the release of glucose from the liver.

The sympathetic responses to pain may be physiologically helpful in the short term but become deleterious if excessive or prolonged. The body cannot sustain this level of activation for long periods. Eventually, physiologic adaptation occurs and the observed sympathetic response to pain abates. Thus the heart rate, blood pressure, and respiratory rate return toward normal or baseline. Signs and symptoms of sympathetic nervous system activation may be an important clue in pain assessment when present, but their absence does not guarantee the absence of pain sensations.

PAIN IN THE YOUNG AND THE ELDERLY

Many myths and misconceptions surround the issue of pain, especially in the very young and the elderly. As a result, often the young and the old receive inadequate treatment of their pain. It was previously thought that neonates were unable to perceive pain. Because their CNSs had not yet fully developed and they were unable to recall painful events, neonates often did not receive pain medication or anesthesia for surgery. It has since been found that infants do indeed have pain perception and that inadequate pain control may lead to persistent behavioral changes and physical changes in the CNS, especially in pain modulation and reactivity. These changes may persist into adulthood.

In the elderly, it has also been theorized that pain perception is decreased. Studies suggest that pain threshold increases with age, but that there is a decrease in the spread or magnitude of brain activation in response to pain. Tolerance to pain appears to remain unchanged as individuals age. Social expectations interfere with the adequacy of pain control in the elderly because pain is often an expected part of aging. Cognitive factors also hinder pain treatment, especially in patients with dementia who are unable to communicate their need for pain medication. However, it has been found that pain has a significant effect on an elderly individual's quality of life. No matter what the age of the patient, adequate pain control is important to his or her care.

TREATMENT MODALITIES

PHARMACOLOGIC AND NONPHARMACOLOGIC PAIN MANAGEMENT

Many pain management strategies are available. By understanding the basic mechanisms of pain transmission, one can readily identify potential

sites where various types of treatment modalities could interrupt pain transmission and perception. Pain management interventions can be directed at three points: (1) interrupting peripheral transmission of nociception, (2) modulating pain transmission at the spinal cord level, and (3) altering the perception and integration of nociceptive impulses in the brain.

Interrupting Peripheral Transmission of Pain

Modalities that interrupt the peripheral transmission of nociceptive impulses are often the first step in controlling pain. The basic action of splinting an injured limb or area of the body alters the peripheral transmission of pain by minimizing or reducing tissue injury. Applying heat or cold to an injured area also helps reduce peripheral nociception by altering blood flow to the area or by reducing swelling.

Pharmacologic treatments such as NSAIDs or local anesthetic agents also exert their analgesic effects by interrupting peripheral transmission at an early stage. NSAIDs and local anesthetic agents are used as a primary intervention for pain management. Inhibition of prostaglandin production by NSAIDs reduces the number of pain chemicals available to stimulate nociceptors in the peripheral tissues. NSAIDs include indomethacin, ibuprofen, naproxen, sulindac, piroxicam, ketorolac, and many others. Blocking the production and action of prostaglandins is not without side effects. For example, prostaglandins are responsible for maintenance of the gastric mucosa, and blocking their actions can result in gastrointestinal bleeding. Prostaglandin inhibition can also lead to decreased platelet aggregation and renal insufficiency. Knowledge of the risks and prescribing guidelines is essential for safe patient care, especially for prolonged periods. Many of these agents are available over the counter, and patient teaching regarding benefits and precautions may be needed.

Local anesthetic agents can be applied either to nerve endings at the site of injury or to the nerve plexus supplying the area. By providing localized or regional blockade, peripheral pain transmission is interrupted. Local anesthetic agents diminish or block conduction of the nociceptive impulses by blocking sodium influx during phase 0 of the action potential. The degree of blockade achieved with local agents depends on the amount of drug applied and hence the extent of sodium channel blockade. Local infiltration of a wound or surgical site with local anesthetic agents such as bupivacaine or lidocaine is a common practice, even when the patient is also receiving a general anesthetic.

Modulating Pain Transmission at the Spinal Cord

Numerous procedures and agents are used to modulate pain transmission at the level of the spinal cord. Nonpharmacologic techniques that inhibit pain transmission include several types of cutaneous stimulation. Cutaneous stimulation activates and recruits large sensory fibers that can block the central progression of nociceptive transmission at the interneurons. Examples of cutaneous stimulation include transcutaneous electrical nerve stimulation, massage, acupuncture, application of heat or cold, and therapeutic touch.

Pharmacologic measures that act at the level of the spinal cord include epidural and intrathecal analgesia. Spinal analgesia can be achieved with opioids, local anesthetics, and α-adrenergic blocking agents. Intraspinal opioids work by binding with opioid receptors in the posterior horn of the spinal cord, thereby decreasing the release of neurotransmitters such as substance P. Intraspinal local anesthetic agents block nerve conduction at the posterior nerve root. Epidural administration of an α-adrenergic blocking agent such as clonidine is thought to achieve analgesic effects by blocking sympathetically mediated pain transmission.

Dorsal column stimulators, sometimes used in chronic pain management, also work at the level of the spinal cord to "close the pain gate" by modulating descending input from the brain to the spinal cord.

Altering the Perception and Integration of Pain

The traditional modality for managing moderate-to-severe pain is the administration of systemic opioids. This pharmacologic intervention has stood the test of time. Opioids work at specific receptor sites that are located throughout the body but are highly concentrated in the brain. Opioid analgesic agents such as morphine and other derivatives alter the perception of pain by the brain. Opioid analgesics have similar mechanisms of action but vary widely in potency. This difference in potency has led to the development of equianalgesic tables to help clinicians prescribe these drugs appropriately (Table 47.6).

Opioid administration is associated with numerous side effects that may limit effectiveness (nausea, vomiting, respiratory depression, constipation). Long-term use of opioids leads to physical dependence and tolerance. Although the incidence of opioid addiction in persons experiencing acute pain is very low, fears about addiction contribute to inadequate pain therapy. Physical dependence is characterized by withdrawal symptoms if treatment is stopped abruptly. Tolerance to opioids is characterized by the need for increasing dosages to achieve the same analgesic effect. Dependence and tolerance are expected responses to long-term opioid therapy. Drug addiction is a behavioral pattern characterized by craving and preoccupation with obtaining the drug.

Nonpharmacologic techniques of pain management include such activities and procedures as distraction, guided imagery, relaxation, biofeedback, and hypnosis. With distraction, the number of generalized stimuli reaching the brain increases. Because the brain has a limited capacity to sort and attend to multiple and varied stimuli, it is less able to integrate the pain experience when other competition is present. Imagery may alter the perception of painful stimuli in the higher centers of the brain and produce relaxation as well as analgesia. Biofeedback is a conditioned response that can be learned as a pain-control strategy. Biofeedback is thought to control pain by increasing blood flow (usually as a consequence of relaxation) to targeted body areas. The increased blood flow decreases the concentration of pain-inducing chemicals in the area. Biofeedback may also increase the amount of endorphins produced and released.

TABLE 47.6 Equianalgesic Table for Common Opioid Analgesics

	APPROXIMATE EQUIANALGESIC DOSE	
Drug	**Oral**	**Parenteral**
Morphine	30 mg q3–4h	10 mg q3–4h
Codeine	60 mg q4h	30–50 mg q4h
Controlled-release morphine (MS Contin)	30–60 mg q12h	—
Hydrocodone	5–10 mg q4–6h	—
Hydromorphone (Dilaudid)	2–4 mg q4h	0.5–0.8 mg q4h
Meperidine (Demerol)	300 mg q2–3h	75 mg q3h
Methadone	20 mg q6–8h	10 mg q6–8h

Data from *Equianalgesic dosing of opioids for pain management.* Available at PL Detail-Document, Equianalgesic Dosing of Opioids for Pain Management. Pharmacist's Letter/Prescriber's Letter. August 2012.

A combination of nonpharmacologic and pharmacologic strategies may help reduce the need for high doses of medications. The choice of drug therapy should correspond to the severity of the pain. It has been recommended that mild pain be managed with nonopioid analgesics such as NSAIDs or acetaminophen, whereas moderate pain may require low-potency opioids such as codeine. Severe pain requires larger and more potent doses of opioids like morphine and fentanyl. The value of combination therapy in blocking pain transmission at multiple sites has been recognized. In the arena of chronic pain management, further research is delving into the central and peripheral sensitization theories. Questions remain as to why these develop and which patient populations are at highest risk for developing these complications. Great opportunity lies in the development of novel therapies for the control of chronic pain.

KEY POINTS

- Treatment is aimed at moderating pain transmission at specific points along the pain pathways. Potential sites of pain moderation are at the peripheral nociceptor, spinal cord, and brain.
- Nociceptor activation can be altered by prostaglandin inhibitors (NSAIDs), heat and cold, and local anesthetics that block sodium influx through fast channels.
- Spinal cord transmission can be altered by cutaneous stimulation (gate control theory), intraspinal analgesics (opioids, local anesthetics, α-adrenergic blockers), and dorsal column stimulators.
- The perception of pain can be altered within the brain by systemic opioids and by nonpharmacologic means such as hypnosis, distraction, and biofeedback.

SUMMARY

The human experience of pain, although unpleasant, is a normal and expected phenomenon in response to injury. Pain sensations alert the individual to a physiologic problem and help ensure that timely treatment is sought. However, prolonged severe pain serves no good purpose and can be physiologically and psychologically harmful. Appropriate efforts to alleviate pain may enhance recovery from illness and prevent the development of some types of chronic pain syndromes. As understanding of pain physiologic mechanisms grows, treatment strategies can more effectively combine the best of pharmacologic and nonpharmacologic therapies.

RESOURCES

Pain Physiology

Barrett KE, Boitano S, Barman SM, Brooks HL: Chapter 8. Somatosensory neurotransmission: touch, pain, and temperature. In Barrett KE, Boitano S, Barman SM, Brooks HL, editors: *Ganong's review of medical physiology*, ed 24, New York, NY, 2012, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=393&Sectionid=39736748. (Accessed 13 December 2015).

Grosser T, Smyth E, FitzGerald GA: Chapter 34. Anti-inflammatory, antipyretic, and analgesic agents; pharmacotherapy of gout. In Brunton LL, Chabner BA, Knollmann BC, editors: *Goodman & Gilman's the pharmacological basis of therapeutics*, ed 12, New York, NY, 2011, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=374&Sectionid=41266242. (Accessed 13 December 2015).

Hall JE, editor: *Guyton and Hall textbook of medical physiology*, ed 13, Philadelphia, 2016, Elsevier.

Iasp-pain.org. IASP Taxonomy - IASP, 2015. Available at: http://www.iasp-pain.org/Taxonomy?nav/temNumbers=576#pain. (Accessed 13 December 2015).

McCaffery M: *Nursing practice theories related to cognition, bodily pain and man-environment interactions*, Los Angeles, 1968, University of California, (master's thesis).

Melzack R, Wall PD: Pain mechanisms: a new theory. *Science* 150(699): 971–979, 1965.

Melzack R, Wall PD, editors: *The challenge of pain*, ed 2, Harmondworth, UK, 1988, Penguin.

Ropper AH, Samuels MA, Klein JP: Chapter 8. Pain. In Ropper AH, Samuels MA, Klein JP, editors: *Adams & Victor's principles of neurology*, ed 10, New York, NY, 2014, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=690&Sectionid=50910867. (Accessed 13 December 2015).

Schaible H: Nociceptive neurons detect cytokines in arthritis. *Arthritis Res Ther* 16(5):2014. doi:10.1186/s13075-014-0470-8.

Schoell ED, Bingel U, Eippert F, et al: The effect of opioid receptor blockage on the neural processing of thermal stimuli. *PLoS ONE* 5(8):e12344, 2010. Available at: www.plosone.org.

Pain Syndromes

Beal M, Hauser SL: Trigeminal neuralgia, Bell's palsy, and other cranial nerve disorders. In Kasper D, Fauci A, Hauser S, et al, editors: *Harrison's principles of internal medicine*, ed 19, New York, NY, 2015, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=1130&Sectionid=79756043. (Accessed 13 December 2015).

Burmeister J, Holle D, Bock E, et al: Botulinum neurotoxin type A in the treatment of classical Trigeminal Neuralgia (BoTN): study protocol for a randomized controlled trial. *Trials* 16(1):550, 2015. doi:10.1186/s13063-015-1052-z.

Burstein R, Noseda R, Borsook D: Migraine: multiple processes, complex pathophysiology. *J Neurosci* 35(17):6619–6629, 2015. doi:10.1523/jneurosci.0373-15.2015.

Choy ES: Chapter 14. The patient with diffuse pain. In Imboden JB, Hellmann DB, Stone JH, editors: *Current diagnosis & treatment: rheumatology*, ed 3, New York, NY, 2013, McGraw-Hill. http://accessmedicine.mhmedical.com.proxy.heal-wa.org/content.aspx?bookid=506&Sectionid=42584897. (Accessed 13 December 2015).

Doan L, Manders T, Wang J: Neuroplasticity underlying the comorbidity of pain and depression. *Neural Plast* 2015:1–16, 2015. doi:10.1155/2015/504691.

Fashner J, Bell A: Herpes zoster and postherpetic neuralgia: prevention and management. *Am Fam Physician* 83(12):1432–1437, 2011.

Harding A, Clark L: Pediatric migraine. *Nurse Pract* 39(11):22–31, 2014. doi:10.1097/01.npr.0000454980.88918.f0.

Isasi C, Colmenero I, Casco F, et al: Fibromyalgia and non-celiac gluten sensitivity: a description with remission of fibromyalgia. *Rheumatol Int* 34(11):1607–1612, 2014. doi:10.1007/s00296-014-2990-6.

Johnson R, McElhaney J: Postherpetic neuralgia in the elderly. *Int J Clin Pract* 63(9):1386–1391, 2009. doi:10.1111/j.1742-1241.2009.02089.x.

Kandasamy R, Price T: The pharmacology of nociceptor priming. *Handb Exp Pharmacol* 227:15–37, 2015. doi:10.1007/978-3-662-46450-2_2.

Lisi L, Aceto P, Navarra P, Dello Russo C: mTOR kinase: a possible pharmacological target in the management of chronic pain. *Biomed Res Int* 2015:1–13, 2015. doi:10.1155/2015/394257.

Schreiber A: Diabetic neuropathic pain: physiopathology and treatment. *World J Diabetes* 6(3):432, 2015. doi:10.4239/wjd.v6.i3.432.

Smith H, Harris R, Clauw D: Fibromyalgia: an afferent processing disorder leading to a complex pain generalized syndrome. *Pain Physician* 14:E217–E245, 2011.

Tesfaye S, Boulton A, Dickenson A: Mechanisms and management of diabetic painful distal symmetrical polyneuropathy. *Diabetes Care* 36(9):2456–2465, 2013. doi:10.2337/dc12-1964.

Vranken J: Elucidation of pathophysiology and treatment of neuropathic pain. *Cent Nerv Syst Agents Med Chem* 12(4):304–314, 2012. doi:10.2174/187152412803760645.

Who.int: WHO | Headache disorders, 2015. Available at: http://www.who.int/mediacentre/factsheets/fs277/en/. (Accessed 13 December 2015).

Wolfe F, Brähler E, Hinz A, Häuser W: Fibromyalgia prevalence, somatic symptom reporting, and the dimensionality of polysymptomatic distress: results from a survey of the general population. *Arthritis Care Res* 65(5):777–785, 2013. doi:10.1002/acr.21931.

Zychowska M, Rojewska E, Przewlocka B, Mika J: Mechanisms and pharmacology of diabetic neuropathy – experimental and clinical studies. *Pharmacol Rep* 65(6):1601–1610, 2013. doi:10.1016/s1734-1140(13)71521-4.

Age Considerations

Hatfield LA: Neonatal pain: what's age got to do with it? *Surg Neurol Int* 5(Suppl 13):S479–S489, 2014. doi:10.4103/2152-7806.144630.

Mohn-Brown E, Burke KM, Eby L, editors: *Medical-surgical nursing care*, ed 3, Upper Saddle River, NJ, 2011, Pearson Education.

Paladini A, Fusco M, Coaccioli S, et al: Chronic pain in the elderly: the case for new therapeutic strategies. *Pain Physician* 18:E863–E876, 2015.

48

Neurobiology of Psychotic Illnesses

Ann Futterman Collier and Samantha Cody Russell

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- What are the "positive" and "negative" symptoms of schizophrenia?
- What genetic, gestational, and neurologic risk factors are related to schizophrenia?
- How are the dopamine D_1 and D_2 receptors related to positive and negative symptoms of schizophrenia?
- How is schizophrenia managed?
- What are the hallmark symptoms of major depression?
- What is the neurobiology of major depression?
- What are the subtypes of bipolar disorder?
- How are major depression and bipolar disorder managed?

CHAPTER OUTLINE

Schizophrenia, 972
Etiology and Neurobiology, 972
Dopamine effects, 972
GABAergic interneuron origin, 973
Genetic effects, 974
Gestational effects, 974
Marijuana use in adolescents and schizophrenia, 974
Neurologic effects, 975
Clinical Manifestations, 976
Pharmacologic Treatment, 976
Nonpharmacologic Treatment, 977
Major Depressive and Persistent Depressive Disorders, 978
Etiology and Neurobiology, 979
Clinical Manifestations, 980
Pharmacologic Treatment, 981
Nonpharmacologic Treatment, 981
Bipolar Disorder, 982
Etiology and Neurobiology, 982
Clinical Manifestations, 983
Pharmacologic Treatment, 983
Nonpharmacologic Treatment, 984
Population Considerations, 984
Women and Mental Illness, 984
Cultural Considerations, 984
Geriatric Considerations, 984

Psychotic disorders are characterized by altered thought processes that interfere with perceptions of reality. Current internationally recognized clinical diagnostic schemes rely on categorical systems to define mental disorders and provide standardized criteria for each diagnosis. The two most widely established systems are (1) *ICD-10,* Chapter V: Mental and Behavioral Disorders, which is part of the *International Classification of Diseases* published by the World Health Organization, and (2) the *Diagnostic and Statistical Manual of Mental Disorders (DSM),* published by the American Psychiatric Association. Although historically there have been significant differences between these systems, in recent years there has been a convergence so that both sets of codes are broadly comparable. In October 2015, the entire U.S. health care system transitioned to using the ICD-10-CM code set, with the aim of providing concise diagnostic information and allowing more access to care for those suffering from mental health conditions. More than 300 different psychiatric disorders are listed in the fifth text revision of *DSM (DSM-5),* which was published in May 2013. Future DSM versions will continue to be guided more heavily by research evidence and attempt to maintain continuity with previous editions.

The *National Survey on Drug Use and Health (NSDUH)* defines "serious mental illness" (SMI) as a mental, behavioral, or emotional disorder (excluding developmental and substance use disorders) that is diagnosable currently or within the past year; is of sufficient duration to meet diagnostic criteria specified within the DSM systems; and results

TABLE 48.1 Lifetime Prevalence of Serious Mental Illness (in U.S.)

Name of Disorder	Overall Classification	12-Month Prevalence (%)	Average Age of Onset (yr)	Treatment and Services Use (Any Type) (%)
Social phobia	Anxiety disorders	6.8%	13	45.6%
Major depression	Mood disorder	6.7%	32	56.8%
Attention deficit disorder with hyperactivity (ADHD)	Attention deficit disorder (13–18 yr old)	9.0% (lifetime)	Not reported	Not reported
Posttraumatic stress disorder (PTSD)	Anxiety disorders	3.5%	23	57.4%
Generalized anxiety disorder	Anxiety disorders	3.1%	31	52.3%
Avoidant personality disorder	Personality disorders	5.2%	Not reported	Not reported
Panic disorder	Anxiety disorders	2.7%	24	65.4%
Bipolar disorder	Mood disorders	2.6%	25	55.5%
Obsessive-compulsive disorder	Anxiety disorders	1.0%	19	Not reported
Borderline personality disorder	Personality disorders	1.6%	Not reported	42.4%
Schizophrenia	Schizophrenia	1.1%	Not reported	64.3%
Antisocial personality disorder	Personality disorders	1.0%	Not reported	46.1%
Autism spectrum disorders	Autism spectrum disorders	1.5%	8	Not reported

From National Institutes of Mental Health (NIMH). Available at www.nimh.nih.gov/statistics/index.shtml. Accessed April 8, 2017.

in serious functional impairment that substantially interferes with or limits one or more major life activities. Using the SMI nomenclature, approximately 4.2% of all adults in the United States meet criteria for SMI. It is also known that 26.2% of the U.S. adult population will have experienced an SMI in the previous year, and 22.3% of these cases will be classified as "severe." In addition, 46.3% of adolescents between 13 and 18 years of age will be diagnosed with an SMI. The research also shows that SMIs occur more frequently in women than in men, are more prevalent in individuals between the ages of 18 and 49, and are least likely to be seen in American individuals of Asian descent. Table 48.1 provides a review of the most common SMIs by prevalence and the average age of onset for each disorder. Some SMIs have very high prevalence but only moderate treatment service use. Alternatively, other SMIs have low prevalence but very high treatment service utilization. Frequently occurring SMI disorders are reviewed in Chapters 48 and 49.

Psychosis is a term used to describe a serious and debilitating mental state or condition. The most restricted definition refers to delusions and prominent hallucinations, which occur in the absence of insight into their pathologic basis. The broadest definition of psychosis includes other positive symptoms that are characteristics of schizophrenia, such as disorganized speech and grossly disorganized or catatonic behavior. Hallmark symptoms of psychosis are delusions, hallucinations, cognitive disorganization, and altered reality. These symptoms characterize a small number of specific mental disorders; however, a wide range of different physical and mental conditions can produce psychotic symptoms. The neurobiological basis of psychosis can be summarized as acute or chronic alterations in neuron anatomy and physiology and cellular biochemical processes. Although psychosis occurs most often with schizophrenia and mood disorders, it may also manifest with substance disorders, delirium, dementia, amnestic disorders, and acute stress disorder. The groundbreaking research in brain imaging techniques and psychopharmacology now allows precise definitions of the various biochemical pathways associated with psychosis. This chapter addresses three disorders that are associated with psychosis: schizophrenia, major depressive disorder (MDD), and bipolar disorder (BD).

SCHIZOPHRENIA

Schizophrenia refers to a chronic, remitting, and relapsing psychotic disorder that is associated with significant impairment in social and vocational functioning. Literally, the term is defined as "split mind," and once was believed to be a disorder that caused the personality to split into multiple subtypes. Schizophrenia is now correctly understood as a split or separation among normally well-synchronized brain functions. This loss of synchronized brain functioning leads to thoughts, behaviors, and feelings that are disordered, disorganized, and disconnected from reality—a condition generally referred to as *psychosis*.

The global incidence rate of schizophrenia has consistently been estimated to be about 1% of the world population, and is fairly equally distributed across genders. Unfortunately, schizophrenia appears to be associated with an average lifespan reduction of 15 to 25 years. Women show symptom onset and are diagnosed typically between the ages of 25 and 35 years, men between the ages of 15 and 25 years. Women with schizophrenia appear to have better outcomes than men. Possibly, the later age of onset, the protective nature of female hormones such as estrogens, or an improved drug response accounts for this difference. People with schizophrenia may be at increased risk for type 2 diabetes and cardiovascular disorders, perhaps attributable to the side effects of antipsychotic medications, poorer overall physical health, less healthy lifestyles, and substandard health care. However, the newer antipsychotic medications alone do not appear to account for the increased incidence of diabetes and cardiovascular disease.

Etiology and Neurobiology

Dopamine effects

Several decades ago it was hypothesized that abnormalities in dopaminergic pathways in specific regions of the brain were the cause of schizophrenia. This conclusion was reached after noting that dopamine antagonists reduce symptoms of schizophrenia, whereas dopamine agonists (mimetic) produce schizophrenic symptoms. Dopamine-specific neurons in the brain primarily are located in the ventral tegmentum of the mesencephalon, medial and superior to the substantia nigra. These regions, as a whole, are referred to as the *mesolimbic dopaminergic system*. The long nerve fibers leaving this system mainly project into the medial and anterior portions of the limbic system. The limbic system contains three powerful centers of behavior control: the nucleus accumbens, the amygdala, and the anterior caudate nucleus.

Generally speaking, decreased neurotransmission and connectivity is the neurobiological basis of schizophrenia. Frankle and colleagues reported that the core biochemical process of schizophrenia involved

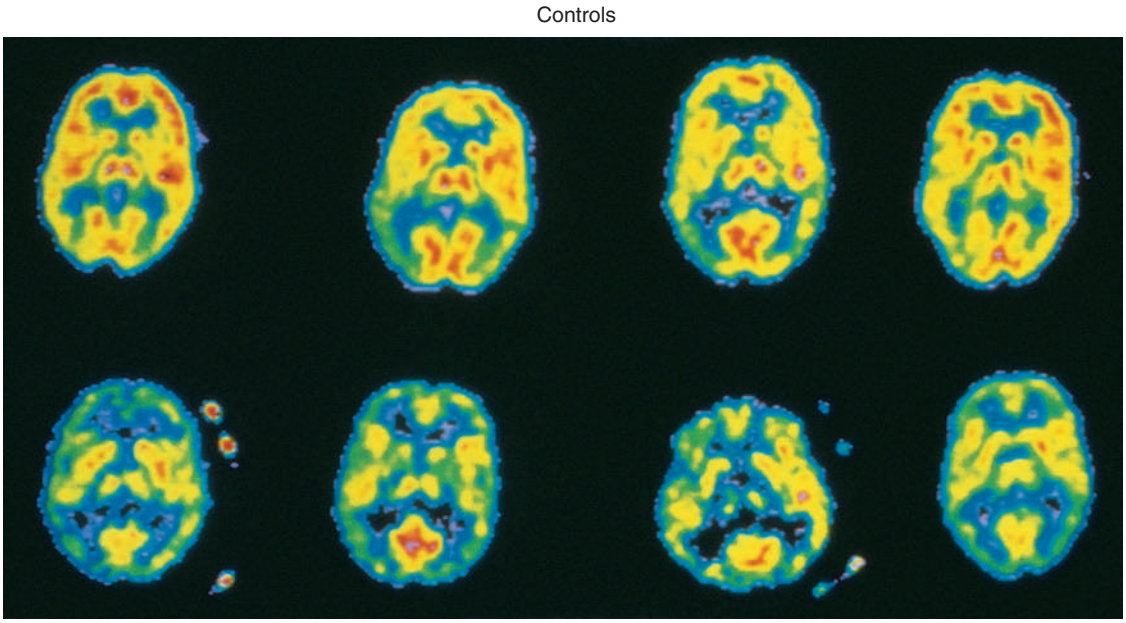


FIG 48.1 Individual variation in positron emission tomography (PET) scans. Four normal individuals *(top row)* and four schizophrenics *(bottom row)* show range of hypofrontality and diminished basal ganglia metabolism. (From Buchsbaum MS, Haier RJ: Functional and anatomic brain imaging: impact on schizophrenia research, *Schizophr Bull.* 1987;13(1):115–132. Reproduced by permission of Monte S. Buchsbaum, MD.)

an excess of subcortical dopaminergic transmission at dopamine D_2 receptors and a deficit of glutamate transmission at *N*-methyl-D-aspartate (NMDA) receptors. The dopamine pathogenesis of schizophrenia can be thought of as disordered synaptic organization. In the brain, normal synaptic organization is necessary for normal communication among brain cells. Specific schizophrenia symptoms have been associated with neurotransmission dysregulation that diminishes or elevates dopamine activity. Hyperdopaminergic states have been associated with "positive" symptoms of schizophrenia, such as hallucinations, delusions, and psychosis. Alternatively, hypodopaminergic states have been associated with "negative" symptoms of schizophrenia, such as cognitive difficulties, lack of energy and motivation, and depression. Dopamine D_2 receptors have been associated with positive symptoms and psychosis, whereas dopamine D_1 receptors have been associated with negative symptoms.

Dopamine receptor activity and synaptic transmission are subject to a variety of mediators such as brain-derived neurotrophic factor (BDNF), a neurotrophin that increases synaptic activity and neurotransmitter output. Dopamine synaptic activity also has been closely linked with stress-related cortisol activity and drugs of abuse (cocaine, amphetamines, morphine, nicotine, and ethanol). Although dopamine dysregulation clearly is the driving force behind the neurochemical processes of psychosis, other experts suggest considering the role of the biopsychosocial contextual factors in shaping and aggravating the expression of biochemical dysregulation.

Positron emission tomography (PET) brain images have enabled researchers to further examine the overall role of dopamine activity in psychosis and schizophrenia. PET studies have demonstrated low glucose metabolism rates in the frontal cortex and dopamine regions of the brains of persons with schizophrenia (Fig. 48.1). Of the many different types of dopamine brain cell receptors identified, dopamine D_2 receptors once again were found to be strongly associated with symptoms of schizophrenia. Similar research has shown that dopamine D_2 receptors are particularly responsive to antipsychotic drugs.

Thus the dopamine hypothesis, which postulates that schizophrenia symptoms result from presynaptic dysregulation of dopamine transmission, continues to be the focus of neurobiological studies. The specific neuropathologic mechanism of schizophrenia appears to be caused by a functional excess of postsynaptic dopamine receptor activity and dopamine receptor hypersensitivity, either alone or in combination.

GABAergic interneuron origin

The systematic administration of *N*-methyl-D-aspartic acid-type receptor (NMDAR) antagonists has been shown to reliably induce schizophrenic-like symptoms in otherwise healthy patients. This effect is thought to occur by the induction of the hypofunction of NMDA glutamate receptors, as the administration of excessive NMDAR antagonist inhibits the receptor's function. Although it is clear that this process does produce schizophrenic-like symptoms, the mechanism that produces the symptoms that follow NMDAR antagonist administration is unclear. One potential hypothesis of the physiologic cause of these symptoms is that of a GABAergic origin.

GABAergic interneurons are thought to play an instrumental role in postnatal maturation of neural circuitry. These circuits are highly immature at birth, and their refinement (which occurs primarily through GABAergic inhibition) continues well into the postadolescent period. If the maturation of GABAergic neurons is impaired, it is plausible that psychiatric disorder (primarily schizophrenia) could develop as the result. More specifically, NMDAR hypofunction at GABAergic interneurons produces schizophrenic symptoms by disrupting GABAergic function. Under close observation, the presence of NMDAR hypofunction appears to produce a fast-spiking firing pattern in GABAergic interneurons, disrupting the control and synchronizing of disparate cortical circuits. As corticolimbic fast-spiking interneurons have been shown to be highly vulnerable to not only genetic predispositions but also to early environmental exposures (including oxidative stress and

excitotoxicity), the physiologic events described here may help to explain the significant contribution of NMDAR antagonists to the presence of schizophrenic symptoms.

Genetic effects

A family history of schizophrenia is identified in many patients diagnosed with schizophrenia; however it can develop in persons with no family history of the disease. Initial interest in identifying possible genetic contributions to schizophrenia was based on early research on monozygotic (identical) and dizygotic twins by studying twins born to parents with schizophrenia but reared apart from their parents. Results suggested that nearly 50% of these offspring developed schizophrenia. Children of two parents with schizophrenia have a 40% to 68% risk of developing the illness, whereas children with one parent with schizophrenia have a 9% to 16% risk. The risk of a non–twin sibling of a brother or sister with schizophrenia developing the disease is slightly less, at 8% to 14%. Many persons with genetic risk do not develop schizophrenia, indicating that factors other than genetics must impact disease expression.

Several additional conditions are thought to be involved in the transformation from genetic risk to actual illness. Factors such as prenatal infections, malnutrition, birth complications, and brain injury have been associated with the development of schizophrenia in persons who have increased genetic risk. Moreover, different gene locations are also relevant to the illness onset. Experts question whether inherited genetic risk for schizophrenia could explain observed differences in incidence and prevalence rates based on gender and race. Lastly, genetic models of schizophrenia typically do not address risk in terms of the subtype of schizophrenia, such as paranoid schizophrenia as opposed to schizophrenia without paranoia. Given the high standards of proof required for gene typing, a purely genetic explanation of an illness as complex as schizophrenia seems unlikely. The increased genetic risk appears to be a critical part of a puzzle composed of many pieces. Thus for individuals who are genetically predisposed, some aspects of schizophrenia will probably be determined by biopsychosocial characteristics. Researchers continue to study the actual versus potential genetic risks of schizophrenia.

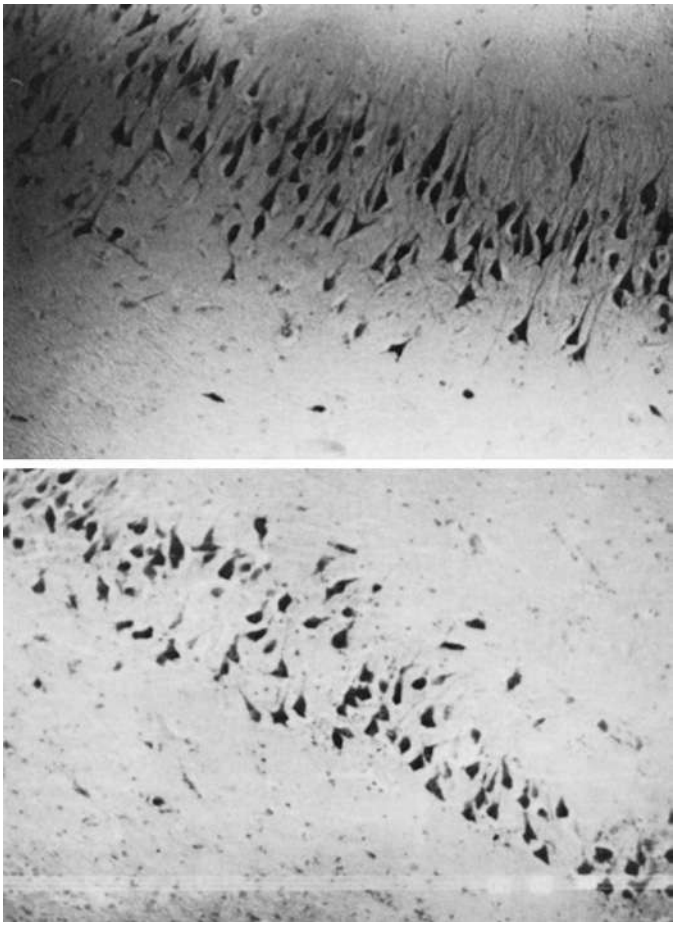


FIG 48.2 Photographic comparison of hippocampal tissue at CA 2/3 interface in a control *(top)* and in a person with chronic schizophrenia *(bottom).* (Original magnification of the Nissl-stained tissue, ×100. From Kovelman JA, Scheibel AB: A neurohistological correlate of schizophrenia, *Biol Psychiatry.* 1984;19:1601–1621.)

Gestational effects

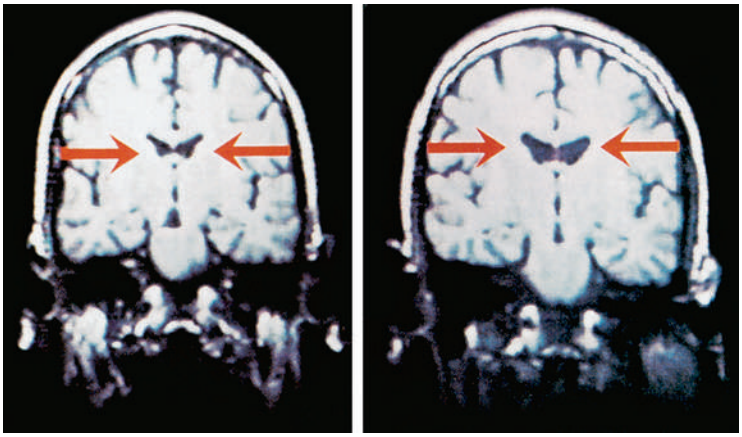
Early findings noted that in persons with schizophrenia, pyramidal cells in the hippocampus were not lined up like a "picket fence," as they were in control subjects (Fig. 48.2). Instead, the cells appeared to be rotated at 70-degree to 90-degree angles. Pyramidal cells migrate during the second trimester of gestation and later become fastened by neuronal cell adhesion molecules (N-CAMs). Researchers focused on the possibility of pyramidal cell misalignment and lost N-CAM adhesive effects as links between gestation and schizophrenia. Researchers noted that in pregnant women living in Scandinavia and England who were exposed to the 1957 flu epidemic during their second trimester, 300% more of their children were diagnosed with schizophrenia than those of women who experienced flu during the first or third trimester.

These findings were considered significant for two reasons: (1) neuronal migration peaks during the second trimester, and (2) the influenza virus is one of very few viruses that produce capsular neuraminidase, an enzyme that can change the adhesive properties of N-CAMs. Hippocampal, parahippocampal gyrus, and amygdala neurons process information and emotional expression. Other researchers have suggested that delivery complications could be an additional factor, playing a mediating role between prenatal influenza exposure and later development of schizophrenia.

More recent studies of gestational abnormalities focus on specific brain regions and stages of prenatal neurodevelopment. This research is exemplified in a well-controlled Canadian study that examined the association of hippocampal formation abnormalities and first adult episode of schizophrenia. The main functions of the hippocampus are learning and new-memory formation. Both abilities may be lost when hippocampal functioning is impaired. Macroscopic cell abnormalities, such as fewer synapse connections and diminished synapse activity, result in reduced hippocampal volume. Reduced hippocampal volume has been associated with severe stress, mood disorders, and schizophrenia. The Canadian researchers hypothesized that schizophrenia was associated with incomplete formation of the hippocampus during the second trimester of development. They used magnetic resonance images to compare the brains of newly diagnosed patients with healthy matched controls. Although the number of participants in the study was too small to allow for generalizations to be drawn, the magnetic resonance images of the newly diagnosed patients clearly showed enlarged hippocampal fissures or disrupted hippocampal formation. Interestingly, obstetric complications during pregnancy were not significant factors.

Marijuana use in adolescents and schizophrenia

From the prenatal period of development through childhood and adolescence, the brain remains in a state of experience-guided development,



Unaffected Affected

FIG 48.3 Loss of brain volume associated with schizophrenia is clearly shown by magnetic resonance images comparing the size of ventricles (butterfly shaped, fluid-filled spaces in the midbrain) of 44-year-old male identical twins, one of whom has schizophrenia *(right).* The ventricles of the person with schizophrenia are larger, suggesting structural brain changes associated with the illness. Note that such magnetic resonance images cannot be used to diagnose schizophrenia in the general population because of normal genetic variation in ventricle size; many unaffected people have large ventricles. (From Fortinash KM, Worret PA, editors: *Psychiatric mental health nursing,* ed 5, St Louis, 2012, Mosby, p 264. Courtesy Daniel R. Weinberger, MD, Chief Researcher, Clinical Brain Disorders Branch, National Institute of Mental Health, Bethesda, MD.)

rendering the brain more vulnerable (compared with a mature brain) to environmental hazards. Among several environmental insults shown by research to adversely affect brain development is exposure to tetrahydrocannabinol (or THC), and particularly during the aforementioned vulnerable period of development. THC is the primary active ingredient in marijuana, which is a commonly used drug in the United States. According to the 2013 NSDUH, nearly 46% of U.S. teens will have tried marijuana at least once in their lifetime, and 23% of high school seniors are current users of marijuana (with 6% using the drug daily).

Though several studies suggest a relationship between the use of marijuana among adolescents and the development of mental illness, the strongest evidence is to the development of psychotic disorders like schizophrenia and in those with a preexisting genetic vulnerability. One study suggested that the genetic variant in the *AKT1* gene influences the risk of individuals developing psychosis if they use cannabis regularly. This particular gene codes for an enzyme that affects dopamine signaling in the striatum, which is an area of the brain that is flooded with dopamine in the presence of particular stimuli. Another study suggested that those who habitually used marijuana during adolescence and who also carried the variant for the gene catechol-o-methyltransferase (COMT) are at increased risk for developing schizophrenia. COMT is primarily responsible for the degradation of neurotransmitters such as dopamine and norepinephrine. This research strongly underscores the notion that the developmental vulnerability that is present at a young age (i.e., during adolescence) in concert with certain genetic predispositions may pose serious risks for individuals who use marijuana often.

Neurologic effects

Observed neuroanatomic differences in persons with schizophrenia led researchers to study anatomic and functional abnormalities in the limbic region and frontal lobe parts of the brain. Structural abnormalities in these brain regions would suggest that abnormal functioning might contribute to the disrupted cognitive processes or symptoms of schizophrenia. Beginning in 1976, computed tomography studies revealed enlarged brain ventricles in persons with schizophrenia. A groundbreaking 1980 study showed that the neurochemical basis of schizophrenia might involve two processes: dopamine neurotransmission dysregulation and abnormal cerebral structure.

Magnetic resonance imaging (MRI) studies of persons with chronic schizophrenia indicated larger-than-normal lateral and third ventricles and reduced temporal lobe gray matter (Fig. 48.3). MRI findings also showed reduced frontal lobe blood flow and relative decreases in frontal lobe metabolic activity. Prefrontal cortex structure and functioning deficits were consistently observed when the subject was simultaneously placed under stress. The stress used in the study was psychological, such as contingency planning exercises or divergent thinking during performance of a cognitive task that utilized specific regions of the prefrontal cortex.

Subsequent studies of brain structure in persons with severe schizophrenia replicated the earlier findings of abnormal limbic-cortical structures and smaller, misarranged hippocampal pyramidal cells. This includes replication of MRI findings of reduced bilateral temporal lobe volume, decreased hippocampal volume, and reduced volume in the parahippocampal gyrus region of the brain. Studies of brain regions other than the limbic system revealed frontal lobe structural alterations in the dorsolateral area of the prefrontal cortex and in the cingulate and motor cortices. The finding of decreased frontal lobe glucose metabolism associated with schizophrenia also has been replicated (Fig. 48.4).

Postmortem brain tissue studies indicated fewer nicotinic receptors present in the hippocampus of schizophrenics. Tobacco dependence is a common secondary disorder with schizophrenia. The finding of fewer nicotinic receptors was of particular interest to researchers in that previous studies had shown that smoking could temporarily normalize auditory sensory gating that typically becomes impaired with schizophrenia. Since then, researchers found that a neurophysiologic deficit at chromosome 15, at the *OC7* nicotinic receptor gene, may partially explain the inheritance of this neurophysiologic symptom. Although

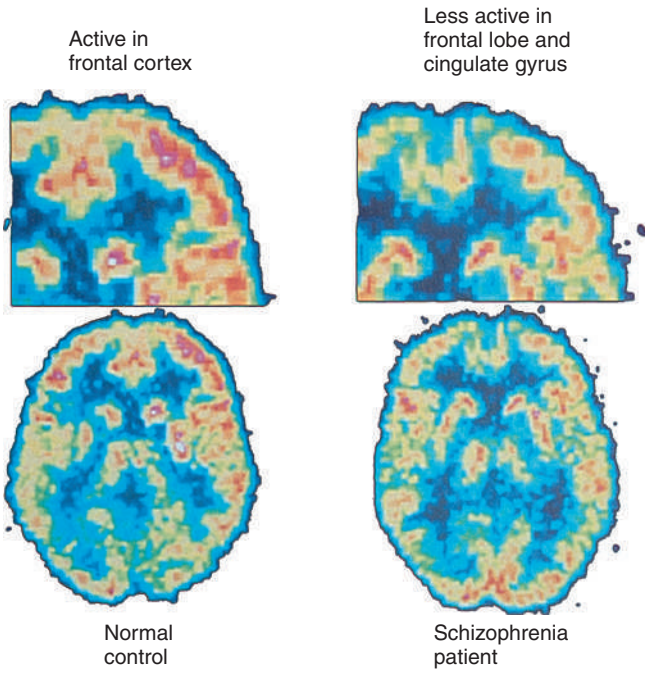


FIG 48.4 Positron emission tomography (PET) scan with 18F-deoxyglucose shows metabolic activity in a horizontal section of the brain in a control subject *(left)* and in an unmedicated client with schizophrenia *(right)*. Reduced red and yellow indicate lower activity in the white matter of the schizophrenia patient. The frontal lobe is magnified to show reduced frontal activity in the prefrontal cortex of the client with schizophrenia. (From Fortinash KM, Worret PA, editors: *Psychiatric mental health nursing,* ed 5, St Louis, 2012, Elsevier, p 279. Courtesy Monte S. Buchsbaum, MD, Mt. Sinai School of Medicine, New York.)

the development of schizophrenia is associated with a number of specific abnormalities, early neurodevelopmental alterations that result in dysfunction of the limbic and prefrontal regions of the brain appear to be critical.

Clinical Manifestations

Schizophrenia is characterized by positive (reality distortion and disorganization), negative, cognitive, and mood symptoms. The types and severity of symptoms differ among patients and change over the course of the illness. Positive symptoms typically reflect an excess or distortion of normal functions. In contrast, negative symptoms reflect a decrease or loss of normal functions. *Positive symptoms* include the psychotic dimension, or distortions in thought content (delusions) and perception (hallucinations), as well as the disorganization dimension, or disorganization in speech and behavior. The resulting positive symptoms also include the inability to self-monitor behavior, which results in grossly disorganized or catatonic behavior. *Delusions,* or systematic, fixed, false beliefs, usually involve themes of persecution, reference, somatization, religiosity, or grandiosity. *Hallucinations* are sensory perceptions with no apparent stimulus. They occur in any sensory system (such as auditory, visual, olfactory, gustatory, or tactile), but usually auditory hallucinations are the norm. Auditory hallucinations are commonly experienced as voices, distinct from the person's own thoughts and out of the range of normal experience. *Disorganized thinking* is usually evaluated by an individual's speech and is frequently characterized by frequent derailment or loose associations, invented words, tangential ideas, and, when most severe, incomprehensible speech. Grossly disorganized behavior can range from childlike silliness to unpredictable agitation and impairs the individual's ability to complete tasks of daily living. When individuals display catatonic motor behaviors, they show a decrease in reactivity to environmental events to such an extreme that they can maintain a rigid posture and resist efforts to be moved. The positive symptoms of schizophrenia are thought to result from excessive dopamine D_2 receptor activity in the brain. *Negative symptoms* are considered to be restricted affect, or avolition and asociality. Negative symptoms represent deficits in functioning and can be more difficult to recognize than positive symptoms. Negative symptoms of schizophrenia are considered to be associated with dopamine D_1 receptor activity in the brain. Poor *cognitive functioning* in schizophrenia includes difficulties with memory, attention (e.g., poor concentration, distractibility, selective attention), and decision-making.

According to the workgroup responsible for developing *DSM-5,* the active phase of schizophrenia is characterized by two or more of the following symptoms that must be present for a significant portion of time during a 1-month period (or less if successfully treated): (1) delusions, (2) hallucinations, (3) disorganized speech, (4) grossly abnormal psychomotor behavior, and (5) negative symptoms. At least one of these symptoms should include delusions, hallucinations, or disorganized speech. In addition, for a significant portion of the time since the onset of the disturbance, schizophrenia is characterized by social/occupational dysfunction in one or more major areas: work, interpersonal relations, or self-care. All of these are markedly below the level achieved before the onset of the illness (or when the onset is in childhood or adolescence, failure to achieve the expected level of interpersonal, academic, or occupational achievement). Regarding duration, typically continuous signs of the disturbance exist for at least 6 months, with at least 1 month of symptoms (or less if successfully treated) that meet criteria for the active phase (delusions, hallucinations, disorganized speech) and may include periods of prodromal or residual symptoms. During these prodromal or residual periods, the signs of the disturbance may be manifested by only negative symptoms or by two or more of the active-phase symptoms in an attenuated form (e.g., odd beliefs, unusual perceptual experiences). Finally, for an individual to receive the diagnosis of schizophrenia, the diagnoses of schizoaffective disorder and mood disorder with psychotic features must be ruled out. In addition, symptoms cannot be due to the direct physiologic effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition.

Schizophrenia was previously divided into five subtypes in *DSM-IV-TR*: paranoid, disorganized, catatonic, undifferentiated, and residual. In each subtype there is predominant symptomatology at the time of evaluation; however, these subtypes were eliminated in the DSM-5 as these categories showed minimal reliability, limited diagnostic stability, and poor validity. The DSM-5 takes a more dimensional approach to assessing the severity of symptoms associated with schizophrenia in an effort to capture the heterogeneity of symptom type and intensity expressed across individuals with psychotic disorders.

Pharmacologic Treatment

When untreated, schizophrenia is associated with increased mortality, impaired vocational and social functioning, and reduced quality of life. Unfortunately, the extent to which treatment does actually improve life span and psychosocial functioning is not specifically clear. Because antipsychotic medications do decrease the likelihood of relapse, they are encouraged.

When effective, antipsychotic medication can be expected to diminish or remit hallucinations or reduce their impact on functioning. The ability to reason should improve; ambivalence, delusions, and suspiciousness should be greatly reduced; agitation and confusion should be

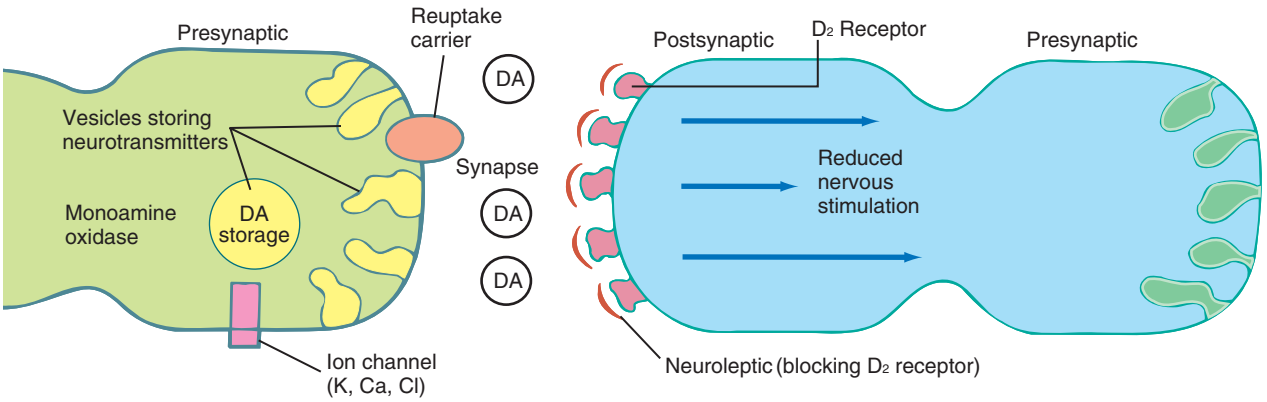


FIG 48.5 Neuroleptic (antipsychotic) action. Dopamine (DA) action at the synapse is modified by neuroleptics, which block postsynaptic receptor sites to reduce nervous stimulation (reducing symptoms of schizophrenia). (Adapted from Fortinash KM, Worret PA, editors: *Psychiatric mental health nursing,* ed 5, St Louis, 2012, Elsevier, p 265.)

relieved; and social behavior should improve. Antipsychotic medications are designed to have specific effects on targeted neurotransmitters. Newer antipsychotic medications, typically called *second-generation agents (SGAs)* (atypical), are intended to generally manage psychosis as well as both positive and negative schizophrenia symptoms. Older antipsychotic medications, traditionally called *first-generation agents (FGAs)* (conventional), were less effective in managing negative schizophrenia symptoms. They mainly acted as dopamine antagonists, blocking the dopamine receptors (D_2). This action diminishes the amount of dopamine received by the receptor sites (Fig. 48.5).

All FGAs and SGAs appear to be equally effective in reducing the positive symptoms and disorganization associated with schizophrenia, hence blocking the dopamine D_2 receptor. Neither antipsychotic agent appears to improve the cognitive symptoms in patients with schizophrenia; in fact, findings suggest that they worsen cognitive impairment. Although the maximum drug response may not be reached for many months, patient response over the first 2 to 4 weeks of antipsychotic medication use will typically predict long-term response. Antipsychotic response will vary as a result of the stage of illness, with first-episode patients responding faster and at a higher rate than those at later stages. In addition, both FGAs and SGAs play a substantial role in decreasing the likelihood of relapse.

FGAs and SGAs differ most notably in their side effect profiles and their potential for interacting with other medications. FGAs (e.g., chlorpromazine) had relatively nonspecific neurotransmitter effects and numerous side effects, with the most difficult being extrapyramidal symptoms (EPS). Haloperidol was the first FGA that had significantly fewer anticholinergic and hypotensive side effects. By 1990 SGAs became widely available; the SGAs have generally shown a lower risk of EPS but a higher risk of other metabolic adverse effects. The first SGA, clozapine, was quickly followed by a generation of new atypical antipsychotics (risperidone, olanzapine, quetiapine, ziprasidone), each promising still fewer side effects, better relief of both positive and negative schizophrenic symptoms, and minimal risk of EPS, specifically those associated with tardive dyskinesia. Some worrisome side effects are clinically significant weight gain, glucose dysregulation, and dyslipidemia. Experts have speculated that hyperglycemia may have to do with dopamine receptor involvement in the regulation of insulin secretion. Numerous explanations for weight gain as a major side effect of antipsychotic medication continue to be developed and tested. The glutamate neurobiological model of psychosis and schizophrenia, an alternative to the dopamine model, may be the basis for the next new generation of antipsychotic medications.

The glutamate deficit model attempts to focus attention on the cause of excessive dopamine receptor activity rather than the excessive activity itself. γ-Aminobutyric acid (GABA) is the most important inhibitory brain neurotransmitter. GABA synthesis depends on and is controlled by the enzyme glutamic acid decarboxylase (GAD). GAD activity is modulated by the glutamate receptor NMDA. GAD dysregulation is thought to lead to insufficient GABA activity and, consequently, excessive dopamine activity. Reduced GAD activity has been observed in the dorsolateral prefrontal cortex and hippocampus of patients with schizophrenia. The interesting observation for pharmacologic researchers is that a broad range of different drugs has been shown to be capable of affecting GAD activity. Although the glutamate deficit model is not new, the model still may lead to a class of antipsychotic medications unlike any previous generation.

Dopamine D_2 continues to remain the lead neurotransmitter target of the atypical antipsychotics. Researchers found that the most common side effects of antipsychotic medication could be reduced if the drug's impact on dopamine D_2 receptors was not excessive. Lower receptor occupancy, less receptor affinity, and faster release of the receptor were methods shown to be associated with fewer side effects. More recently, the aim for effective antipsychotic medication is to stabilize rather than reduce dopamine activity. The newest SGAs, such as aripiprazole, paliperidone, iloperidone, asenapine, and lurasidone, show greater affinity for serotonin receptors (negative symptoms) and moderate affinity for dopamine and norepinephrine receptors (positive symptoms). Effective antipsychotic medications have significant dopamine effects; however, schizophrenia symptoms are highly complex and likely to involve other neurotransmitters, particularly serotonin and norepinephrine (Table 48.2).

Nonpharmacologic Treatment

Psychosocial treatments that offer integrated care are essential for all patients with schizophrenia. At a minimum this should include case management in order to maintain the system of care, permit the most efficacious treatment in the least restrictive setting, and optimize quality of life and social function. However, cognitive-behavioral therapy (CBT), cognitive remediation and rehabilitation, social skills–based therapies, vocational rehabilitation, and family therapy have all proven to be effective supplemental care for schizophrenia.

TABLE 48.2 Neurotransmitters in Schizophrenia: Type and Function

Neurotransmitter	Type	Function
Dopamine	Catecholamine	Regulates motor behavior in extrapyramidal nerve tracts and also transmits in cortex. Increases vigilance and may increase aggression. Excess may produce psychosis; deficiency may cause movement disorders (EPS).
Serotonin	Indolamine	Brainstem transmitter; modulates mood; lowers aggressive tendencies. Deficiency may be responsible for some forms of schizophrenia.
Acetylcholine	Cholinergic	Transmits at nerve–muscle connections (central nervous system and autonomic nervous system). Deficiency may increase confusion and acting-out behavior. Controls EPS.
Norepinephrine	Catecholamine	Transmits in sympathetic nervous system. Induces "fight or flight" syndrome (hypervigilance). May be insufficient in clients with schizophrenia who display anhedonia (loss of pleasure).
Cholecystokinin	Peptide	Excites limbic neurons. Deficiency is related to avolition (lack of motivation) and flat affect.
Glutamate	Amino acid	Excitatory neurotransmitter. Impairment in *N*-methyl-D-aspartate affects glutamate metabolism, which can lead to problems with cognition, delusions, and possibly some negative symptoms of schizophrenia.
γ-Aminobutyric acid (GABA)	Amino acid	Inhibitory neurotransmitter; predominantly a brain transmitter. Promotes balance between dopamine and glutamate and thus inhibits impulsive behaviors.

EPS, Extrapyramidal symptoms.
From Fortinash KM, Worret PA: *Psychiatric mental health nursing,* ed 5, St Louis, 2012, Elsevier, p 265.

KEY POINTS

- *DSM* and *ICD* are the two main systems for the classification of mental illness; *DSM-IV-TR* was revised as *DSM-5,* which was published in 2013. There are more than 300 different types of mental illness classifications today.
- Schizophrenia is characterized by altered perceptions of reality and disordered thinking. Genetic predisposition and environmental factors are thought to interact to produce biological changes in the brain, particularly in the hippocampus, temporal lobes, and dopaminergic pathways that project to the limbic system. Exposure to influenza virus during the fifth to sixth months of gestation appears to predispose to schizophrenia.
- The average age at onset for schizophrenia is 15 to 25 years for men and 25 to 35 years for women. There is a higher incidence in industrialized societies.
- The positive symptoms of schizophrenia are thought to be due to excessive dopamine D_2 receptor activation in the brain. Disorganized thinking (inability to connect thoughts logically), disorganized speech (rambling, tangentiality), delusions (fixed system of false beliefs), and hallucinations (sensory perceptions when no apparent stimulus exists) are typical positive symptoms. Delusions are often persecutory, grandiose, or controlling. Hallucinations are most often auditory but may also be visual, olfactory, gustatory, or tactile. Positive symptoms respond to drugs that decrease dopamine activity in the brain (e.g., olanzapine, quetiapine).
- Negative symptoms are thought to be mediated by dopamine D_1 receptors in the brain. Drugs that block D_1 receptors (e.g., clozapine) may alleviate some of the negative symptoms, which include social withdrawal, flat affect, and poverty of speech; and ritualistic posturing.
- Hypofunction of NMDA receptors at GABAergic interneurons may disrupt the control and synchronization of cortical circuitry, which may describe the physiologic mechanism by which excessive administration of NMDAR antagonists can produce schizophrenic-like symptoms.
- Evidence suggests marijuana use during adolescence and genetic predispositions, specifically involving the *ATK1* and *COMT* genes, may be linked to the presence of some psychotic illnesses, including schizophrenia.

MAJOR DEPRESSIVE AND PERSISTENT DEPRESSIVE DISORDERS

The DSM-5 revised its categories of depression to include persistent depressive disorder, which now encompasses both major chronic depressive disorder and dysthymic disorder. The diagnostic criteria required for a diagnosis of persistent depressive disorder includes a depressed mood that lasts for at least 2 years. Furthermore, the DSM-5 includes several new depressive disorders, including premenstrual dysphoric disorder and disruptive mood dysregulation.

MDD, once referred to as *endogenous depression* (or depression that arises from innate characteristics of the person), is now understood to be a complex illness involving inherited genetic susceptibility and symptoms associated with specific alterations in brain structures and functioning. MDD is a highly common disorder in the general population, affecting nearly 3% to 5% of the U.S. adult population at any given time. With a 17% lifetime risk, it is often seen frequently in general medical settings. The disorder commonly co-occurs with a number of medical illnesses, which can complicate diagnosis and potentially impair the patient's ability to adhere to medical care. Nearly half of all cases of MDD are identified and diagnosed in primary care settings. About 51% of people with MDD receive psychological treatment, with only 21% receiving care that is consistent with American Psychological Association guidelines.

Some people experience a seasonal pattern to depression (i.e., seasonal affective disorder *[SAD]*), where depression is more likely to occur in the fall or winter months. This occurs in one third of the cases of depression. MDD can develop as a serious secondary illness or illness complication. For example, as early as 1937, researchers were able to show that cardiac patients with severe depression had higher cardiac death rates than their counterparts. Recent studies confirm that depression continues to be a significant risk factor for 1-year mortality rates in myocardial infarction patients. Similarly, systematic reviews and longitudinal studies validate that patients with comorbid renal disease, cardiovascular disease, and cancer demonstrate significant improvement in overall morbidity and mortality when depressive symptoms are effectively treated.

Significant rates of MDD are observed across all ages, races, education, and income groups. Globally, females are diagnosed with MDD two times more often than males. Depressed individuals with a family history of MDD have a onefold to twofold increased risk for developing dysthymia and anxiety disorders. The complexity of managing severe depression in adolescents and older adults has made these population groups particularly concerning. As a result of the potentially serious disability, morbidity, and mortality risks directly associated with MDD, the World Health Organization has ranked major depression among the top five global health problems.

Etiology and Neurobiology

Improved neurobiological research techniques have allowed for advances in the understanding of the neurobiological mechanisms of MDD. Nevertheless, the specific *cause* of MDD remains unknown. Table 48.3 summarizes mood disorder risk factors.

Cognitive processing models of depression describe dysfunctional thoughts and beliefs that occur during depression. These dysfunctional thoughts are believed to result from underlying schemas, or ways in which attention, memory, and information are organized with a negative bias. The depressed person focuses on negative information, emotions, and memories and selectively attends to negative material with automatic negative thoughts. This reinforces the continuation of the dysfunctional depressive schemas, creating a self-referent bias toward negative thoughts and expectations. When the depressed person ruminates, everything is interpreted through a negative filter. Research suggests that depressive states are indeed associated with slower cognitive processing, impaired attention, and bias toward negative stimuli; these appear to abate as depression goes into remission.

Neurobiological changes associated with MDD are thought to involve neurotransmission dysregulation, altered hippocampal and prefrontal cortex cell structure and functioning, and impaired hypothalamic–pituitary–adrenal (HPA) system activation (Table 48.4). Based on observations of low central nervous system levels of serotonin (Fig. 48.6) in persons with severe symptoms of depression, the basic neurobiology of depression has been hypothesized to be reduced brain serotonin neurotransmission activity, either through excessive presynaptic uptake or stress-related down-regulation of postsynaptic receptors. Chronic or persistent vulnerability to depression is thought to be related to decreased hippocampal volume or capacity and suppressed hippocampal neurogenesis.

The prefrontal cortex is also thought to play a role in the neurobiology of depression, as with the ventromedial cortex. Studies utilizing MRI show a reduction in the brain volume of depressed patients compared with healthy controls, with particularly marked reductions in the anterior cingulate and orbitofrontal cortex and moderate reductions in the putamen, caudate, and hippocampus. PET studies have revealed abnormalities in cerebral blood flow and glucose metabolism in numerous

TABLE 48.3 Risk Factors for Major Depression and Bipolar Disorders

Risk Factor	Major Depression	Bipolar Disorder
Lifetime prevalence (%)	Overall: 5.2–17.1 Females: 7.4–21.3 Males: 2.8–12.7	Bipolar I: 0.9–1.6 Bipolar II: 0.3–3 Spectrum: 3–6
Female/male ratio (U.S.)	1.7–2.6	No significant differences
Age at onset (yr) (U.S.)	23.8–25.6	18–27
Social class	No clear relationship	Slight increase in upper classes
Marital status	Divorced and separated have 2× increase compared with married or never married	More frequent in people with multiple divorces and unmarried
Race and ethnicity	No relationship	No relationship
Family history	Ranges from 14.7–24.2 per 100 in probands; depends on severity and age of onset of depression For monozygotic twins, ranges from 31%–71%; for dizygotic twins, ranges from 20%–25%; higher for females	Bipolar I 7× more frequent in relatives of proband Twin studies suggest 50%–60% increase

From Hirschfeld RMA, Weissman MM: Risk factors for major depression and bipolar disorder. In Davis KL, Charney D, Coyle JT, Nemeroff C, editors: *Neuropsychopharmacology: the fifth generation of progress,* Brentwood, TN, 2002, American College of Neuropsychopharmacology, pp 1017–1025.

TABLE 48.4 Prefrontal Cortex and Serotonin Interconnections: Implications in Depression

Interconnected Brain Structures	Hypothesized Role of These Interconnections in Depression
Prefrontal cortex	Covering the frontal lobes, it is unique within the central nervous system for its strong interconnections with all other areas of the brain; it receives information that has already been processed by other sensory areas and then merges this information with other emotional, historical, or relevant information, thus attending to both feelings and intellect.
Limbic system structures	Prefrontal cortex modulates limbic system activities (emotional and instinctive) by way of these three structures:
• Hippocampus	• Major importance in cognitive function, including memory
• Amygdala	• Major importance in modulating feelings such as aggression, anger, love, and shyness
• Cingulate gyrus	• Involved in motivation and interest
Brainstem	Responsible for regulating the general state of arousal and tone of brain function; also the location of structures that manufacture various neurotransmitters, such as serotonin (5-HT), norepinephrine (NE), and dopamine (DA).
Raphe nuclei	Located in the brainstem, they manufacture 5-HT; they also modulate excessive stimuli and the organization and coordination of appropriate responses to these stimuli.
Hypothalamus	This interconnection allows for direct prefrontal input into neuroendocrine function via the hypothalamic–pituitary axes.
Suprachiasmatic nucleus	Located in the hypothalamus, it regulates circadian (24-hr) rhythms and circannual rhythms; thus it is also implicated in seasonal affective disorder.

From Stuart GW, editor: *Principles and practice of psychiatric nursing,* ed 9, St Louis, 2009, Elsevier, p 292.

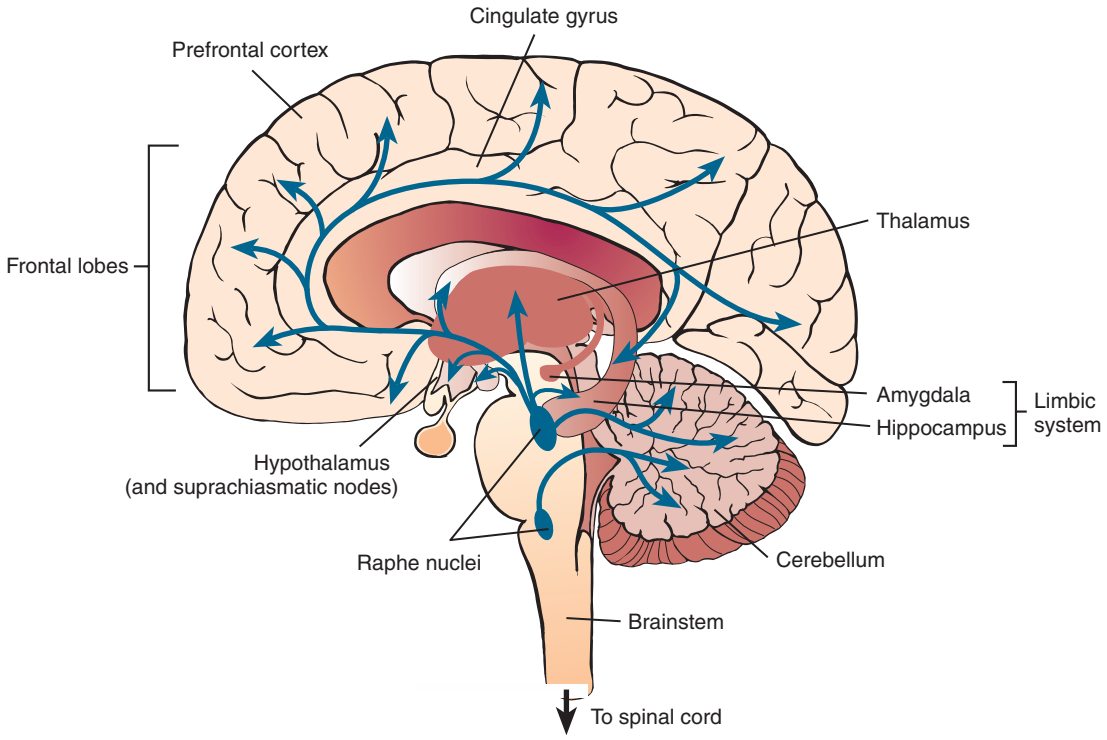


FIG 48.6 The serotonin neurotransmitter system implicated in depression. (From Stuart GW, editor: *Principles and practice of psychiatric nursing,* ed 9, St Louis, 2009, Elsevier, p 290.)

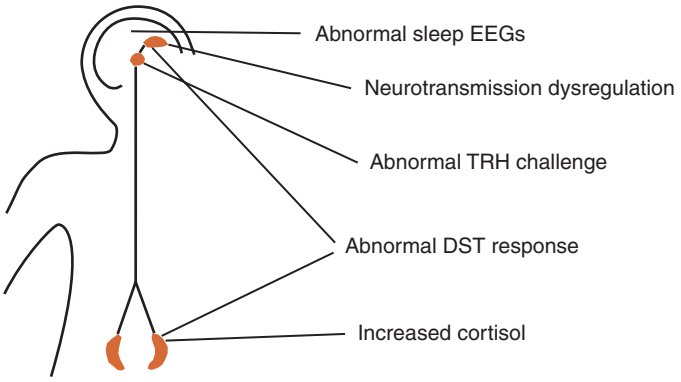


FIG 48.7 Biological factors related to depression. *EEGs,* Electroencephalograms; *DST,* dexamethasone suppression test; *TRH,* thyroid-releasing hormone. (From Stuart GW, editor: *Principles and practice of psychiatric nursing,* ed 9, St Louis, 2009, Elsevier, p 293.)

prefrontal cortical and limbic structures that are implicated in emotional processing.

Some of the biological factors associated with depression are illustrated in Fig. 48.7. HPA axis dysfunction has been associated with depression. Although no single gene has been identified as a cause of MDD, both family studies and twin studies suggest an increased risk of heritability. Genetic susceptibility to depression may also mediate the link between stress and depression.

Other theorists propose that increased stress hormone levels can lead to significant decreases in the expression of BDNF. Reduced BDNF activity has been shown to lead to hippocampal cell atrophy and, consequently, reduced neurotransmission activity in this area of the brain. According to the researchers, effective antidepressant medications seemed to improve BDNF activity and thereby promote the growth and survival of serotonin neurons in this vital brain region. Significant reduction in BDNF activity as a result of severe stress has also been demonstrated in research animals. The HPA model of depression supports the development of novel treatments that can target cortisol activity (e.g., mifepristone) rather than serotonin activity (e.g., selective serotonin reuptake inhibitors [SSRIs]).

Circadian rhythms (synchronizing cycles) are closely associated with symptoms of major depression. The pineal gland in the brain produces the hormone melatonin. Brain melatonin levels can fluctuate significantly, with annual and daily increases and decreases in light and dark periods. Melatonin helps regulate circadian rhythms; in turn, these cycles govern body temperature changes and the urge to sleep. Depressed persons have been shown to suffer from low melatonin levels when their symptoms include disturbed sleep. In theory, natural light acts as a zeitgeber, or a biological clock synchronizer, that is based on the 24-hour day–night cycle. It has been hypothesized that persons with SAD who are exposed to additional natural light will experience improvement in their sleep–wake cycle and, consequently, improved mental health.

Clinical Manifestations

To be diagnosed with MDD, an individual must experience five (or more) of the following criteria during the same 2-week period; this must also represent a change from previous functioning. In addition, at least one of the symptoms must be either depressed mood or loss of interest or pleasure. *Depressed mood,* occurring most of the day, nearly every day, is indicated by either subjective report (e.g., feels sad or empty) or observation made by others (e.g., appears tearful). In children and adolescents, depressed mood can be irritable mood. Depressed mood associated with MDD is qualitatively different from normal sadness or grief associated with loss; it is typically experienced as painful, numbing, and bottomless. Markedly *diminished interest or pleasure* in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation made by others),

is another hallmark symptom of MDD. Even if the depressed person does engage in activities that were previously enjoyable, interest and pleasure are not experienced. There is also significant *appetite disturbance* that results in weight loss (when not trying to diet) or weight gain (e.g., a change of more than 5% of body weight in a month). Depression is also associated with *insomnia or hypersomnia* nearly every day. Sleep symptoms include difficulty falling asleep and/or staying asleep, early-morning awakening, frequent awakenings, and waking feeling extremely tired. *Psychomotor agitation or retardation* can occur nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down). This can make normal daily activities difficult to perform. There is frequently *fatigue or loss of energy* nearly every day. Depressed people may associate their severe or sudden fatigue with serious physical illness such as cancer or cardiovascular disease. The fatigue can also make ordinary daily activities, once performed automatically, nearly impossible. Individuals experience *feelings of worthlessness or excessive or inappropriate guilt* (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick). Negative self-appraisals range from pointless guilt to self-hate. In some cases, the guilt reaches delusional proportions or is markedly disproportionate to actual misdeeds or perceived failings. Negative thinking, expressed as negative views of self, life, and the future, is a common symptom of depression. There is often a *diminished ability to think or concentrate, or indecisiveness,* nearly every day (either by subjective account or as observed by others).

Finally, individuals with MDD have *recurrent thoughts of death (not just fear of dying),* display recurrent suicidal ideation without a specific plan, adopt a specific plan for committing suicide, or actually complete a suicide attempt. The clinician should ask directly about suicidal ideation to determine the presence of these thoughts, the intention to act on these ideas, and the extent of plans or preparation for suicide. If a method to complete suicide is identified, the lethality, the patient's expectation about the lethality, and the accessibility to the means of committing the plan (e.g., availability of firearms or medications) should be assessed. The client should also be interviewed to determine psychiatric risk factors, including agitation, pervasive insomnia, and impulsiveness; comorbidities, including substance abuse, psychosis (especially with command hallucinations), or personality disorder; family history of suicide; previous history of suicide attempts; and history of recent exposure to suicide.

Psychosis associated with MDD is thought to result from extreme symptoms, prolonged symptom duration, or comorbid illness complications. Hallucinations (especially auditory), delusions (especially nihilistic or somatic), and disorganization may become prominent in psychotic depression and require treatment with antipsychotic medications. Usually the psychosis is egosyntonic, meaning that the hallucinations or delusions are consistent with their negative view of themselves, the world, and the future. Severely depressed older adults are at increased risk of depression-associated psychosis.

Pharmacologic Treatment

Most current-day antidepressants act by improving brain norepinephrine and serotonin activity (Fig. 48.8). Earlier generations of antidepressants had less specific effects on these neurotransmitters and significantly more side effects. Monoamine oxidase inhibitors (MAOIs) blocked the destruction of norepinephrine and serotonin once they were released into the synaptic cleft. Tricyclic antidepressants (TCAs) blocked the reuptake of norepinephrine and serotonin, thereby allowing more neurotransmitter activity. SSRIs include medications such as fluoxetine, paroxetine, and sertraline. Response rates vary across studies, but range from 60% to 75%. No single SSRI is more effective than another. Although the side-effect profile of SSRIs represents real improvement over earlier antidepressants, SSRIs are not free of side effects. Serotonin syndrome, a serious side effect that results from excessive serotonin activity, includes altered mental status, restless agitation, myoclonus, hyperreflexia, sweating (diaphoresis), shivering, tremor, gastrointestinal upset, ataxia, and headache. Sexual dysfunction, evidenced by loss of interest, impaired arousal, and anorgasmia, is also thought to result from SSRI-induced excessive serotonin activity. Less severe but equally troublesome side effects include gastrointestinal upset, headache, allergy, dry mouth, constipation, urination difficulties, diaphoresis, and significant weight gain. In general, it takes 3 to 6 weeks for depressed patients to notice improvements in symptoms after beginning any type of pharmacotherapy.

The SSRI cellular and neurochemical mechanisms continue to be the subject of a great deal of basic and clinical research. It would appear that the SSRIs share the basic action of boosting neurotransmission activity of mood-related monoamines. More recent research findings suggest that SSRIs may also alter the genetics-based expression of BDNF and cell neurogenesis in the hippocampus.

Nonpharmacologic Treatment

Electroconvulsive therapy (ECT) has been shown to be useful for patients who have inadequate response to two to three other antidepressant

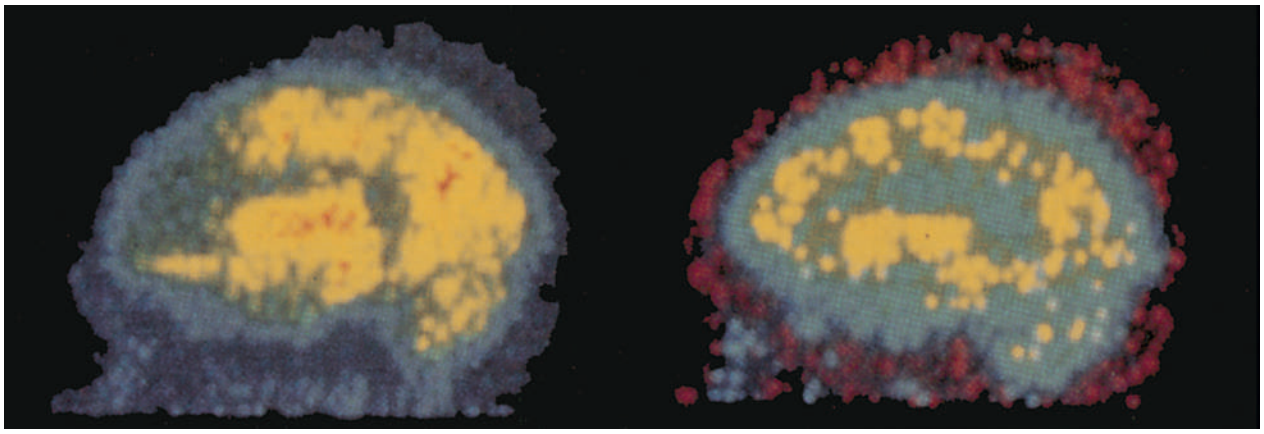


FIG 48.8 Positron emission tomography (PET) scan of glucose use in depressed subject *(figure on left)* showing frontal hypometabolism *(left side of figure).* This improves after treatment with antidepressant medication *(figure on right);* note increased glucose metabolism in frontal lobe *(left side of figure).* (From Stuart GW, editor: *Principles and practice of psychiatric nursing,* ed 9, St Louis, 2009, Elsevier, p 291.)

treatment modalities or require intensive therapy because of psychotic features or active suicidality; ECT is also efficacious in pregnant women who cannot receive pharmacotherapy. *Phototherapy,* which consists of 30 minutes per day of direct facial exposure to a 10,000-lux intensity full-spectrum white-light box, can augment antidepressant therapy or be used alone, especially when a seasonal pattern has been identified. Several *psychotherapeutic interventions* have been shown to be effective with depression, namely, CBT and interpersonal therapy. With mild-to-moderate uncomplicated MDD, these psychotherapies can be used as the first-line treatment option. When used in conjunction with pharmacologic agents, a more rapid therapeutic response can be found, especially when a client has severe, recurrent MDD. In addition, behavioral therapies combined with maintenance pharmacotherapy appear more successful at preventing relapse than use of pharmacotherapy alone.

BIPOLAR DISORDER

BD is a highly complex mood disorder characterized by recurring symptoms of depression and elation that can become severe enough to produce psychosis. The most recognizable course of illness with BD is the sudden onset of severe mania lasting from weeks to months. Although mania is the hallmark of BD, the depressive phase actually represents the greatest burden on patients, and it is more common for BD patients to present with depressive symptoms. BD patients experience three times more days with depression than do persons with any other mood disorders and frequently have a continuous presence of subthreshold depressive symptomatology. Because they are less likely to volunteer information about manic or hypomanic symptoms, accurate diagnosis of BD when the initial symptom profile is depression can be quite difficult or delayed. When the initial symptoms of mania are severe enough to produce psychosis, BD can be mistaken for schizophrenia.

Within the last decade, there has been a marked increase in the rate at which children are being assigned the diagnosis of BD. BD in children is characterized by irritability, cyclical mood changes, and associated attentional deficit disorder with hyperactivity. The clinical course is often more chronic and undulating, with fewer discrete mood episodes. Some researchers have suggested that severe, nonepisodic temporal dysregulation is characteristic of pediatric BD.

The two basic forms of bipolar disorder are bipolar I and bipolar II. A diagnosis of bipolar I disorder requires any past or present history of a full manic episode. There is usually a history of depressive episode(s) meeting criteria for MDD, but depression is not a requirement for the diagnosis of bipolar I disorder. Bipolar II disorder requires any past or present history of hypomania (never full mania) and a history of depression consistent with MDD. Thus what differentiates bipolar I from bipolar II disorder is the intensity of manic symptoms: bipolar I has a history of at least one full manic episode; bipolar II has a history of hypomanic episodes (never a full manic episode). Three additional types of BD have been proposed (types II1/2, III, and IV), which further differentiate types I and II and better represent the clinical complexity of bipolar symptoms. This suggests the concept of a bipolar spectrum and allows for earlier and more accurate diagnosis and treatment of persons who clearly suffer from mood dysregulation but who do not meet the narrow symptom criteria for bipolar I or bipolar II disorder. There are no known gender differences in the incidence of BD. The risk in the general population is 1%.

Etiology and Neurobiology

Decades of clinical and genetic research findings indicate that the increased risk of developing BD is both inherited and acquired, with genetic risk reported to range from 60% to 85%. Fig. 48.9 illustrates a unified model of mood disorders. Efforts to identify the neurobiological factors that characterize BD have yielded mixed results. First, the similarities between psychosis associated with mania, psychosis associated with depression, and psychosis associated with schizophrenia are a major hurdle. Second, the depressive symptoms associated with BD differ very little from the depressive symptoms that characterize MDD (although response to treatment differs). Third, when bipolar symptom onset is gradual, changes in behavior can be mistaken for personality disorder or substance use disorder.

Life stress and neuroticism are both robust risk factors for the later development of mood disorders (BD, in particular), and especially for those who have close family members who suffer from any type of mood disorder. In an effort to further explore the relationship between genetic vulnerability, disposition, environmental exposure to stress, and the development of BD, one study investigated the relationship between neuroticism, familial risk, and frontolimbic serotonin 2A (5-HT_{2A}) receptor binding. The research suggested that the personality trait neuroticism (which is thought to reflect stress vulnerability) was positively correlated with frontolimbic 5-HT_{2A}. Researchers first administered personality inventories (including the assessment of the personality characteristic neuroticism) to a group of healthy twins who had a co-twin with or without a history of BD. Participants then underwent a [^{18}F] altanserin positron emission tomography examination. Those participants whose co-twin had a history of a mood disorder (in other words, who were at "high risk") and who endorsed stronger neurtocisim characteristics were much more likely to show frontolimbic 5-HT_{2A} binding. These findings suggest a possible neurobiological mechanism between the genetic risk and personality features that links those attributes to the manifestation of BD.

The neurotransmission model of BD follows the basic catecholamine hypothesis for major depression and schizophrenia. Neurotransmission activity deficits are thought to promote depression symptoms associated with BD, whereas excessive activity is believed to promote symptoms of mania and psychosis. There is a general assumption that the depression

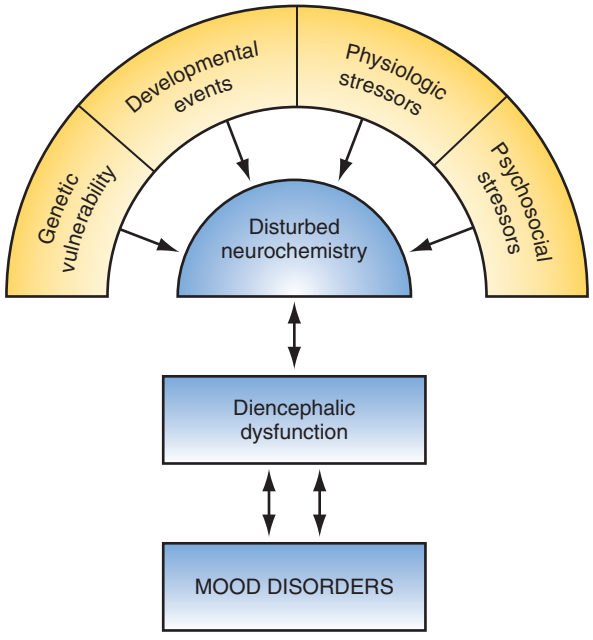


FIG 48.9 Unified model of mood disorders. (Redrawn from Stuart GW, Sundeen SJ: Disturbances of mood. In Stuart GW, Sundeen SJ, editors: *Principles and practice of psychiatric nursing,* ed 4, St Louis, 1991, Elsevier, p 429.)

and mania of BD are the result of two different neurobiological processes. Alternatively, BD is seen as the result of a neurotransmission dysregulation that leads to mood stabilization. Given the fact that it is not uncommon in persons with BD to have a mixed mood state that includes symptoms of both mania and depression, the mood destabilization hypothesis is of great interest. Both models focus on serotonin, norepinephrine, and dopamine as the neurotransmitters involved.

Neuroimaging studies of brain structural changes associated with BD indicate findings potentially similar to those of major depression. For example, ventricular enlargement; increased numbers of T2 signal hyperintensities; and tissue loss in the basal ganglia, lateral and mesial temporal structures, and cortical regions have been observed. At the same time, other researchers have reported no significant tissue changes associated with BD. Experts suggest that tissue changes probably are linked with greater mania symptom severity and duration. For example, symptom onset in older age has been related to greater tissue changes. Neurobiological studies of the evidence of tissue changes associated with BD thus far have not linked specific changes with depression, elation, or mania. Nevertheless, impaired functioning has consistently been associated with tissue changes. Findings such as these lend important support to the hypothesis of impaired emotion processing and impaired regulation of emotional behavior as the probable neurobiological mechanisms of BD.

Clinical Manifestations

Mania is characterized by an abnormally and persistently elevated, expansive, or irritable mood that lasts for 1 week or more (or any duration if hospitalization is required). During the period of mood disturbance, the person also typically experiences three to four of the following symptoms: *inflated self-esteem or grandiosity* (e.g., self-appraised importance, claims of limitless expertise, and insulting derogatory statements). This perceived self-importance can reach delusional proportions, resulting in the person's attempting to act on his or her perceived importance, genius, and infallibility. Serious negative consequences of grandiose actions, as well as financial and criminal schemes are not uncommon. Other manifestations that may become evident during the manic phase of BP include *decreased need for sleep; increased talkativeness or pressure* to keep talking; *flight of ideas or subjective experience that thoughts are racing; distractibility; increased goal-directed activity* (such as social, work or school, sexual) or *psychomotor agitation;* and *excessive involvement in pleasurable activities that have a high potential for painful, long-term consequences* (e.g., unrestrained buying spree, sexual indiscretions). The mood disturbance must be severe enough to cause marked impairment in functioning and/or maintaining relationships with others, or necessitate hospitalization to prevent harm to self or others; in addition, psychotic features in the mood disturbance may be evident.

The *difference between mania and hypomania* can be characterized by a difference in severity and duration. For example, although the symptoms are the same, hypomanic symptoms persist for 4 days or more (instead of 7). Typically, hypomania does not include psychosis or impaired functioning. Instead, the individual experiencing hypomania has a sudden onset of increased energy, expanded self-esteem, and decreased anxiety; these symptoms typically are reported to have improved the affected individual's productivity and are experienced as an acceptable natural high. Unlike mania, it is possible for an episode of hypomania to run its course without being recognized as hypomania.

Hallucinations and delusions can occur with the psychosis that typically develops with severe, prolonged mania. Unlike the hallucinations and delusions associated with schizophrenia, the alterations in perceptions and thinking that occur with mania generally are mood congruent. In other words, grandiose mood is likely to be mirrored by grandiose delusions and hallucinations.

Pharmacologic Treatment

BD is a complex condition. Although the "depression" in BD appears similar to unipolar depression, it is different with regard to pathology, outcome, and management. Caution needs to be taken to avoid manic switches or rapid-cycling induction with antidepressants. Appropriate pharmacologic treatment for BD requires accurate diagnosis. Before BD diagnosis and treatment are considered, the possibility of a primary medical condition or drug reaction as the cause of depression, elation, mania, or psychosis must be considered. Because of the many overlapping symptoms, it is not uncommon for persons with BD to be misdiagnosed as depressed or schizophrenic for many years before a diagnosis of BD is made.

Even though most BD patients present with symptoms of depression and do not volunteer information about mania, treatment guidelines typically focus on the management of hypomania and mania, rather than depression. This focus may exist due to the lack of proven efficacy and substantial controversy regarding the use of antidepressant drugs. In 2007 a large clinical trial revealed no benefit associated with the administration of paroxetine or bupropion to a mood stabilizer; another study reported that paroxetine achieved no better recovery from depression in BD depressive states than a placebo. Though it would be considered premature to declare antidepressants as ineffective in treating BD, a mounting body of evidence supports the ineffectiveness of paroxetine specifically for treating depressive symptoms in BD patients.

Most treatment guidelines since 2005 emphasize using antidepressant medications in combination with antimanic agents. In general, medications presently indicated for use as mood stabilizers include lithium, anticonvulsants, and atypical antipsychotics. For patients not yet in treatment for BD, with severe mania or mixed episodes, lithium (most typically) or valproates are commonly used in combination with an antipsychotic to stabilize mood. Many individuals with severe BD symptoms require additional antidepressant and antipsychotic medications to achieve optimal symptom management. Antipsychotic medications have been found to be effective in remitting symptoms of psychosis as well as in preventing their recurrence. Managing the depression symptoms of BD can prove more complicated. In some cases antidepressant medications can trigger mood switching or destabilize mood; as such, antidepressant monotherapy is not recommended.

Growing numbers of anticonvulsants now are being considered for use as mood stabilizers. These include various reformulations of carbamazepine, divalproex, and lamotrigine. Several treatment guidelines suggest that lamotrigine can also be used as the first-line choice. Many of the pharmacodynamic and pharmacokinetic properties of the commonly used mood-stabilizing medications have been well defined. A variety of potential mechanisms of action have been proposed for lithium. The neurotransmission effects of lithium have been attributed to calcium-dependent cell wall depolarization. Through this process, dopamine and norepinephrine are released, and major secondary messenger neurotransmitter signals are released. Lithium is absorbed in the gastrointestinal tract but is not metabolized; more than 90% is excreted by the renal system. Any interference with the excretion of lithium (e.g., fluid volume depletion, angiotensin-converting enzyme inhibitors, diuretics, and ibuprofen) can lead to rapidly increasing plasma lithium levels and toxicity. Continuous patient education and routine plasma lithium level measurements help to reduce the risk of toxicity. Early symptoms of lithium toxicity include confusion, nausea, and fatigue. Prelithium assessment of liver, renal, and thyroid functioning is required. In some cases, lithium leads to hypothyroidism that requires treatment. Lastly, lithium treatment for BD requires multiple daily doses, generally twice a day or more often. Multiple daily doses can exacerbate patient ambivalence about taking lithium. Euphoria, mania, and long

symptom-free periods typically are misperceived as signs that lithium no longer is needed. There is some evidence that the degree of response to lithium treatment is familial.

Divalproex often is the next best choice when lithium cannot be taken. This anticonvulsant has been found to have neuroprotective effects similar to those of lithium. Unlike lithium, divalproex is highly bioavailable, is metabolized by the liver, produces an active metabolite, has a long half-life, and can be used to manage acute mania. Persons with impaired liver functioning or liver disease cannot take divalproex, and the routine evaluation of plasma levels is required to reduce the risk of toxicity.

Several atypical antipsychotic medications show promise as potentially effective mood-stabilizing medications. Of these, olanzapine, risperidone, quetiapine, and ziprasidone have received considerable attention. These medications exhibit a range of dopaminergic, serotonergic, and norepinephrinergic effects that improve many bipolar symptoms, particularly when the symptoms include full mania rather than hypomania. Once-daily dosing is an important advantage. The two atypical antipsychotics that have emerged as the best treatment options for bipolar depression are either quetiapine monotherapy or quetiapine with an olanzapine–fluoxetine combination; if psychotic features are present, then an antipsychotic such as olanzapine, quetiapine, or risperidone is suggested. It is critical to discontinue medications that no longer seem effective, although most treatment guidelines provide little information about a specific time frame before interrupting or switching medications.

Nonpharmacologic Treatment

Although pharmacotherapy is the first-line treatment choice, supplemental psychotherapy can be extremely helpful in the treatment of BP. Psychoeducational approaches have been shown to reduce relapse rate by educating the patient about the illness and medication, gaining insight into early signs of relapse, and promoting regular sleep–wake cycles. CBT that focuses on treatment adherence and addressing barriers to treatment has been associated with fewer bipolar episodes, reduced hospitalizations, and reduced episode duration. Family-focused treatment has also proven to be effective. It is usually easier to engage patients during their depressive cycles than manic, because the mania itself is associated with lack of insight and motivation.

POPULATION CONSIDERATIONS

Women and Mental Illness

Globally, women are more than twice as likely as men to suffer from depression (including unipolar depression, dysthymia, and BD) and anxiety disorders (including panic disorder, posttraumatic stress disorder, generalized anxiety disorder, social anxiety, and phobias) (Table 48.5). These gender differences begin in adolescence and continue throughout midlife, occurring throughout childbearing years. There is very likely both a biological basis for this gender difference and an inherent vulnerability associated with the conditions of women's lives, including marital status; work and roles in society; and exposure to sexual abuse, assault, and physical violence. Biologically, women and men have different physiologic responses to stress, and the HPA and sympathoadrenomedullary systems are known to play important roles in both depression and anxiety disorders. In addition, serotonin 1A (or 5-HT_{1A}) receptors have been implicated in both depression and anxiety, and there are known gender differences in 5-HT_{1A} receptor and 5-HT binding potentials. Reproductive hormones also play an important role, because they affect HPA responsiveness, glucocorticoid feedback sensitivity, and brain GABA activity and possibly destabilize homeostatic systems in vulnerable women, exacerbating anxiety and depression. Thus, as hormonal levels change throughout the menstrual cycle (e.g., after parturition and during menopause), the balance and homeostasis of serotonin-specific neurotransmitter function are disrupted, and hence affect mood.

Cultural Considerations

There are known differences in how cultures and regions around the world view mental health and, as such, how they express concerns about the body, self, and emotions. There are also aberrant behavioral patterns indigenous to specific cultures that do not easily fit into *DSM* or *ICD* diagnostic categories. As such, aberrant behaviors that are considered an "illness" by the indigenous populations are referred to as *culture-bound syndromes* in DSM-5. DSM-5 provides greater cultural sensitivity throughout the manual by replacing criteria to reflect cross-cultural variations of mental illnesses presentations. Specific diagnostic criteria were altered so that they may apply across diverse cultures, and idioms of distress are explicated from differing cultural perspectives. To be labeled as having a culture-bound syndrome in *DSM-5,* the disorder must be a discrete, well-defined syndrome; recognized as a specific illness in the given culture; expected, recognized, and sanctioned as a response to certain precipitants in the culture; and have a higher incidence or prevalence in societies where the disorder is culturally recognized, compared with other societies. Furthermore, the cultural formulation interview guide is a novel, evidence-based psychiatric tool presented in the DSM-5 that assists clinicians in making person-centered evaluations that are culturally informed, making diagnosis and treatment more effective.

In addition to culture-bound syndromes, it is important for the clinician to consider the cultural identity of the individual and how that identity will influence cultural explanations of mental illness and their personal psychosocial environment. Cultural elements will also come into play in the relationship between the patient and the clinician and in the patient's participation in diagnosis and care (e.g., adherence to psychotropic medication regimen). The clinician must consider all of these factors when formulating a treatment plan to be a culturally competent mental health provider.

Geriatric Considerations

Schizophrenia and delusional disorder may continue into old age or may appear later in life. Late-onset schizophrenia, which emerges after age 45 years, occurs more often in women than in men. Psychiatric admissions of older individuals are characterized by significantly more major depression and less dysthymia than that seen in younger adults. The incidence of BD in the geriatric age group should approach that of lifetime risk. It is extremely rare for BD to emerge after age 60 years, and all BD patients who survive to old age continue to be vulnerable to that illness.

Mood disorders in the elderly are likely to be associated with concomitant illnesses or their treatments.

Management of late-appearing psychoses is complicated by a number of factors. Older patients are more sensitive to medication than younger patients, and their response varies more than that of younger patients; they may suffer cognitive impairment; they may have visual or auditory impairment; they may be taking drugs for other chronic disorders; and they may forget to take their medications or take incorrect doses. In addition, older patients may suffer side effects such as tardive dyskinesia, a disorder related to antipsychotic drug dosage and duration characterized by involuntary chewing motions and darting of the tongue. It is a generally accepted practice that antidepressants should be prescribed at lower dosages and titrated upward more slowly in the elderly (often referred to as "start low and go slow"). Some investigators have argued that this strategy delays response and they suggest that as an alternative different TCAs should be used (such as nortriptyline). There has been little research on the effectiveness of SSRIs in the elderly; ECT has been shown to be effective and safe.

TABLE 48.5 Global Health Considerations: Prevalence of Depression and Anxiety in Women

Country/ Region	Prevalence of Depression and Anxiety	Possible Cultural Factors That Contribute to Depression and Anxiety
Sub-Saharan Africa	18.2% Uganda 30.8% Zimbabwe	Female genital mutilation; abuse of women; PMDD (premenstrual dysphoric disorder) Multiple roles women fulfill in society: wives, mothers, caretakers, employees Continuing impact of HIV/AIDS
East Asia/China	24.2% older Chinese female	Female infanticide High suicide rate in Chinese Stressors of multiple roles (wives, mothers, caretakers, employees) Major social and economic changes
North America	8%–11% perinatal depression 6%–13% postpartum depression 36% all anxiety disorders (U.S.) 20% major depression	High incidence of rape Stressors of multiple roles (wives, mothers, caretakers, employees) Strong inverse relationship between social position and mental health outcomes Large socioeconomic inequalities linked to health inequalities
Western and Central Europe	18.2% Europe 44% Russia 40% Poland 34% Czech	High incidence of rape Stressors of multiple roles (wives, mothers, caretakers, employees) Strong inverse relationship between social position and mental health outcomes Large socioeconomic inequalities linked to health inequalities
South and Southeast Asia	16.2% antenatal depression India 26.3% depression in Asia	Dowry death; domestic violence; other violence against women Burnings: in 1990 87,000 women in India died in fires Stressors of multiple roles (wives, mothers, caretakers, employees) Major social and economic changes
Central and South America	Unclear; probably similar	Domestic violence, other violence against women Stressors of multiple roles (wives, mothers, caretakers, employees) Major social and economic changes
Middle East and North Africa	13%–18% in Middle East and North Africa	Traditional arranged marriage; preference for males Stressors of multiple roles (wives, mothers, caretakers, employees)
Oceania	36% primary care have symptoms of psychological disorders; 20.5% of those have depression and anxiety symptoms	Australia: strong inverse relationship exists between social position and mental health outcomes Social disadvantage and discord, exposure to adverse life events

Data from *Action steps for improving women's mental health,* U.S. Department of Health and Human Services, Office on Women's Health, National Institutes of Health, National Institute of Mental Health, Rockville, MD, 2009. Available at http://store.samhsa.gov/shin/content/OWH09-PROFESSIONAL/OWH09-PROFESSIONAL.pdf. Accessed April 8, 2017. Chandra PS, Herrman H, Fisher J et al: *Contemporary topics in women's mental health: global perspectives in a changing society,* West Sussex, UK, 2009, Wiley-Blackwell. Gomel MK: *A focus on women. Report for the Division of Mental Health and Prevention of Substance Abuse,* World Health Organization, Geneva, 1997. Available at http://whqlibdoc.who.int/hq/1997/WHO_MSA_NAM_97.4.pdf. Accessed April 8, 2017.

KEY POINTS

- Mood disorders are due to disordered affect. Included in this category of psychoses are BD (periods of mania and depression) and MDD. The average age at onset is 30 years for BD and 40 years for MDD. Depression affects women twice as often as men.
- A biochemical basis for BD is supported by observations that levels of brain monoamines (norepinephrine, serotonin) are below normal or the ratio of norepinephrine to serotonin is altered. Depression is thought to occur when serotonin and norepinephrine activity in the brain is low. Mania may be due to a relative excess of norepinephrine in the context of low serotonin or acetylcholine activity. Plasma membrane transport of small molecules such as lithium also is different. BD has a familial pattern of expression, suggesting a genetic cause.
- Psychosocial factors that may affect the development and expression of mood disorders include loss (real, anticipated, or perceived) and low self-esteem. Sleep disorders accompany both extremes of mood. Depression is associated with altered rapid eye movement sleep and decreased slow-wave sleep. Mania is associated with short periods of sleep and reduced fatigue.
- Depression is manifested by low energy, inability to experience joy, difficulty initiating tasks, reduced decision-making ability, difficulty sleeping, poor appetite, weight loss, and decreased libido. Thoughts may focus on guilt, futility, emptiness, hopelessness, helplessness, and suicide.
- The management of MDD is aimed at increasing norepinephrine and serotonin activity in the brain. MAOIs reduce the rate of neurotransmitter destruction; TCAs inhibit reuptake; newer agents (fluoxetine) selectively prevent serotonin reuptake.
- Mania is manifested by high energy; inflated self-esteem; hyperactivity; inability to focus or concentrate; low sensitivity to fatigue, injury, or pain; rapid or incoherent speech; hallucinations; delusions; increased appetite and libido; decreased sleep; poor judgment; and poor impulse control. Mania is managed with lithium, a compound that inhibits the action of norepinephrine and serotonin in the brain.
- In general, BD is managed with mood stabilizers, including lithium, anticonvulsants, and atypical antipsychotics. Anticonvulsants as mood stabilizers are aimed at increasing norepinephrine and serotonin activity in the brain. MAOIs reduce the rate of neurotransmitter destruction; TCAs inhibit reuptake; newer agents (fluoxetine) selectively prevent serotonin reuptake. Depression, a common symptom in BD, is managed by using antidepressant medications only in combination with antimanic agents.

SUMMARY

The clinical symptoms, causes, and neurobiological mechanisms of schizophrenia, MDD, and BD have been presented. Significant risk of psychosis is a common characteristic shared among these seemingly unrelated disorders. Brief overviews of the neurobiological basis of medications used to manage the symptoms of these disorders reflect the tremendous neurobiological advances that have been accomplished. Nevertheless, these three illnesses are potentially disabling conditions that all too often rob the affected person of academic goals, meaningful work, close relationships, self-actualization, and, in some circumstances, survival.

RESOURCES

Epidemiology and Classification

American Psychiatric Association: *DSM-5 development,* April 23, 2012. Available at www.dsm5.org/proposedrevision/pages/default.aspx.:Accessed.

American Psychiatric Association: *Diagnostic and statistical manual of mental disorders (DSM-IV-TR)*, ed 4, text revision, Washington, DC, 2000, The American Psychiatric Association.

International Statistical Classification of Diseases and Related Health Problems: *10th revision (ICD-10) Version for 2010: Mental and behavioural disorders*, Available at http://apps.who.int/classifications/icd10/browse/2010/en#/v. (Accessed 27 April 2012).

National Council on Behavioral Health: *Coding Behavioral Health Services*, Available at http://www.thenationalcouncil.org/topics/coding-behavioral-health-services/. (Accessed 22 October 2015).

National Institutes of Mental Health: *Serious mental illness in USA adults*, Available at www.nimh.nih.gov/statistics/index.shtml. (Accessed 25 April 2012).

United States Department of Health and Human Services: *Substance Abuse and Mental Health Services Administration*, Center for Behavioral Health Statistics and Quality. National Survey on Drug Use and Health, 2013. ICPSR35509-v3. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2015-11-23. http://doi.org/10.3886/ICPSR35509.v3.

Schizophrenia

Aquila R: Management of weight gain in patients with schizophrenia. *J Clin Psychiatry* 63(Suppl 4):33–36, 2002.

Bresee LC, Majumdar SR, Patten SB, et al: Prevalence of cardiovascular risk factors and disease in people with schizophrenia: a population-based study. *Schizophr Res* 117(1):75–82, 2010.

Caspi A, Moffitt TE, Cannon M, et al: Moderation of the effect of adolescent-onset cannabis use on adult psychosis by a functional polymorphism in the catechol-O-methyltransferase gene: longitudinal evidence of a gene X environment interaction. *Biol Psychiat* 57(10): 1117–1127, 2015.

Diagnostic and Statistical Manual of Mental Disorders. Highlights of changes from DSM-IV-TR to DSM-5, *American Psychiatric Association*, Retrieved from: http://www.dsm5.org/Documents/changes%20from%20dsm-iv-tr%20to%20dsm-5.pdf.

Di Forte M, et al: Confirmation that the AKT1 (rs2492732) genotype influences the risk of psychosis in cannabis users, *Biol Psychiat. Biol Psychiatry* 72(10):811–816, 2012.

Dixon L, Weiden P, Delahanty J, et al: Prevalence and correlates of diabetes in national schizophrenia samples. *Schizophr Bull* 26(4):903–912, 2000.

Evans JD, Heaton RK, Paulsen JS, et al: The relationship of neuropsychological abilities to specific domains of functional capacity in older schizophrenia patients. *Biol Psychiatry* 53(5):422–430, 2003.

Fadem B: *High-yield behavioral science*, ed 2, Philadelphia, 2001, Lippincott Williams & Wilkins.

Frankle WG, Lerma J, Laruelle M: The synaptic hypothesis of schizophrenia. *Neuron* 39(2):205–216, 2003.

Goggi J, Pullar IA, Carney SL, et al: Signaling pathways involved in the short-term potentiation of dopamine release by BDNF. *Brain Res* 968(1): 156–161, 2003.

Harrop C, Thrower P: *Why does schizophrenia develop at late adolescence? A cognitive-developmental approach to psychosis*, West Sussex, England, 2003, Wiley.

Holden C: Neuroscience: deconstructing schizophrenia. *Science* 299(5605): 333–335, 2003.

Jampala VC, Taylor MA, Abrams R: The diagnostic implications of formal thought disorder in mania and schizophrenia: a reassessment. *Am J Psychiatry* 146(4):459–463, 1989.

Javitt DC, Coyle JT: Decoding schizophrenia. *Sci Am* 290(1):48–55, 2004.

Jessen F, Scheef L, Germeshausen L, et al: Reduced hippocampal activation during encoding and recognition of words in schizophrenia patients. *Am J Psychiatry* 160(7):1305–1312, 2003.

Kalkman HO, Loetscher E: GAD(67): the link between the GABA-deficit hypothesis and the dopaminergic and glutamatergic theories of psychosis. *J Neural Transmission* 110(7):803–812, 2003.

Kapur S: Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. *Am J Psychiatry* 160(1):13–23, 2003.

Kapur S, Seeman P: Does fast dissociation from the dopamine D(2) receptor explain the action of atypical antipsychotics? A new hypothesis. *Am J Psychiatry* 158(3):360–369, 2001.

Keltner NL: Neuroreceptor function and psychopharmacologic response. *Issues Ment Health Nurs* 21(1):31–50, 2000.

Keltner NL, Folks DG: *Psychotropic drugs*, ed 4, St Louis, 2005, Elsevier.

Magueresse CL, Monyer H: GABAergic interneurons shape the functional maturation of the cortex. *Neuron* 77:388–405, 2013.

Milev P, Ho BC, Amdt S, et al: Initial magnetic resonance imaging volumetric brain measurements and outcome in schizophrenia: a prospective longitudinal study with 5-year follow-up. *Biol Psychiatry* 54(6):608–615, 2003.

Minzenberg MJ, Yoon JH, Carter CS: Schizophrenia. In Hales RE, Yudofsky SC, Gabbard GO, editors: *Textbook of psychiatry*, ed 5, Arlington, VA, 2008, American Psychiatric Publishing, pp 407–456.

Nakazawa K, Zsiros V, Jiang Z, et al: GABAergic interneuron origin of schizophrenia pathophysiology. *Neuropharm* 62:1574–1583, 2012.

Nasrallah HA: Atypical antipsychotic-induced metabolic side effects: insights from receptor-binding profiles. *Mol Psychiatry* 13(1):27–35, 2007.

Nasrallah HA, Smeltzer DJ: *Contemporary diagnosis and management of the patient with schizophrenia*, Newtown, PA, 2002, Handbooks in Health Care.

Portin P, Alanen YO: A critical review of genetic studies of schizophrenia: II. Molecular genetic studies. *Acta Psychiatr Scand* 95(2):73–80, 1997.

Rinomhota AS, Marshall P: *Biological aspects of mental health nursing*, Edinburgh, 2000, Churchill Livingstone.

Saal D, Dong Y, Bonci A, et al: Drugs of abuse and stress trigger a common synaptic adaptation in dopamine neurons. *Neuron* 37(4):577–582, 2003.

Saha S, Chant D, McGrath J: A systematic review of mortality in schizophrenia—is the differential mortality gap worsening over time? *Arch Gen Psychiatry* 64(10):1123–1131, 2007.

Schmitz Y, et al: Presynaptic regulation of dopaminergic neurotransmission. *J Neurochem* 87(2):273–289, 2003.

Schoepf D, Potluri R, Uppal H, et al: Type-2 diabetes mellitus in schizophrenia: increased prevalence and major risk factor of excess mortality in a naturalistic 7-year follow-up. *Eur Psychiatry* 27(1):33–42, 2012.

Smith GN, Lang DJ, Kopala LC, et al: Developmental abnormalities of the hippocampus in first-episode schizophrenia. *Biol Psychiatry* 53(7): 555–561, 2003.

Susser E, Lin SP, Brown AS, et al: No relation between risk of schizophrenia and prenatal exposure to influenza in Holland. *Am J Psychiatry* 151(6): 922–924, 1994.

Tandon R: Antipsychotics in the treatment of schizophrenia: an overview. *J Clin Psychiatry (Suppl)* 72:4–8, 2011.

Tandon R, Nasrallah HA, Keshavan MS: Schizophrenia, "just the facts" 4. Clinical features and conceptualization. *Schizophrenia Res* 110(1-3):1–23, 2009.

Thara R, Padmavati R: Psychotic disorders and bipolar affective disorders. In Chandra PS, Herrman H, Fisher J, et al, editors: *Contemporary topics in women's mental health: global perspectives in a changing society*, West Sussex, UK, 2009, Wiley-Blackwell, pp 9–36.

Venables PH, et al: Prenatal influenza exposure and delivery complications: implications for the development of schizophrenia. *Fam Community Health* 30(2):151–159, 2007.

Volkow ND, Baler RD, Compton WM, Weiss SRB: Adverse health effects of marijuana use. *New Engl J Med* 370:2219–2227, 2014.

Depression

Beck AT, Rush AJ, Shaw BF, et al: *Cognitive therapy of depression*, New York, 1979, Guilford.

Brent DA, Birmaher B: Clinical practice: adolescent depression. *N Engl J Med* 347(9):667–671, 2002.

Caspi A, Sugden K, Moffitt TE, et al: Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science* 301(5631): 386–389, 2003.

Drevets WC: Functional neuroimaging studies of depression: the anatomy of melancholia. *Annu Rev Med* 49:341–361, 1989.

Driessen E, Hollon SD: Cognitive behavioral therapy for mood disorders: efficacy, moderators and mediators. *Psychiatr Clin North Am* 33(3): 537–555, 2010.

Frasure-Smith N, Lespérance F, Gravel G, et al: Social support, depression, and mortality during the first year after myocardial infarction. *Circulation* 101(16):1919–1924, 2000.

Gelenberg A, Freeman M, Markowitz J, et al: *Practice guideline for the treatment of patients with major depression*, ed 3, Toronto, Ontario, Canada, 2010, Contact Resource Center.

González HM, Vega WA, Williams DR, et al: Depression care in the United States: too little for too few. *Arch Gen Psychiatry* 67(1):37–46, 2010.

Guro Å, Hammar Å: Is impairment in cognitive inhibition in the acute phase of major depression irreversible? Results from a 10-year follow-up study. *Psychol Psychother Theory Res Pract* 84(2):141–150, 2011.

Halaris A: A primary care focus on the diagnosis and treatment of major depressive disorder in adults. *J Psychiatr Pract* 17(5):340–350, 2011.

Insel TR, Charney DS: Research on major depression: strategies and priorities. *JAMA* 289(23):3167–3168, 2003.

Jacobs BL, van Praag H, Gage FH: Depression and the birth and death of brain cells. *Am Sci Online* 88(4):340–347, 2000.

Joska JA, Stein DJ: Mood disorders. In Hales RE, Yudofsky SC, Gabbard GO, editors: *Textbook of psychiatry*, ed 5, Arlington, VA, 2008, American Psychiatric Publishing, pp 457–503.

Kadri N, Alami K: Depression and anxiety among women. In Chandra P, Herrman H, Fisher J, et al, editors: *Contemporary topics in women's mental health: global perspectives in a changing society*, Hoboken, NJ, 2009, Wiley-Blackwell, pp 37–64.

Kessler RC, Berglund P, Demler O, et al: The epidemiology of major depressive disorder—results from the National Comorbidity Survey Replication (NCS-R). *JAMA* 289(23):3095–3105, 2003.

Koolschijn MP, van Haren Neeltje EM, Gerty JLM, et al: Brain volume abnormalities in major depressive disorder: a meta-analysis of magnetic resonance imaging studies. *Hum Brain Mapp* 30:3719–3735, 2009.

Manji HK, et al: Enhancing neuronal plasticity and cellular resilience to develop novel improved therapeutics for difficult-to-treat depression. *Biol Psychiatry* 53(8):707–742, 2003.

McGuffin P, Rijsdijk S, Andrew M, et al: The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. *Arch Gen Psychiatry* 60:497–502, 2003.

National Association for the Mentally Ill: *Major depression fact sheet*, Arlington, VA. Available at www.nami.org/Template.cfm?Section=Depression&Template=/ContentManagement/ContentDisplay.cfm&ContentID=88956. (Accessed 26 April 2012).

Pearson G: Psychopharmacology. In Mohr WK, editor: *Psychiatric-mental health nursing*, ed 6, Philadelphia, 2006, Lippincott Williams & Wilkins.

Phillips ML, Drevets WC, Rauch SL, et al: Neurobiology of emotion perception II: implications for major psychiatric disorders. *Soc Biol Psychiatry* 54(5):515–528, 2003.

Phillips WJ, Hine DW, Thorsteinsson EB: Effortful information processing in patients with major depression—a 10-year follow-up study. *Clin Psychol Rev* 30(6):2010.

Posener JA, DeBattista C, Williams GH, et al: 24-hour monitoring of cortisol and corticotropin secretion in psychotic and nonpsychotic major depression. *Arch Gen Psychiatry* 57(8):755–760, 2000.

Racagni G, Brunell N: Physiology to functionality: the brain and neurotransmitter activity. *Int Clin Psychopharmacol* 14(Suppl 1):S3–S7, 1999.

Sheline YI: Neuroimaging studies of mood disorder effects on the brain. *Biol Psychiatry* 54(3):338–352, 2003.

Shimizu E, Hashimoto K, Okamura N, et al: Alterations of serum levels of brain derived neurotrophic factor (BDNF) in depressed patients with or without antidepressants. *Biol Psychiatry* 54(1):70–75, 2002.

Stockmeir CA, Shi X, Konick L, et al: Neurokinin-1 receptors are decreased in major depressive disorder. *Neuroreport* 13(9):1223–1227, 2002.

Sullivan P, Kendler K: Genetic case-control studies in neuropsychiatry. *Arch Gen Psychiatry* 58:1015–1024, 2001.

Verhagen M, van der Meij A, Franke B, et al: Familiality of major depressive disorder and gender differences in comorbidity. *Acta Psychiatr Scand* 118(2):130–138, 2008.

Young E, Korszun A: Psychoneuroendocrinology of depression: hypothalamic-pituitary-gonadal axis. *Psychiatr Clin North Am* 21(2): 309–323, 1998.

Zubenko GS, Hughes HB, 3rd, Stiffler JS, et al: D2S2944 identifies a likely susceptibility locus for recurrent, early-onset, major depression in women. *Mol Psychiatry* 7(5):460–467, 2002.

Bipolar Disorder

Akiskal HS: The bipolar spectrum in psychiatric and general medical practice. *Prim Psychiatry* 11(9):30–35, 2004.

Biederman J, Faraone SV, Wozniak J: Further evidence of unique developmental phenotypic correlates of pediatric bipolar disorder: findings from a large sample of clinically referred preadolescent children assessed over the last 7 years. *J Affect Disord* 82:S45–S58, 2004.

Blader JC, Carlson GA: Increased rates of bipolar disorder diagnoses among U.S. child, adolescent, and adult inpatients, 1996-2004. *Biol Psychiatry* 62(2):107–114, 2007.

Cruceanu C, Alda M, Rouleau G, et al: Response to treatment in bipolar disorder. *Curr Opin Psychiatry* 24(1):24–28, 2011.

Fountoulakis KN: The contemporary face of bipolar illness: complex diagnostic and therapeutic challenges. *CNS Spectr* 13(9):763–774, 2008.

Frokjaer VG, et al: Familial risk for mood disorder and the personality risk factor, neuroticism, interact in their association with frontolimbic serotonin 2A receptor binding. *Neuropsychopharm* 35:1129–1137, 2010.

Geddes JR, Miklowitz DJ: Treatment of bipolar disorder. *Lancet* 381(9878):1672–1682, 2013. doi:10.1016/SO140-6736(13)60857-0.

Hajek T, Gunde E, Slaney C: Striatal volumes in affected and unaffected relatives of bipolar patients—high-risk study. *J Psychiatric Res* 43(7): 724–729, 2009.

Judd LL, Schettler PJ, Akiskal HS, et al: Residual symptom recovery from major affective episodes in bipolar disorders and rapid episode relapse/recurrence. *Arch Gen Psychiatry* 65(4):386–394, 2008.

Keck PE, McElroy SL: Clinical pharmacodynamics and pharmacokinetics of antimanic and mood-stabilizing medications. *J Clin Psychiatry* 63(Suppl 4):3–11, 2002.

Lam DH, Watkins ER, Hayward P, et al: A randomized controlled study of cognitive therapy for relapse prevention for bipolar affective disorder: outcome of the first year. *Arch Gen Psychiatry* 60:145–152, 2003.

Milkowitz DJ, Hooley JM: Developing family psychoeducational treatments for patients with bipolar and other severe psychiatric disorders: a pathway from basic research to clinical trials. *Marital Fam Ther* 24:419–435, 1998.

Moreno C, Laje G, Blanco C, et al: National trends in the outpatient diagnosis and treatment of bipolar disorder in youth. *Arch Gen Psychiatry* 64(9): 1032–1039, 2007.

Nivoli AMA, Colom F, Murru A, et al: New treatment guidelines for acute bipolar depression: a systematic review. *J Affect Disord* 129(1-3):14–26, 2011.

Stahl SM: *Essential psychopharmacology: neuroscientific basis and practical applications*, ed 2, Cambridge, UK, 2000, Cambridge University Press.

Tugrul K: The nurse's role in the assessment and treatment of bipolar disorder. *J Am Psychiatr Nurs Assoc* 9(6):180–186, 2003.

49

Neurobiology of Nonpsychotic Illnesses

Ann Futterman Collier and Samantha Cody Russell

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- What neurobiological alterations have been associated with panic disorder, generalized anxiety disorder, obsessive-compulsive disorder, and posttraumatic stress disorder?
- What neurobiological alterations have been associated with attention-deficit/hyperactivity disorder and autism spectrum disorder?

CHAPTER OUTLINE

Anxiety Disorders, 989
Panic Disorder, 990
Etiology and Neurobiology, 990
Clinical Manifestations, 991
Treatment, 991
Generalized Anxiety Disorder, 991
Etiology and Neurobiology, 991
Clinical Manifestations, 992
Pharmacologic Treatment, 992
Nonpharmacologic Treatment, 992
Obsessive-Compulsive Disorder, 992
Etiology and Neurobiology, 992
Clinical Manifestations, 993
Treatment, 993
Posttraumatic Stress Disorder, 994
Etiology and Neurobiology, 994
Clinical Manifestations, 994
Pharmacologic Treatment, 995
Nonpharmacologic Treatment, 995
Neurodevelopmental Disorders, 995
Attention-Deficit/Hyperactivity Disorder, 996
Etiology and Neurobiology, 996
Clinical Manifestations, 996
Pharmacologic Treatment, 996
Nonpharmacologic Treatment, 997
Autism Spectrum Disorder, 997
Etiology and Neurobiology, 997
Clinical Manifestations, 997
Treatment, 997

The neurobiological mechanisms of mental disorders that do not cause psychosis may prove to be more similar than different from those of mental disorders associated with psychosis. These conditions already show similarity in that they are categorized by altered neuronal structures and functioning, genetic risk factors, and neurotransmission dysregulation. A major difference between psychotic and nonpsychotic illnesses is that individual variations in nonpsychotic conditions can be extensive. The greater individual variations increase the difficulty of defining the hallmark symptomatology and neurobiological basis of the condition. In addition, much has been learned about the neurobiological impact of stress response systems in psychotic illnesses, whereas the impact of stress response systems in nonpsychotic illnesses is less well understood. In the past, simple contrasts, such as comparing schizophrenia with eating disorders, invited the false assumption that one disorder may be more or less serious than another. For the affected person, such comparisons are unhelpful. Mental disorders, by definition, can cause profound suffering and impairment.

ANXIETY DISORDERS

This section presents four anxiety disorders: panic disorder (PD), generalized anxiety disorder (GAD), obsessive-compulsive disorder (OCD), and posttraumatic stress disorder (PTSD). Although anxiety disorders have many similar physical symptoms, they differ greatly in terms of symptom onset triggers, symptom duration, and symptom management. Anxiety disorders are primarily characterized by physical symptoms, so physical illnesses (e.g., hyperthyroidism) and medication reactions (e.g., antidepressants, steroids, anticholinergic medications) must be ruled out before a diagnosis of anxiety disorder can be made. Together with depression, anxiety disorders are the most common of

all psychiatric illnesses, and aside from social phobia and OCD, occur two times more often in women than in men.

Panic Disorder

In the *Diagnostic and Statistical Manual of Mental Disorders,* Fifth Edition (DSM-V), PD is characterized by the presence of recurrent, unexpected panic attacks (PAs) followed by at least 1 month of persistent concern about having another PA, worry about the possible implications or consequences of the PAs, or a significant behavioral change related to the attacks. PAs themselves are discrete periods of intense fear or discomfort, in which four (or more) of the following symptoms develop abruptly and reach a peak within 10 minutes: palpitations, pounding heart, or accelerated heart rate; sweating; trembling or shaking; sensations of shortness of breath or smothering; feeling of choking; chest pain or discomfort; nausea or abdominal distress; feeling dizzy, unsteady, lightheaded or faint; derealization or depersonalization; fear of losing control or going crazy; fear of dying; paresthesias; and chills or hot flashes. PAs are acute episodes of anxiety symptoms that are unexpected, sudden, and recurrent and generate intense feelings of fear. Sudden symptom onset can cause affected persons to seek emergency health care for what they believe is a cardiac arrest, respiratory arrest, or "nervous breakdown." PAs can be situation bound, and hence occur in other anxiety disorders such as agoraphobia, specific phobias, social phobia, OCD, and PTSD. It is important to determine that the PAs are not due to the physiologic effects of a substance or other general medical condition, or are accounted for better by another mental disorder.

Depending on the subtype, PD is diagnosed two to three times more often in women than in men. Rates of 1% to 2% are most commonly reported. Initial onset usually occurs in late adolescence or young adulthood, with a mean age of onset of 26.6 years. Initial symptom onset in older adults is less typical. Non-Hispanic whites are the most likely to develop PD compared with African Americans and black Caribbeans, although there are higher levels of functional impairment in Caribbean blacks when PD develops.

Etiology and Neurobiology

The risk of anxiety disorder symptom onset has been associated with moderate genetic, psychological, and biological system alterations. Family, twin, and adoptive family studies have consistently shown a strong genetic liability for these disorders. However, experts now speculate that the etiology question is no longer nature versus nurture. Current models seek to explain the *interactions* of nature and nurture that can create susceptibility to anxiety. Brain regions that underpin the experiences of fear, anxiety, and stress are thought of as circuits that can be shaped and altered by a wide range of forces. Neurobiological conditioning is one force shown to affect such circuits and thus is of particular importance to understanding the development of anxiety disorders.

Brain serotonin activity has been revealed to interact with both genes and environment, and these interactions contribute to what has come to be referred to as *synaptic plasticity.* In this way, gene activity is linked with brain cell neurochemistry and psychological characteristics such as temperament. More specifically, evidence of genetic variability in negative emotions, such as anxiety, has been found in studies of serotonin transporter genes. In this research, gene variability is in gene allele length. For example, family studies have shown that siblings with serotonin transporter genes with the short-form gene allele had higher neuroticism scores than their siblings with the long-form gene allele. Research aimed at gene typing mental disorders clearly is in its infancy, and findings such as these are inconclusive, but they demonstrate the possibility of genetics-based neurobiological models of anxiety. Additional genetic risk factors may involve brain-derived neurotrophic factor and catechol-O-methyltransferase (COMT), which could be associated with increased sensitivity to anxiety, especially in females. In addition, COMT rs4680 may affect white matter connectivity in panic disorder.

Susceptible persons who breathe air with high levels of carbon dioxide will experience an acute onset of panic anxiety symptoms. A small study of persons with panic anxiety disorder, using an infusion of doxapram (respiratory stimulant) to cause profound hyperventilation examined the effectiveness of cognitive interventions to reduce respiratory anxiety symptoms. Cognitive interventions were designed to minimize misinterpretation of drug-induced hyperventilation as a sign of danger and thereby reduce the odds of the respiratory stimulant triggering panic anxiety. Breath-by-breath analyses of the patients and healthy controls were performed. The researchers hypothesized that if the respiration anxiety symptoms were the result of dysregulation within the brain respiratory control center, cognitive interventions would not be particularly effective. They found that less fearful thinking did reduce panic, but some respiratory anxiety symptoms persisted despite less fearful thinking. In other words, respiratory anxiety symptoms appeared to result both from anticipatory anxiety and from dysregulation within the brain respiratory center.

The surges of physiologic activation and physiologic instability are thought to be the hallmark neurobiological processes underlying panic anxiety disorder. Multiple organ systems, including the cardiovascular and respiratory systems, are thought to be involved. Observations such as these help to explain why, for example, caffeine triggers panic anxiety symptoms in susceptible individuals. Physiologic instability may prove to be the key to identifying a genetics-based marker for susceptibility to panic. However, the biopsychological marker for the disorder is likely to be the overinterpretation of physical anxiety symptoms (e.g., sudden increase in heart rate) as life threatening. This thinking, referred to as *learned panic,* is thought to result from inordinately high levels of life stress in early childhood.

The presence of overwhelming life stress can increase the level of circulating glucocorticoids (stress hormones) and stimulate the release of glutamate (which inhibits neurogenesis). Early childhood life stress, specifically abuse and neglect, has been studied as a possible predictor of various adult-onset anxiety disorders. These models are based on altered serotonin, norepinephrine, and dopamine neurotransmission; glutamate release; and physiologic instability. Repeated and prolonged childhood exposure to overwhelming stress is thought to create adult susceptibility to anxiety disorders. The leading theory is that early life stress leads to an *overspecialized* or excessive stress response.

Early life stress is thought to produce adult susceptibility to anxiety by altering critical neuron structures and functioning during this critical stage of human growth and development. Brain regions most vulnerable to alteration as a result of early life stress include the hippocampus (glucocorticoid receptors), amygdala (γ-aminobutyric acid [GABA] and benzodiazepine receptors), corpus callosum (glial cells critical to myelination), cerebellar vermis (glucocorticoid receptors), and the prefrontal cortex (glucocorticoid receptors, dopamine projections, and inhibition of hypothalamic–pituitary–adrenal [HPA] axis activation). When the developing brain of a young child is exposed to overwhelming life stress, the stress response system appears to adapt by overbuilding or building additional brain stress response pathways. Later, under less stressful adult circumstances, this overbuilt stress response system becomes maladaptive. Like a very large overpowered car on a small, winding road, the overbuilt stress response system could become the source of physiologic instability that has come to be associated with panic anxiety.

Newer models of cognitive vulnerability suggest that children who go on to develop anxiety disorders have parents who model the dangerousness of anxiety symptoms as well as the need to escape from these symptoms. In addition, their parents reinforce the sick role in children when they experience arousal reactive symptoms. As these

children develop increasing anxiety about bodily symptoms, they also experience a sense of low perceived control over their anxiety. These factors, taken together, likely create cognitive vulnerability for developing PD as well as other anxiety disorders.

Clinical Manifestations

With PD, no single experience consistently triggers symptom onset, although the first attack(s) frequently occur(s) during a life-threatening illness or accident, loss of a close interpersonal relationship, or separation from family. After that, they can occur during any routine activity, from reading to driving. When patients experience their first few PAs they frequently think they are either having a heart attack or losing their mind, and it is common to seek emergency medical treatment. *Physical* symptoms are most prominent and emphasized and include respiratory distress, heart palpitations, tachycardia, pounding heart, chest pain, smothering or choking sensation, dizziness, lightheadedness, faintness, sweating, trembling, shaking, hot flashes, chills, numbness, tingling, nausea, abdominal distress, and urinary frequency. *Psychological* and *cognitive* symptoms include expressed fears of dying, fear of cardiac arrest, fear of losing control, fear of nervous breakdown, derealization, depersonalization, and perceptual distortions. *Behavioral* symptoms include hyperkinesis, pressured speech, and exaggerated startle response. The attacks usually last between 5 and 20 minutes, although they can last as long as an hour. They can happen in a wavelike manner, so that they occur successively, or as described earlier, as part of another clinical disorder. Some people experience such severe anticipatory anxiety that it is hard to separate when the attack starts and ends, so that the PA is experienced as continuous.

PD is characterized by two important psychological symptoms: anticipatory anxiety and avoidance anxiety. *Anticipatory anxiety* refers to fearful expectation of panic anxiety onset. People with the disorder tend to develop a morbid dread of events or experiences that they come to believe *might* trigger panic anxiety. *Avoidance anxiety* refers to personal strategies used to increase feelings of control and thereby decrease the risk of panic anxiety. Persons with PD strive to avoid situations and circumstances they associate with the development of their symptoms. This helps to explain why PD and agoraphobia (the phobic avoidance of public spaces beyond personal control, e.g., airports and shopping malls) often coexist.

Treatment

PD can be effectively managed with cognitive-behavioral therapy (CBT) aimed at reducing fearful thinking and desensitization of cognitive and physical stress responses. Aerobic exercise may also facilitate the effectiveness of CBT. When panic symptoms are disabling, medication for symptom management is recommended. Long-acting benzodiazepines, such as clonazepam, are the sedatives of choice when short-term calming and symptom relief are mandatory. Tolerance to benzodiazepines develops with continuous use regardless of dosage. Misuse of benzodiazepines represents a serious health hazard. Although these drugs are useful in blocking the PA, they do not always decrease the anticipatory anxiety and avoidance, especially when drug regimens are initially undertaken. Many atypical psychiatric medications that can target serotonin, dopamine, or norepinephrine receptors have been clinically tested and shown to be effective treatment for anxiety symptoms. Examples of such medications found to be helpful include paroxetine, sertraline, citalopram, and fluoxetine. Unless contraindicated, β-blocker medications that dampen physical anxiety symptoms may also be helpful.

Generalized Anxiety Disorder

GAD is characterized by chronic and persistent worry, as well as physical anxiety symptoms. The anxiety is excessive, pervasive, difficult to control, and associated with marked distress or impairment. Typically, the patient experiences multiple anxiety symptoms, including restlessness, fatigue, impaired concentration, irritability, muscle tension, muscle pain, and disturbed sleep. Lacking clear symptom onset patterns, GAD is easily overlooked or misdiagnosed. GAD is a chronic condition, and earlier onset is associated with greater overall impairment. There is high comorbidity with substance use and other anxiety disorders, depression, and personality disorders. Higher rates of GAD occur in women than men. The prevalence of GAD does not decline with age and appears to account for most of the anxiety disorders in the elderly.

Etiology and Neurobiology

GAD differs from other anxiety disorders in that the cognitive, psychological, and behavioral symptoms are relatively constant. Psychoanalysts developed most of the original etiologic theories concerning persistent worry. What now is referred to as *persistent worry* was then described as *anxious expectation.* Even at that early stage of discovery, it was apparent that generalized anxiety rarely occurred without comorbid conditions such as depression. This *psychodynamic* view of generalized anxiety prevailed until the late 1980s and early 1990s. More recently, *cognitive theorists* suggest that GAD may originate from difficulty with early attachment to the primary caretaker. Conceptually, worry is seen as an avoidance strategy for negative affect; as a distraction from realistic and proximal threats that need immediate solutions; and as a coping method to "prevent" the feared outcome, such as occurs with magical thinking. Unlike PD, where the worry is more typically about physical catastrophes, in GAD, worries are more about interpersonal confrontation, competence, and acceptance. Technologically advanced research methods have now made it possible to precisely define GAD symptoms in terms of their actual qualities, intensity, and duration, but physical GAD symptoms continue to be viewed as *somatic* expressions of psychological problems.

Unlike other anxiety disorders, GAD onset typically is gradual with symptom duration measured in years. As has been observed with other anxiety disorders, vulnerability to GAD likely is inherited. Twin and family study findings indicate a 30% increase in risk of GAD among the relatives of persons with the disorder. Efforts to describe the neurobiological basis of GAD will no doubt be greatly advanced by theoretical models of inherited vulnerability as well as improved understanding of GAD symptoms. The most important unanswered question likely will have to do with the fact that the alterations in brain structure and functioning shown to be associated with GAD also are consistent with alterations observed with other disorders.

Preliminary positron emission tomography (PET) studies measuring brain glucose metabolism rates in GAD have shown higher-than-normal rates in patients at rest. With GAD, apparently some brain regions undergo both increases and decreases in glucose metabolism rates. This mixed response is most apparent in the frontal and cingulate areas of the cortex, the brain region associated with worry and hypervigilance. Findings such as these lend support to the basic GAD explanatory hypothesis of anxiety symptoms as manifestations of hyperactive brain circuits. Just the opposite condition (hypoactive brain circuits) is thought to be the fundamental basis of depressive disorders.

Neurotransmitter findings with GAD, as with other disorders that produce mood, thinking, and behavior symptoms, point to alterations in GABA receptors, benzodiazepine receptors, norepinephrine systems, serotonin systems, HPA axis activation, and plasma cortisol levels. Thus far, no specific alterations that can consistently explain GAD symptoms have been identified. Nevertheless, one interesting observation shows considerable promise. At rest, no obvious alterations in neurotransmission are noted in GAD patients. Only when subjected to laboratory activities designed to induce stress responses is significant

neurotransmitter overactivity observed. The evidence of norepinephrine receptor down-regulation lends additional support to this stress response model. Attempting to modulate overactive responses, receptor down-regulation is thought to occur automatically when subjected to prolonged, recurrent, excessive, or hyperactive neurotransmitter activity.

Whereas norepinephrine activity is thought to be associated with physical symptoms of GAD, anticipatory and avoidance symptoms are thought to be associated with activity along a key serotonin pathway linking the amygdala and frontal cortex. As might be expected, insufficient serotonin activity in specific brain regions is thought to be associated with GAD. Given the obvious symptom overlap between stress and anxiety, hyperactivity within the HPA axis and high plasma cortisol levels continue to be leading models in GAD research. Much of the difficulty in defining the neurobiological basis of GAD has to do with significant individual variations in GAD symptomatology. Uncontrollable worry (frontal cortex) is the only symptom likely to show meaningful consistency over time and from individual to individual. A second major research difficulty has to do with the frequency with which GAD symptoms co-occur with depression symptoms—so much so that some experts now view mixed depression-anxiety as a specific disorder.

Clinical Manifestations

The symptoms of GAD typically fall into two categories: apprehensive expectation and worry, and physical symptoms. The *worry* is often about minor issues, where the person anticipates the worst possible outcome and finds it difficult to control. No areas of life are excluded, and worry is not limited to any single area of concern (e.g., children). *Physical* symptoms vary greatly, but people often feel "keyed up," which results in muscle tension, lightheadedness, sweating, palpitations, dizziness, and stomach distress. Concentration is typically severely impaired, and irritability is common. Diffuse anticipatory anxiety, avoidance anxiety, and dysphoria are common. Behavioral symptoms include severe sleep disturbance and fatigue. Although much of the behavior associated with GAD is likely to be the result of maladaptive methods of coping with physical and psychological GAD symptoms, impaired social, academic, and employment functioning are common.

Pharmacologic Treatment

Effective psychological and drug treatments for GAD can be relatively complex. When alcohol is comorbid, it may limit the effectiveness of treatment and/or delay the onset of benefits. The delayed onset of benefits from treatment is likely to be seen when GAD symptoms are long-standing and highly disabling. Lastly, neurobiological research findings indicate that effective drug treatment is likely to require one or more medications that are reliable modulators of multiple neurotransmission systems across multiple brain regions. Medications shown to relieve and in some cases remit GAD symptoms include long-acting benzodiazepines (e.g., clonazepam), partial serotonin (i.e., 5-hydroxytryptamine, 5-HT_{1A}) agonists (e.g., buspirone), tricyclic antidepressants (TCAs; e.g., imipramine), selective serotonin reuptake inhibitors (SSRIs; e.g., sertraline), serotonin-norepinephrine reuptake inhibitors (e.g., venlafaxine), and long-acting β-blockers (e.g., propranolol). Pregabalin may be helpful for acute symptoms and relapse prevention in GAD, as well as for patients with treatment-resistant GAD.

Nonpharmacologic Treatment

CBT, especially when combined with relaxation training, appears to be more effective than nondirective and supportive therapy for the treatment of GAD. In addition, CBT is superior to behavioral therapy or relaxation therapy alone. Biofeedback (enabling patients to see or hear feedback about their physiologic state, such as their heart rate or muscle tension), when combined with CBT, can also be helpful.

Obsessive-Compulsive Disorder

OCD is now classified in DSM-5 under a new category of disorders called *Obsessive Compulsive and Related Disorders*. In the past, OCD was included under *Anxiety Disorders* because anxiety is frequently associated with the OCD symptoms and because yielding to compulsions temporarily decreases anxiety. Recent research suggests that OCD more aptly belongs with a separate grouping of compulsive spectrum disorders. Typically the disorder is characterized by persistent, involuntary thoughts that then provoke anxiety and involuntary anxiety management rituals. Unlike the acute onset and short duration of panic anxiety symptoms or the chronic symptoms of GAD, the obsessions and compulsions that characterize OCD are localized but nevertheless affect all areas of functioning.

OCD appears to be a chronic condition for most people. More than 50% of patients diagnosed with OCD have a chronic and progressive course, 25% to 33% have a fluctuating course, and fewer than 15% have a phasic course with periods of complete remission. Occasionally, there can be sudden onset of symptoms, especially when there is a neurologic basis for the illness. Predictors for poor prognosis include an early age of onset, longer duration of the illness, presence of obsessions and compulsions, poorer baseline social functioning, and presence of magical thinking.

People with OCD typically strive to avoid disclosing their symptoms to relatives, friends, and health professionals. As such, accurate incidence and prevalence statistics are nearly impossible to determine. The lifetime prevalence rate for OCD in the general U.S. population appears to be approximately 2%; approximately 30% to 50% of people with OCD have childhood onset, typically occurring before 10 years of age. The risk in first-degree relatives of probands with OCD is significantly higher. Although there is no gender difference in the prevalence rates of OCD, there is some indication that OCD might have its onset or worsen during pregnancy. In addition, of children diagnosed with OCD, 70% are males.

Etiology and Neurobiology

Neurobiological research findings indicate that there is a strong genetic or inherited risk of OCD. Twin and family studies show a greater degree of concordance for OCD among monozygotic twins compared with dizygotic twins. Once an adult family member has been diagnosed with the disorder, child relatives are more likely to be diagnosed. Depression, anorexia, and Tourette syndrome are common OCD comorbid disorders. Close links between OCD and the hereditary neurologic disorder Tourette syndrome have been reported, but it is not clear whether this link represents an increased risk of OCD or whether Tourette syndrome and OCD are related in some other way.

OCD studies using PET brain scans have shown significant increases in glucose metabolism rates in the frontal lobes, caudate nucleus, and cingulate gyrus regions of the brain (Fig. 49.1). These brain regions are directly associated with response to strong emotions. However, several OCD models of altered brain functioning have been hypothesized. One model proposes that a causal pathway for OCD exists between the frontal cortex region and basal ganglia region of the brain. This model draws on the observation that similar illnesses with well-defined etiologic factors have been shown to involve both cognitive and motor brain regions. Predictably, serotonin activity dysregulation and dysfunction also represent possible OCD models.

As shown in Fig. 49.1, PET scans of the brain of a person with OCD reveal significant increases in glucose metabolism activity in the prefrontal cortex brain region. Magnetic resonance imaging (MRI) findings have suggested widely distributed cellular abnormalities such as significantly lower amounts of total white matter (connection fibers) and greater cortex cell volume. More recent models of OCD focus on the possibility

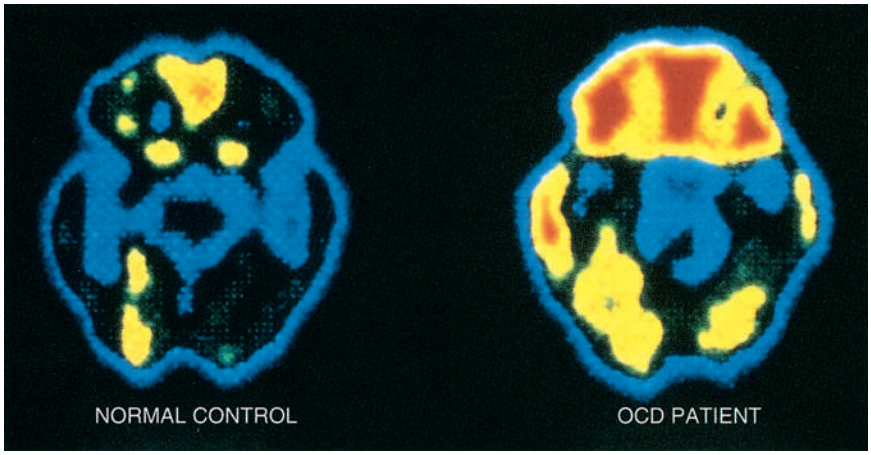


FIG 49.1 Hyperactivity of the orbitofrontal cortex has been a consistent finding in more than a decade of brain imaging research on patients with obsessive-compulsive disorder (OCD). These positron emission tomographic images are from the initial report of this finding by a UCLA group. This excessive metabolic activity could generate spurious "error detection" signals that result in patients with obsessive-compulsive disorder experiencing repetitive adventitious feelings that "something is wrong." (Originally adapted from Baxter LR Jr, et al: Local cerebral glucose metabolic rates in obsessive-compulsive disorder: a comparison with rates in unipolar depression and in normal controls. *Arch Gen Psychiatry* 4(3):211–218, 1987. As published in Schwartz JM: Obsessive-compulsive disorder, *Sci Med* 4:16, 1997.)

of deficient serotonin inhibitory action in the basal ganglia region of the brain, which then permits excessive release of dopamine, a stimulating neurotransmitter.

PET, functional MRI (fMRI), and single-photon emission computed tomography studies of persons with OCD have confirmed the correlation between OCD symptoms and abnormal brain circuit activity in the orbitofrontal cortex, caudate nucleus, anterior cingulate cortex, and thalamus. An fMRI study of medicated OCD patients and comparison subjects explored the possibility of meaningful links between anterior cingulate cortex activity and the severity of OCD symptoms by observing this activity under laboratory stress designed to trigger symptom onset. Hyperactivity was observed in one brain region (anterior cingulate cortex) and was significantly related to error making and expressed doubt, suggesting a neurobiological model for the disabling self-corrective urges associated with OCD. This theory is often referred to as the *cortico-striato-thalamo-cortical model of OCD.*

Evidence indicates that gene variants in serotonin and dopamine increase the risk of OCD. In addition, research from neuroimaging, animal models, and candidate gene and treatment studies suggest that there is a glutamate-signaling dysfunction in OCD. For example, nonmedicated patients with OCD appear to have higher levels of glutamate in the cerebrospinal fluid than healthy normal controls, as well as in psychiatrically matched controls. Although these high levels of glutamate may be a consequence of the illness rather than a cause, more recent genetic research suggests it may play a causal factor.

Research on cognitive thought processes suggests that cognitions do contribute to the maintenance of the disorder, even if they are not the genesis. The following are some examples of faulty cognitions observed in patients with OCD: responsibility and overestimation of threat; perfectionism and intolerance of uncertainty; and importance and control of thoughts. There is some evidence that people with OCD have deficits in selective attention, memory bias toward disturbing themes, and decreased confidence in memory.

Clinical Manifestations

The hallmarks of OCD are obsessions or compulsions. *Obsessions* are strong, persistent, intrusive, uncontrollable thoughts. Obsessive thoughts manifest as ideas, images, and urges that dominate normal thinking and functioning. Affected persons recognize that their obsessions are products of their own mind and may judge them as senseless. Nevertheless, they are unable to stop, govern, or resist their obsessions. *Compulsions* are repetitive, ritualistic behaviors (e.g., handwashing, ordering, checking) or mental acts (e.g., praying, counting, repeating words silently) that the person performs with urgency and rigidity. They are typically related to the obsession in terms of content. Symptom clusters are as follows: concern about dirt and contamination, counting, and a third group that is purely obsessional (with no compulsions). In the obsessional-only group, "slowness" is the main symptom, where patients often conduct daily activities at an extremely slow pace. Overall 90% of the patients have features of both obsessions and compulsions. However, 28% are most bothered by obsessions, 20% by compulsions, and 50% by both. A recent meta-analysis with more than 2000 patients found four consistent "syndromes": symmetry/ordering, contamination/cleaning, hoarding, and obsessions/checking. Other subtypes include sexual, religious, aggressive, or somatic obsessions. "Washers" comprise from 25% to 50% of the OCD samples and are concerned with dirt, contaminants, and germs. These people frequently spend hours each day washing hands or showering. "Checkers" compulsively check to see if they have done something, such as run over someone with a car or left the door unlocked. "Hoarding," the inability to dispossess of meaningless, worthless objects, is one example of the more complex compulsions. Some investigators suggest that hoarding behaviors occur in about 25% to 30% of individuals with OCD. Mental compulsions should not be overlooked, because they are quite common and are often undetected because most clinicians only ask about behavioral rituals. Approximately 80% of OCD patients have both behavioral and mental compulsions; these are the third most common type of compulsions. Uncertainty is also a hallmark of the compulsions, which results from a discrepancy between sensory information and internal beliefs.

Treatment

Either exposure therapy and response prevention or CBT, when combined with effective medication treatment, can effectively diminish OCD obsessions and compulsions. Nevertheless, the disorder itself makes

entering treatment extremely difficult, if not impossible. To be successful, CBT must target the person's obsessions and compulsions. This would require full disclosure, and as stated earlier, disclosure is very difficult. Medications that target serotonergic and dopaminergic neurotransmission (e.g., fluvoxamine, paroxetine, sertraline, venlafaxine, fluoxetine) appear to be helpful with OCD, although more than 25% of OCD sufferers receive little or no benefit. Glutamate-modulating medications (e.g., riluzole, memantine) appear to be helpful in treatment-refractory OCD, although to date most studies have relied on small clinical trials. Atypical antipsychotics, such as risperidone, in combination with antidepressants and β-blockers may be considered when OCD symptoms are disabling. Effective OCD symptom relief has been reported with the TCA clomipramine. Some treatment-refractory OCD patients have benefited from deep brain stimulation, a neurosurgical procedure that implants electrodes and targets regions of the brain affected by OCD (i.e., the anterior limb of the internal capsule, the nucleus accumbens and ventral striatum, and the subthalamic nucleus).

Posttraumatic Stress Disorder

First introduced in DSM-III, PTSD validated that the constellation of symptoms experienced by war veterans did, in fact, cause real impairment. Although anxiety is the most prominent symptom of PTSD, depression and dissociation are also common. PTSD is included in the DSM-5 under *Trauma and Stressor Related Disorders,* in contrast to its previous categorization as an anxiety disorder in the DSM-IV TR. DSM-5 identifies the triggers of PTSD as exposure to or actual threatened death, serious injury, or sexual violation either by directly experiencing the traumatic event, witnessing the event, learning the event occurred to a close family member or friend, and/or experiencing firsthand repeated or extreme exposure. These disturbances must cause clinically significant impairment (e.g., in the individual's social interactions, their capacity to work) in order for the individual to be diagnosed with PTSD. DSM-5 pays particular attention to the behavioral symptoms that often accompany PTSD with four distinct diagnostic categories: reexperiencing, avoidance, negative cognitions and mood, and arousal.

Although 70% of adults in the United States have reported experiencing some type of traumatic event at some point in their life, only 8% of the U.S. population have PTSD at any given time, whereas almost 50% of outpatient mental health patients report PTSD. PTSD is typically a chronic condition. One longitudinal study with young adults reported that more than half of the sample showed no signs of remission after a 3- to 4-year period. Other studies have indicated that as long as 5 years after trauma, 82% of the PTSD sample was not in remission. Women appear to take four times longer to recover from PTSD than men and are two times more likely to meet the criteria for PTSD than men. Although this may be attributed to the higher rates of trauma and sexual violence that women have experienced as children, it is probably also because women respond to trauma differently than men (e.g., women may be more likely to blame themselves for the event than men). There appears to be a strong relationship between PTSD and suicidal thoughts and behaviors, and PTSD is associated with poorer outcome posttreatment.

Etiology and Neurobiology

Acute stress disorder is highly predictive of PTSD. Additional psychological risk factors include the following: past history of trauma (and PTSD), especially childhood trauma; depression; anxiety disorders; comorbid personality disorders; familial history of anxiety disorders, PTSD, and disrupted parental attachments; and severity of exposure to trauma. High intelligence appears to be protective of PTSD. DSM-5 is considering a special section for preschool children because their manifestation of the disorder is very different from that observed in adults.

One theory that has found support both in animal models and with patients is that during or after trauma, the stress response becomes dysregulated and chronic autonomic hyperactivity occurs. The result is hyperarousal and intrusive recollections, or the so-called *positive symptoms* of PTSD. Evidence suggests that there is limbic hyperactivity in the amygdala and cingulate, cortical hyporesponsivity in both prefrontal and Broca areas, HPA axis dysregulation, noradrenergic activation, and heightened physiologic response to traumatic stimuli. Endogenous opioid systems also appear to be overreactive in PTSD; that is, individuals who have sustained prolonged or repeated trauma have endogenous opiates readily released with any reminder of the trauma, which leads to analgesia and psychic numbing. Less is known about the role of the serotonergic system in PTSD, but the theory is that irritability and outbursts are related to serotonergic deficit. Some investigators have suggested that there are separate subgroups of PTSD subjects, where some show serotonin deficits and others exhibit noradrenergic sensitization.

Multiple studies on brain neuroanatomy and neurocircuitry suggest that PTSD patients show decreased right and left hippocampus volumes compared with traumatized and nontraumatized control subjects. It remains unclear whether hippocampi size plays a role in contributing to preexisting vulnerability to PTSD, acts as a developmental determinant, or is the outcome of traumatic stress. Limited research is available to compare brain neuroanatomy before and after PTSD treatment; some results do show that treatment using SSRIs promotes hippocampal neurogenesis and increased volume.

Cognitive and behavioral models of PTSD suggest that classical conditioning may play a role. Behaviorally, patients with PTSD could have higher sympathetic system arousal at the time of conditioning, which then allows them to be more likely to be classically conditioned than trauma-exposed individuals without PTSD. Some studies suggest a higher incidence of PTSD after severe traumatic brain injury with loss of consciousness and few traumatic memories; this suggests that trauma itself could mediate PTSD at an implicit level. There are also impairments in explicit memory associated with PTSD that could be attributable to hippocampal toxicity.

Research on gender differences in the expression of PTSD suggests that women's response to stress is probably different from men's stress response, which sets the stage for women's higher incidence of PTSD. For example, women have lower cortisol levels, reduced HPA responses, and decreased serotonin transporter gene promoter polymorphism than men. They also appear to have higher levels of nonsulfated dehydroepiandrosterone (DHEA) than men, who have higher levels of sulfated DHEA. The increased exposure of women to adverse events is also associated with depression and anxiety, and probably mediated by the corticotropin-releasing factor.

Clinical Manifestations

There are five main characteristics of PTSD. First, the person must have been *exposed to a traumatic event* (actual or threatened) by personally experiencing the event(s), witnessing the event(s), learning about the event(s), or experiencing repeated or extreme exposure to the event's details. Second, the person must experience *intrusive symptoms* associated with the traumatic event(s), including spontaneous or cued recurrent, involuntary, and intrusive distressing memories of the traumatic event(s); recurrent distressing dreams related to the event(s); dissociative reactions such as flashbacks that feel as if the traumatic event(s) was (were) recurring; intense or prolonged psychological distress at exposure to internal or external cues that symbolize or resemble an aspect of the traumatic event(s); and marked physiologic reactions to reminders of the traumatic event(s).

Third, the person persistently *avoids all stimuli* associated with the traumatic event(s), including internal reminders (thoughts, feelings, or

physical sensations) and external reminders (people, places, conversations, activities, objects, situations) that arouse recollections of the traumatic event(s). Fourth, there are *negative cognitive and mood changes* associated with the traumatic event(s), including the inability to recall the event(s) (typically dissociative amnesia; not due to head injury, alcohol, or drugs); persistent and exaggerated negative expectations about oneself, others, or the world; persistent distorted blame of self or others about the cause or consequences of the traumatic event(s); pervasive negative emotional state (e.g., fear, horror, anger, guilt, or shame); markedly diminished interest or participation in significant activities; feelings of detachment or estrangement from others; and persistent inability to experience positive emotions (e.g., unable to have loving feelings, psychic numbing). Finally, there is a *change in arousal and reactivity* associated with the traumatic event(s), including irritability, aggression, recklessness, self-destructiveness, hypervigilance, exaggerated startle response, problems with concentration, and sleep disturbance. None of these symptoms can be the direct physiologic effects of a substance (e.g., medication or alcohol) or a general medical condition (e.g., traumatic brain injury, coma).

Although both men and women report hyperarousal, reexperiencing, avoidance, and numbing, it is more common for women to feel anxious, to have more trouble feeling emotions, and to avoid things that remind them of the trauma. Women are also more likely than men to experience depression and general anxiety, whereas men with PTSD are more likely than women to have problems with alcohol or drugs. In contrast, men with PTSD are more likely to express anger and have anger management issues than women.

Pharmacologic Treatment

The purpose of early pharmacologic treatment interventions after exposure to a trauma is to ameliorate psychological and neurobiological changes that promote PTSD symptoms and behavior. SSRIs are the first-line treatment for PTSD and are associated with marked improvement. To date, most studies on SSRIs have been with fluoxetine, sertraline, and fluvoxamine. Newer studies have reported success with venlafaxine extended release and mirtazapine. TCAs have not been reported to be very effective, but adrenergic blockers show some improvement with both PTSD and acute stress disorder symptoms. Treatment of PTSD is not always effective with SSRIs, however; therefore research has begun to explore different pharmacologic agents in managing the illness.

Although interest in the utilization of antiadrenergic agents in the treatment of PTSD first surfaced in the early 1980s, their use and empirical exploration have not been widely carried out until recently. The alpha-1 postsynaptic receptor antagonist medication prazosin has been found useful in preventing traumatic nightmares. Propranolol, a postsynaptic beta receptor antagonist, has also shown efficacy in treating both the prevention of PTSD-related psychophysiological arousal and in the presence of mental imagery of the traumatic event the individual witnessed. In numerous randomized control trials, atypical antipsychotic agents, such as risperidone and olanzapine, have been administered to PTSD patients with favorable results. Further research is needed to examine the efficacy of these drugs, with several multisite replication studies in progress.

Nonpharmacologic Treatment

CBT approaches have been found to be helpful in the treatment of PTSD. Most treatment protocols include gradual or graded exposure to the trauma with imagination, real-life, and/or virtual reality therapy. Treatment also typically includes cognitive restructuring or reformulations and relaxation techniques. One relatively new treatment is called *eye movement desensitization and reprocessing;* results may be comparable to those achieved from use of exposure and stress inoculation.

KEY POINTS

- Anxiety disorders are characterized by irrational and debilitating fears. The four major categories of anxiety disorders are PD, GAD, OCD, and PTSD. These disorders show some evidence of heritability, and biochemical correlates are suspected. Defects in serotonin pathways have been proposed as etiologic factors. PD, GAD, and PTSD are more likely to occur in females than in males.
- PD is characterized by acute episodes of severe anxiety accompanied by dyspnea, chest pain, and a sense of impending doom. Palpitations, hyperventilation, dizziness, paresthesias, and diaphoresis may occur during an attack, which may last 5 to 30 minutes. Anticipatory anxiety and phobic avoidance may develop in individuals with PD.
- GAD is characterized by a continuous but moderate degree of anxiety without discrete periods of acute attacks. Agoraphobia rarely develops, but chronic headaches, muscle tension, abdominal discomfort, and sleep disturbances are common.
- OCD is characterized by obsessive thoughts and compulsive behavior. Obsessive thoughts manifest as ideas, images, and urges that dominate normal thinking and functioning. Compulsions are repetitive, ritualistic behaviors (e.g., handwashing, ordering, checking) or mental acts (e.g., praying, counting, repeating words silently) that the person performs with urgency and rigidity. They are typically related to the obsession in terms of content. There are four consistent syndromes: symmetry/ordering, contamination/cleaning, hoarding, and obsessions/checking, but they can also include sexual, religious, aggressive, or somatic content. Uncertainty is also a hallmark of the compulsions, which results from a discrepancy between sensory information and internal beliefs.
- PTSD is categorized under the *Trauma and Stressor-Related Disorders* section in the DSM-5. It is usually precipitated by a traumatic event, after which the person typically experiences intrusive symptoms, avoids all stimuli associated with the event(s), has negative cognitive and mood changes, and experiences a change in arousal and reactivity. PTSD is frequently chronic, and women take longer to recover than do men.
- Benzodiazepines, antidepressants, and possibly antianergic agents may be used to manage all anxiety disorders; CBT has proven to be quite successful.

NEURODEVELOPMENTAL DISORDERS

DSM-5 contains a new section called *Neurodevelopmental Disorders;* this includes diagnoses previously listed in DSM-IV under the chapters of *Disorders Usually First Diagnosed in Infancy, Childhood, or Adolescence.* Neurodevelopmental disorders are disabilities experienced by children that are primarily associated with the functioning of the neurologic system and brain. Examples include intellectual disability (previously known as *mental retardation*), attention-deficit/hyperactivity disorder (ADHD), autism spectrum disorder, and learning disabilities. Children with these disorders typically experience problems in language and speech, motor skills, behavior, memory, learning, or other neurologic functions. Although the symptoms and behaviors can change as a person ages, many children with neurodevelopmental disorders will have permanent disabilities. As many as 12% of children in the United States ages 3 to 17 years are affected by at least one neurodevelopmental disorder, such as ADHD, learning disorders, or intellectual disability. Some investigators believe that the prevalence of certain neurodevelopmental disorders, specifically autism and ADHD, has increased during the past 40 years. For example, the percentage of children reported to have ever been diagnosed with autism rose from 0.1% in 1997 to 0.7%

in 2008. Because of a lack of long-term data and changes in awareness and diagnostic criteria, it is difficult to determine whether this is true. Most neurodevelopmental disorders result from a combination of genetic, biological, psychosocial, and environmental risk factors, as well as behavioral risk factors such as alcohol, tobacco, or illicit drug use. There are also known environmental contaminants that can damage a child's developing brain and nervous system, such as lead, methyl mercury, and polychlorinated biphenyls.

Attention-Deficit/Hyperactivity Disorder

ADHD is the most common psychiatric disorder among children. It typically begins in early childhood and continues throughout adolescence; it can be chronic, lasting through adulthood. ADHD is characterized by difficulty staying focused and paying attention, difficulty controlling behavior, and hyperactivity (overactivity). There are three subtypes of ADHD: predominantly hyperactive-impulsive, predominantly inattentive, and combined hyperactive-impulsive and inattentive. Children with ADHD typically have poorer academic performance and higher rates of learning disabilities. It is also common for these children to need special classroom placement, tutoring, or even to repeat a grade. When diagnosing ADHD, it is important to use several sources such as parent, child, and teacher ratings.

DSM-5 more accurately categorizes the experience of adults diagnosed with ADHD, as extensive research has revealed that although this disorder begins in childhood, it continues into adulthood for many individuals. The estimated worldwide prevalence of ADHD is between 5.3% and 7.1% in children and adolescents; it is 3.4% in adults. It accounts for as many as 3% to 50% of all mental health service referrals for children and occurs four times more often in males. In females, the inattentive type occurs more frequently than the hyperactive type. Interestingly, the gender ratio disappears in adults, but this could be due to methodologic differences in studies.

Risk factors for ADHD include family history of ADHD, psychosocial adversity, and comorbidity with oppositional defiant disorder (ODD) (50%) and conduct disorder (CD) (30%), mood disorder (15% to 20%), and anxiety disorders (20% to 25%) as well as tic disorders. As many as two thirds of all individuals diagnosed with ADHD are diagnosed with other psychiatric disorders. When diagnosed with both ADHD and either ODD or CD, the child will typically have greater disability and longer persistence of symptoms. As the child grows into adolescence, ADHD is highly associated with substance use disorders. It can be difficult to differentiate ADHD symptoms from bipolar disorder (BD) symptoms in children because many of the symptoms overlap (e.g., distractibility, impulsivity, hyperactivity, mood swings, and irritability). Usually children with BD have elevated mood, decreased need for sleep, and grandiosity. As many as 75% of the children diagnosed with ADHD continue to have problems into adulthood, with the hyperactivity and impulsivity remitting and the inattention remaining.

Etiology and Neurobiology

A number of biological and environmental factors have been implicated in ADHD, and a biopsychosocial model is best incorporated. Genetic models indicate that 25% to 50% of cases occur in families. First-degree relatives have a 15% to 25% chance of the disorder, and 50% if both parents have the disorder. Research suggests that there are abnormalities in genes coding for proteins in central nervous system dopamine function, and specific genes include the dopamine receptor gene, the dopamine transporter gene, and the dopamine-β-hydroxylase gene. The specific pathophysiology of ADHD remains unclear, but research has shown that the neural circuits of the prefrontal cortex and striatum, as well as the brainstem catecholamine systems that innervate these circuits, show abnormalities. Exposure to toxins during pregnancy does appear to contribute to the development of ADHD, including maternal smoking and lead and alcohol exposure. In addition, low birth weight, antepartum hemorrhage, prolonged labor, and low Apgar scores are associated with ADHD. The role of diet (e.g., preservatives, artificial dyes, food allergies) on hyperactivity remains controversial. Many scientists believe that this disorder is caused by a combination of factors, such as genetic susceptibility with exposure to environmental contaminants, rather than by any one factor.

A meta-analysis published in 2012 reviewed 55 studies; 39 included child participants and 16 included adult participants. This study shed light upon the proposed neurobiology involved in the presentation of ADHD. Hypoactivation in children was found to almost always implicate systems involved in executive functions (the frontoparietal network) and attention, in contrast to hyperactivation, which implicated default, ventral attention and somatomotor networks. Those studies that evaluated adult participants indicated that ADHD-related hypoactivation was strongly associated with frontoparietal system activity, whereas ADHD-hyperactivation was associated with activity in the visual, dorsal attention, and default networks.

Clinical Manifestations

To be diagnosed with ADHD, a child must have symptoms for 6 or more months (symptoms within the cluster) and to a degree that is greater than that found in other children of the same age. For either type of ADHD, symptoms are present by age 12 and are apparent in two or more settings (e.g., at home, school, or work; with friends or relatives; or in other activities). The symptoms must also interfere with or reduce the quality of social, academic, or occupational functioning.

Children who have symptoms of *inattention* are described in the following ways: (a) fails to give close attention to details; (b) has difficulty sustaining attention in tasks or play activities; (c) often does not seem to listen when spoken to directly; (d) frequently does not follow through on instructions; (e) often has difficulty organizing tasks and activities; (f) characteristically avoids, seems to dislike, and is reluctant to engage in tasks that require sustained mental effort; (g) frequently loses objects necessary for tasks or activities; (h) is often easily distracted by extraneous stimuli; and (i) is often forgetful in daily activities, chores, and errands.

Children who have symptoms of *hyperactivity and inattention* are described in the following ways: (a) often fidgets or taps hands or feet or squirms; (b) often restless during activities when others are seated; (c) often runs about or climbs on furniture and moves excessively in inappropriate situations; (d) is often excessively loud or noisy during play, leisure, or social activities; (e) often "on the go," acting as if "driven by a motor"; (f) often talks excessively; (g) often blurts out an answer before a question has been completed; (h) has difficulty waiting his or her turn or waiting in line; (i) often interrupts or intrudes on others; (j) tends to act without thinking; (k) often impatient; (l) uncomfortable doing things slowly and systematically; and (m) finds it difficult to resist temptations or opportunities.

Pharmacologic Treatment

Most ADHD treatments focus on reducing the symptoms in order to improve functioning. Medication treatments especially do not cure the disorder but treat the symptoms only while the medication is taken. Treatment typically includes medication, psychotherapy, education or training, or a combination of these treatment modalities. Medications commonly include psychostimulants that paradoxically have a calming effect and are known to reduce inattention, impulsivity, and hyperactivity, and occasionally OCD and anxiety disorders. Stimulant medications are available in multiple forms (e.g., pill, capsule, liquid, skin patch) and in short-acting, long-acting, or extended-release varieties. The most

commonly used psychostimulant is methylphenidate, which is available in both short-acting and long-acting agents (e.g., Concerta, Metadate CD, and Ritalin LA). Side effects are minor and typically disappear over time; these include delayed sleep, decreased appetite, and sometimes anxiety and irritability. Some children also report mild stomachaches or headaches. For children who do not respond well to psychostimulants, atomoxetine (Strattera) can be used, as well as TCAs, α-agonists, and some anxiolytic medications.

Nonpharmacologic Treatment

Psychosocial interventions are effective in teaching the child behavioral and social skills. These can range from practical assistance (such as helping the child organize tasks or complete schoolwork) to assistance in crisis situations (such as helping the child cope with emotionally difficult events). Behavioral therapies can also teach children to monitor their own behavior by controlling anger or by thinking before acting. In addition, psychotherapy can teach basic social skills, such as how to wait their turn, share toys, ask for help, or respond to teasing.

Autism Spectrum Disorder

Autism was first introduced in the DSM-III in the early 1980s and modified in the DSM-IV in the mid-1990s. In past versions of DSM, autism, Asperger disorder, and child disintegrative disorder were each separate conditions. Because of a wide research base that is both reliable and valid, DSM-5 now groups these and similar disorders together under the category of *Autism Spectrum Disorder* (ASD). ASD is defined by a common set of behaviors: social/communication deficits and fixated interests and repetitive behaviors. For example, ASD patients may be interested in social interactions, but they lack the skills for reciprocal interaction. Normal attachment appears to be impaired, and often there is a lack of social referencing, where others are seen and treated as objects. Individuals with autism frequently show difficulty with change and often perform ritualized, compulsive behaviors or thoughts that are reflected in repetitive questions and physical mannerisms. Clients with ASD also typically have unusual responses to sensory stimuli as well as poor motor imitation, gait, and tone. In the past, language delays were a defining feature of the disorder. The DSM-5 considers language delays only as factors that influence the clinical symptoms; instead, they are placed into the categories of social/communication or restricted, repetitive patterns of behavior. In addition, autistic symptoms must be present in early childhood (although they may not fully manifest until social demands exceed limited capacities), and symptoms must limit and impair everyday functioning. Autism has been associated with intellectual disability because between 66% and 75% of all individuals with autism have intellectual impairment. However, mental retardation and autistic disorder are considered distinct from one another.

Based on past diagnostic criteria, investigators believe that the prevalence for autism is between 0.15 and 34.00 per 10,000, whereas that for Asperger syndrome is between 0.6 and 10.0 per 10,000 persons. Studies in Asia, Europe, and North America have identified individuals with ASD with an average prevalence of about 1%. In March 2014, The Centers for Disease Control and Prevention released new data on the prevalence of autism in the United States identifying 1 in 68 children (1 in 42 boys and 1 in 189 girls) as having ASD. Some studies suggest, however, a prevalence rate as high as 1 in 45 children having been diagnosed with the illness. ASD is almost five times more common in boys than it is in girls, although females appear to be affected more severely and have greater cognitive impairment. Prevalence rates for both disorders have increased since the mid-1990s. Although it is still unclear whether this is due to an actual increase in incidence, some research suggests that it is due to both changing diagnostic practices that improve identification, as well as a true increase in the illness. Scientists suggest further research regarding the etiologic nature of ASD needs to be produced in order to understand its environmental or genetic causes.

Etiology and Neurobiology

To date, there is no known identifiable cause for autism. Popular media sources frequently suggest that measles-mumps-rubella vaccine or thimerosal exposure plays a role in the etiology; however, this has not been supported by research. Some medical illnesses are associated with autism, and these include tuberous sclerosis, fragile X syndrome, maternal rubella, congenital hypothyroidism, phenylketonuria, Down syndrome, neurofibromatosis, and Angelman syndrome. Most scientists agree that genetic vulnerabilities are one of the risk factors that can increase the likelihood that an individual may develop ASD; this includes higher concordance in monozygotic twins and low concordance in dizygotic twins, as well as greater-than-expected incidences of anxiety disorders, major depressive disorder, and motor tics in first-degree family members. In addition, parents of autistic children show higher rates of associated behaviors, such as rigidity, aloofness, anxiety, and restricted friendships. Research suggests that the critical period for developing ASD occurs before, during, and immediately after birth and that children who are born to older parents are at an increased risk for developing ASD. It is fair to say that this is a new area for research, and future editions may shed more light on its neurobiology.

Clinical Manifestations

First, in DSM-5, ASD is defined by *persistent deficits in social communication and social interactions across contexts.* This includes social-emotional reciprocity, deficits in nonverbal communicative behaviors used for social interaction, and deficits in developing and maintaining relationships appropriate to developmental level (beyond those with caregivers). Second, ASD is defined by *restricted, repetitive patterns of behavior, interests, or activities,* including stereotyped or repetitive speech, motor movements, or use of objects; excessive adherence to routines, ritualized patterns of verbal or nonverbal behavior, or excessive resistance to change; highly restricted, fixated interests that are abnormal in intensity or focus; and hyperreactivity or hyporeactivity to sensory input or unusual interest in sensory aspects of the environment. Both the social communication and the restricted, repetitive patterns of behavior are assigned a severity level, ranging from "1" (requiring support), to "2" (requiring substantial support), to "3" (requiring very substantial support). Thus in the past, "milder" autism was often identified as Asperger disorder; under the new rubric, it would be assigned as level 1 ASD.

Treatment

To date, there are no established pharmacologic treatments for ASD. Research in the developmentally disordered population suggests that SSRIs may be effective in treating associated symptoms of ASD such as compulsive and repetitive behavior, behavioral rigidity, and aggression or that α-agonists may be effective in targeting impulsivity, hyperactivity, and poor concentration. There is anecdotal clinical evidence to suggest that individuals with ASD respond to very low doses (i.e., trivial) of psychoactive or antipsychotic medications; clinicians recommend "starting low and going slow."

A strong body of literature supports the use of evidence-based behavioral interventions for children with ASD. These types of treatment are known to focus on social skills acquisition and environmental modifications, such as the use of schedules, visual cueing, and structured settings. One long-established method is the applied behavioral analysis approach, developed by Dr. Ivar Lovas, that starts with ASD children as young as 2 to 3 years of age and provides intensive, daily sessions. The goal of treatment is to focus on expansion of communication,

emergence of new play skills, development of interactive relationships and more appropriate reaction to sensory input, and development of other pivotal skills such as imitation and requesting. Long-term results have proven successful in both raising IQ and adaptive behavior, with a sizable minority of children achieving normal educational and intellectual functioning by 7 years of age, especially with early intervention.

KEY POINTS

- A number of biological and environmental factors have been implicated in both ASD and ADHD, and a biopsychosocial model is best incorporated. Early research on genetic models has suggested a limited basis of contribution to these disorders.
- ADHD subtypes include symptoms of inattention or hyperactivity and impulsivity. The disorder typically continues into adulthood, with symptoms of inattention most common. Psychostimulant medications can effectively reduce the symptoms and improve daily functioning in most children and adults, but there is no "cure" for the disorder. Psychosocial interventions offer both practical assistance (such as help organizing tasks or completing schoolwork) and assistance in crisis situations (such as dealing with emotionally difficult events); in addition, these interventions teach behavioral self-control.
- In the DSM-5, ASD captures the features of autism and Asperger syndrome. Children are given a severity rating from 1 (requiring support) to 3 (requiring substantial support) for symptoms that fall into the categories of (a) deficits in social communication and social interaction and (b) restricted, repetitive patterns of behavior, interests, or activities. There are no established pharmacologic treatments for ASD; however, behavior intervention, especially if provided early, assists children in developing behavioral, communication, and social skills.

SUMMARY

Hypotheses regarding the cause and pathogenesis, clinical manifestations, and implications for management of representative subsets have been presented. It is important to recognize that these disorders represent a mixture of biological, psychological, social, and environmental factors. Researchers have yet to unravel all of the mysteries surrounding the neurobiological mechanisms of nonpsychotic illnesses.

RESOURCES

Anxiety and Panic Disorders

Abelson JL, Weg JG, Nesse RM, et al: Persistent respiratory irregularity in patients with panic disorder. *Biol Psychiatry* 49(7):588–595, 2001.

American Psychiatric Association: *Diagnostic and statistical manual of mental disorders*, 5th ed, Washington, DC, 2013, The Association.

Baldwin DS, Ajel K, Masdrakis VG, et al: Pregabalin for the treatment of generalized anxiety disorder: an update. *Neuropsychiatr Dis Treat* (9):883–892, 2013.

Bunaciu L, Leen-Feldner EW, Blumenthal H, et al: An experimental test of the effects of parental modeling on panic-relevant escape and avoidance among early adolescents. *Behav Ther* 45:517–529, 2014.

Cuijpers P, Sijbrandij M, Koole S, et al: Psychological treatment of generalized anxiety disorder: a meta-analysis. *Clin Psychol Rev* 34:130–140, 2014.

Davidson JR: Pharmacotherapy of generalized anxiety disorder. *J Clin Psychiatry* 62(Suppl 11):46–50, 2001.

Gaudlitz K, Plag J, Dimeo F, et al: Aerobic exercise training facilitates the effectiveness of cognitive behavioral therapy in panic disorder. *Depress Anxiety* 32:221–228, 2015.

Gorman JM, Kent J, Martinez J, et al: Physiological changes during carbon dioxide inhalation in patients with panic disorder, major depression, and premenstrual dysphoric disorder: evidence for a central fear mechanism. *Arch Gen Psychiatry* 58(2):125–131, 2001.

Heim C, Newport DJ, Bonsall R, et al: Altered pituitary-adrenal axis responses to provocative challenge tests in adult survivors of childhood abuse. *Am J Psychiatry* 158(4):575–581, 2001.

Hollander E, Simeon D: Anxiety disorders. In Hales RE, Yudofsky SC, Gabbard GO, editors: *The American Psychiatric Publishing textbook of psychiatry*, Washington, DC, 2008, American Psychiatric Publishing, pp 505–608.

Kim B, Yoo E, Lee J-Y, et al: The effects of the catechol-O-methyltransferase val158met polymorphism on white matter connectivity in patients with panic disorder. *J Affect Disord* 147:64–71, 2013.

Konishi Y, Tanii H, Otowa T, et al: Gene x gene x gender interaction of BDNF and COMT genotypes associated with panic disorder. *Prog Neuropsychopharmacol Biol Psychiatry* 51:119–125, 2014.

Lesch KP: Molecular foundation of anxiety disorders. *J Neural Transm* 108(6):717–746, 2001.

Levine DS, Himle JA, Taylor RJ, et al: Panic disorder among African Americans, Caribbean blacks and non-Hispanic whites. *Soc Psychiatry Psychiatr Epidemiol* 48:711–723, 2013.

Mathew SJ, Coplan JD, Gorman JM: Neurobiological mechanisms of social anxiety disorder. *Am J Psychiatry* 158(10):1558–1567, 2001.

McGinn LK, Nooner KB, Cohen J, Leaberry KD: The role of early experience and cognitive vulnerability: presenting a unified model of the etiology of panic. *Cogn Ther Res* 3:408–519, 2015.

Nutt DJ: Neurobiological mechanisms in generalized anxiety disorder. *J Clin Psychiatry* 62(Suppl 11):22–27, 2001.

Olivares JM, Alvarez E, Carrasco JL, et al: Pregabalin for the treatment of patients with generalized anxiety disorder with inadequate treatment response to antidepressants and severe depressive symptoms. *Int Clin Psychopharmacol* 30:265–271, 2015.

Rickels K, Rynn MA: What is generalized anxiety disorder? *J Clin Psychiatry* 62(Suppl 11):4–12, 2001.

Sheikh JI: Anxiety in older adults. Assessment and management of three common presentations. *Geriatrics* 58(5):44–45, 2003.

Teicher MH, Andersen SL, Polcari A, et al: The neurobiological consequences of early stress and childhood maltreatment. *Neurosci Biobehav Rev* 27(1-2):33–44, 2003.

Wilhelm FH, Trabert W, Roth WT: Physiologic instability in panic disorder and generalized anxiety disorder. *Biol Psychiatry* 49(7): 596–605, 2002.

Obsessive-Compulsive Disorder

Brown WA: Hoarding. *Psychiatr Times* 24(13):50–52, 2007.

Clayton IC, Richards JC, Edwards CJ: Selective attention in obsessive-compulsive disorder. *J Abnorm Psychol* 108(1):171–175, 1999.

de Koning PP, Figee M, van den Munckhof P, et al: Current status of deep brain stimulation for obsessive-compulsive disorder: a clinical review of different targets. *Curr Psychiatry Rep* 13:274–282, 2011.

Fisher PL, Wells A: How effective are cognitive and behavioral treatments for obsessive-compulsive disorder? A clinical significance analysis. *Behav Res Ther* 43(12):1543–1558, 2005.

Kadri N, Alami K: Depression and anxiety among women. In Chandra P, Herrman H, Fisher J, et al, editors: *Contemporary topics in women's mental health: global perspectives in a changing society*, Hoboken, NJ, 2009, Wiley-Blackwell, pp 37–64.

MacDonald PA, Antony MM, MacLeod CM: Memory and confidence in memory judgments among individuals with obsessive compulsive disorder and non-clinical controls. *Behav Res Ther* 35(6):497–505, 1997.
Mataix-Cols D, Rosario-Campos MC, Leckman JF: A multidimensional model of obsessive-compulsive disorder. *Am J Psychiatry* 162(2):228–238, 2005.
Pauls DL: The genetics of obsessive-compulsive disorder: a review. *Dialogues Clin Neurosci* 12:149–163, 2010.
Pauls DL, Abramovitch A, Rauch SL, et al: Obsessive-compulsive disorder: an integrative genetic and neurobiological perspective. *Nat Rev Neurosci* 15(6):410–424, 2014.
Radomsky AS, Rachman S: Memory bias in obsessive-compulsive disorder (OCD). *Behav Res Ther* 37(7):605–618, 1999.
Ruscio AM, et al: The epidemiology of obsessive-compulsive disorder in the National Comorbidity Survey Replication. *Mol Psychiatry* 15(1):53–63, 2010.
Salkovskis PM, Wroe AL, Gledhill A, et al: Responsibility attitudes and interpretations are characteristic of obsessive compulsive disorder. *Behav Res Ther* 38(4):347–372, 2000.
Taylor S, McKay D, Abramowitz JS: Hierarchical structure of dysfunctional beliefs in obsessive-compulsive disorder. *Cogn Behav Ther* 34(4):216–228, 2005.
Ursu S, Stenger VA, Shear MK, et al: Overactive action monitoring in obsessive-compulsive disorder: evidence from functional magnetic resonance imaging. *Psychol Sci* 14(4):347–353, 2003.
Winsber ME, Cassic KS, Koran LM: Hoarding in obsessive-compulsive disorder: a report of 20 cases. *J Clin Psychiatry* 60(9):591–597, 1999.
Wu K, Hanna GL, Rosenberg DR, et al: The role of glutamate signaling in the pathogenesis and treatment of obsessive–compulsive disorder. *Pharmacol Biochem Behav* 100(4):726–735, 2012.
Zohar AH: The epidemiology of obsessive-compulsive disorder in children and adolescents. *Child Adolesc Psychiatr Clin N Am* 1999.

Post-Traumatic Stress Disorder

Anisman H, Pizzino A, Sklar LS: Coping with stress, norepinephrine depletion and escape performance. *Brain Res* 191(2):583–588, 1980.
Bremner JD, Randall P, Scott TM, et al: MRI-based measurement of hippocampal volume in patients with combat-related posttraumatic-stress-disorder. *Am J Psychiatry* 152(7):973–981, 1995.
Bremner JD, Vermetten E: Neuroanatomical changes associated with pharmacotherapy in posttraumatic stress disorder. *Ann N Y Acad Sci* 1032:154–157, 2004.
Bryant RA, Marosszeky JE, Crooks H, et al: Posttraumatic stress disorder after severe traumatic brain injury. *Am J Psychiatry* 157(4):629–631, 2000.
Chung MY, Min KH, Jun YJ, et al: Efficacy and tolerability of mirtazapine and sertraline in Korean veterans with posttraumatic stress disorder: a randomized open label trial. *Hum Psychopharmacol* 19(7):489–494, 2004.
Davidson JR, Baldwin D, Stein DJ, et al: Treatment of posttraumatic stress disorder with venlafaxine extended release: a 6-month randomized controlled trial. *Arch Gen Psychiatry* 63:1158–1165, 2006.
Davidson JR, Conner KM, Hertzberg MA, et al: Maintenance therapy with fluoxetine in posttraumatic stress disorder: a placebo controlled discontinuation study. *J Clin Psychopharmacol* 25:166–169, 2005.
Davidson JR, Rothbaum BO, Tucker P, et al: Venlafaxine extended release in posttraumatic stress disorder: a sertraline- and placebo-controlled study. *J Clin Psychopharmacol* 26(5):473, 2006.
De Boer M, Op den Velde W, Falger PJ, et al: Fluvoxamine treatment for chronic PTSD: a pilot study. *Psychother Psychosom* 57:158–163, 1992.
Department of Veteran Affairs [online]: *Women, trauma, and PTSD*. Available at: www.ptsd.va.gov/public/pages/women-trauma-and-ptsd.asp. (Accessed 2 October 2012).
Endler NS, Kocovski NL: Personality disorders at the crossroads. *J Pers Disord* 16(6):487–502, 2002.
Geracioti TD, Baker DG, Ekhator NN, et al: CSF norepinephrine concentrations in posttraumatic stress disorder. *Am J Psychiatry* 158(8):1227–1230, 2001.
Giustino TF, Fitzgerald PJ, Maren S: Revisiting propranolol and PTSD: Memory erasure or extinction enhancement? *Neurobiol Learn Mem* 130:26–33, 2016.
Heim C, Newport DJ, Heit S, et al: Pituitary-adrenal and autonomic responses to stress in women after sexual and physical abuse in childhood. *JAMA* 284(5):592–597, 2000.
Larsen SE, Stirman SW, Smith BN, Resick PA: Symptom exacerbations in trauma-focused treatments: Associations with treatment outcome and non-completion. *Behav Res Ther* 77:68–77, 2016.
Lindauer RJL, Vlieger EJ, Jalink M, et al: Smaller hippocampal volume in Dutch police officers with posttraumatic stress disorder. *Biol Psychiatry* 56(5):356–363, 2004.
Neylan TC, Metzler TJ, Schoenfeld FB, et al: Fluvoxamine and sleep disturbances in posttraumatic stress disorder. *J Trauma Stress* 14(3): 461–467, 2001.
Olff M, Langeland W, Draijer N, et al: Gender differences in posttraumatic stress disorder. *Psychol Bull* 133(2):183–204, 2007.
Orr SP, Metzger LJ, Lasko NB, et al: De novo conditioning in trauma-exposed individuals with and without posttraumatic stress disorder. *J Abnorm Psychol* 109(2):290–298, 2000.
Ozer EJ, et al: Predictors of posttraumtic stress disorder and symptoms in adults: A meta-analysis. *Psychol Bull* 129:3–36, 2003.
Pae C-U, et al: The atypical antipsychotics olanzapine and risperidone in the treatment of posttraumatic stress disorder: a meta-analysis of randomized, double blind, placebo controlled clinical trials. *Int Clin Psychopharmacol* 23(1):1–8, 2008.
Panagioti M, Gooding P, Tarrier N: Posttraumatic stress disorder and suicidal behavior: a narrative review. *Clin Psychol Rev* 29:471–482, 2009.
Perkonigg A, Pfister H, Stein MB, et al: Longitudinal course of posttraumatic stress disorder and posttraumatic stress disorder symptoms in a community sample of adolescents and young adults. *J Am Acad Child Adolesc Psychiatry* 27:567–572, 2005.
Pitman RK, Vanderkolk BA, Orr SP, et al: Naloxone-reversible analgesic response to combat-related stimuli in posttraumatic-stress-disorder—a pilot-study. *Arch Gen Psychiatry* 47(6):541–544, 1990.
Schnurr P, Friedman M: Treatments for PTSD: Understanding the Evidence. *PTSD Res Quart* 19(3):1050–1835, 2008.
Shay J: Fluoxetine reduces explosiveness and elevates mood of Vietnam combat vets with PTSD. *J Trauma Stress* 5(1):97–101, 1992.
Southwick SM, Krystal JH, Bremner JD, et al: Noradrenergic and serotonergic function in posttraumatic stress disorder. *Arch Gen Psychiatry* 54(8): 749–758, 1997.
Stein DJ, Isper JC, Seedat S: Pharmacotherapy for post-traumatic stress disorder. *Cochrane Database Syst Rev* (1):CD002795, 2006.
Tarrier N: The cognitive and behavioral treatment of PTSD: what is known and what is known to be unknown: how not to fall into the practice gap. *Clin Psychol Sci Pract* 17(2):134–143, 2010.
U.S. Department of Veteran Affairs: *PTSD: National center for PTSD*, Available at: http://www.ptsd.va.gov/professional/PTSD-overview/epidemiological-facts-ptsd.asp. (Accessed 20 November 2015).
van der Kolk BA: The psychobiology and psychopharmacology of PTSD. *Hum Psychopharmacol* 16:S49–S64, 2001.
Vermetten E, Vythilingam M, Southwick SM, et al: Long-term treatment with paroxetine increases verbal declarative memory and hippocampal volume in posttraumatic stress disorder. *Biol Psychiatry* 54(7):693–702, 2003.
Young EA, Breslau N: Cortisol and catecholamines in posttraumatic stress disorder—an epidemiologic community study. *Arch Gen Psychiatry* 61(4):394–401, 2004.
Young EA, Breslau N: Saliva cortisol in posttraumatic stress disorder: a community epidemiologic study. *Biol Psychiatry* 56(3):205–209, 2004.
Zlotnick C, Warshaw M, Shea MA, et al: Chronicity in posttraumatic stress disorder (PTSD) and predictors of course of comorbid PTSD in patients with anxiety disorders. *J Trauma Stress* 12(1):89–100, 1999.

Attention Deficit Hyperactivity

American Psychiatric Association: *Highlights of changes from DSM-IV to DSM-5*, 2013. Available at: http://www.dsm5.org/Documents/changes%20from%20dsm-iv-tr%20to%20dsm-5.pdf.

Arnold LE, Abikoff HB, Cantwell DP, et al: National Institute of Mental Health Collaborative Multimodal Treatment Study of children with ADHD (the MTA)—design challenges and choices. *Arch Gen Psychiatry* 54(9):865–870, 1997.

Biederman J, Wilens TE, Spencer T, et al: Diagnosis and treatment of adults with attention-deficit/hyperactivity disorder [Editorial Material]. *CNS Spectr* 12(4):A1–A14, 2007.

Cortese S, et al: Toward systems neuroscience of ADHD; a meta-analysis of 55 fMRI studies. *Am J Psychiatry* 169(10):1038–1055, 2012.

Gainetdinov RR, Mohn AR, Bohn LM, et al: Glutamatergic modulation of hyperactivity in mice lacking the dopamine transporter. *Proc Natl Acad Sci USA* 98(20):11047–11054, 2001.

Scahill L, Schwab-Stone M: Epidemiology of ADHD in school-age children. *Child Adolesc Psychiatr Clin N Am* 9(3):541, 2000.

United States Department of Health and Human Services. Substance Abuse and Mental Health Services Administration. Center for Behavioral Health Statistics and Quality: *National Survey on Drug Use and Health, 2013*. ICPSR35509-v3. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2015-11-23, 2010. http://doi.org/10.3886/ICPSR35509.v3.

Ursano AM, Kartheiser PH, Barnhill LJ: Disorders usually first diagnosed in infancy, childhood, and adolescence. In Hales RE, Yudofsky SC, Gabbard GO, editors: *The American Psychiatric Publishing textbook of psychiatry*, Washington, DC, 2008, American Psychiatric Publishing, pp 861–920.

Wilens TE, Biederman J, Spencer TJ: Attention deficit/hyperactivity disorder across the lifespan. *Annu Rev Med* 53:113–131, 2002.

Wilens TE, Dodson W: A clinical perspective of attention-deficit/hyperactivity disorder into adulthood. *J Clin Psychiatry* 65(10):1301–1313, 2004.

Autism

Abrahams BS, Geschwind DH: Advances in autism genetics: on the threshold of a new neurobiology. *Nat Rev Genet* 9:341–355, 2008.

D'Souza Y, Fombonne E, Ward BJ: No evidence of persisting measles virus in peripheral blood mononuclear cells from children with autism spectrum disorder. *Pediatrics* 118(6):2608, 2006.

Durkin MS, et al: Advanced parental age and the risk of autism spectrum disorder. *Am J Epidemiol* 168(11):1268–1276, 2008.

Eldevik S, Hastings RP, Hughes JC, et al: Meta-analysis of early intensive behavioral intervention for children with autism. *J Clin Child Adolesc Psychol* 38(3):439–450, 2009.

Eldevik S, Jahr E, Eikeseth S, et al: Cognitive and adaptive behavior outcomes of behavioral intervention for young children with intellectual disability. *Behav Modif* 34(1):16–34, 2010.

Eldevik S, Jahr E, Hastings R, et al: Behavioural intervention for children with autism in local mainstream pre-school settings. *J Appl Res Intellect Disabil* 23(5):440, 2010.

McDougle CJ, Hollway J, Scahill L, et al: Risperidone for the core symptom domains of autism: results from the study by the autism network of the research units on pediatric psychopharmacology. *Am J Psychiatry* 162(6):1142–1148, 2005.

National Center for Health Statistics- Center for Disease Control and Prevention: *Autism Prevalence*, (Accessed 27 March 2014).

National Center for Health Statistics- Centers for Disease Control and Prevention: *Autism*, 2015. Available at: http://www.cdc.gov/ncbddd/autism/data.html.

Piven J, Palmer P, Landa R, et al: Personality and language characteristics in parents from multiple-incidence autism families. *Am J Med Genet* 74(4):398–411, 1997.

Rizwan S, Manning JT, Brabin BJ: Maternal smoking during pregnancy and possible effects of in utero testosterone: evidence from the 2D:4D finger length ratio. *Early Hum Dev* 83(2):87–90, 2007.

Shevell M, Fombonne E: Autism and MMR vaccination or thimerosal exposure: an urban legend? *Can J Neurol Sci* 33(4):339–340, 2006.

Simonoff E: Autism spectrum disorder: prevalence and cause may be bound together. *Br J Psychiatry* 2(201):88–89, 2012. doi:10.1192/bjp.bp.111.104703.

Tsai L, editor: *Autistic disorder*, 3rd ed, Washington, DC, 2003, American Psychiatric Publishing.

50

Structure and Function of the Musculoskeletal System

Carol L. Danning

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- What are the functions of osteoblasts and osteoclasts in bone remodeling?
- What is the relationship between joint structure and joint mobility?
- Why is articular cartilage particularly susceptible to degenerative changes?
- What factors determine tendon strength and compliance?
- How does the striated structure of skeletal muscle relate to its contractile function?
- How does an action potential in the α-motor neuron lead to a contraction in the muscle cells of the motor unit?

CHAPTER OUTLINE

Structure and Function of Bone, 1002
Composition, 1002
Functional Properties, 1003
Growth and Ossification, 1003
Continuous Growth, 1003
Bone Remodeling, 1004
Calcium Homeostasis, 1004
Response to Injury, Stress, and Aging, 1004
Fracture Healing, 1006
Structure and Function of Joints, 1006
Synarthroses, 1007
Fibrous Structure, 1007
Cartilaginous Structure, 1007
Diarthroses, 1007
Synovial Structure, 1009
Range of Movement, 1010
Structure and Function of Articular Cartilage, 1011
Composition, 1011
Functional Properties, 1011
Response to Injury, Stress, and Aging, 1012
Structure and Function of Tendons and Ligaments, 1012
Composition, 1012
Functional Properties, 1013
Response to Injury, Stress, and Aging, 1013
Structure and Function of Skeletal Muscle, 1013
Composition, 1014
Contractile Apparatus, 1014
Mechanics of Muscle Contraction, 1015
Sliding Filament Theory, 1015
Role of Calcium, 1015
Electromechanical Coupling, 1015
Types of Muscle Contraction, 1016
Twitch Contraction, 1016
Concentric, Eccentric, and Isometric Contractions, 1017
Mechanical Principles, 1017
Length–Tension Relationship, 1017
Load–Velocity Relationship, 1017
Force–Time Relationship, 1017
Effects of Temperature Change, 1017
Effects of Fatigue, 1017
Response to Movement and Exercise, 1017

Movement is one of the most characteristic and visible aspects of human life. Ease of movement adds to self-worth and well-being because the ability to move is closely connected to independence. A working knowledge of the system responsible for body movement is imperative to the provider. This chapter examines the basic characteristics of the firm support of bone and joint structures that make motion possible and the properties of skeletal muscles that are responsible for actually moving the body's framework.

STRUCTURE AND FUNCTION OF BONE

The primary purposes of the skeletal system are to protect internal organs, provide bony attachments for muscles and ligaments, present rigid levers to allow functional movement of the body and its separate parts, and store mineral and marrow elements for forming new blood cells. Bone is highly vascular and is metabolically active from birth until death.

Composition

Bone is composed of three main components: an organic matrix, an inorganic mineral content, and water. Accounting for approximately 22% to 25% of bone weight is an organic matrix (called *osteoid*) that is composed mostly of type I collagen fibers (about 94% of matrix). These collagen fibers, occurring in a triple helix arrangement, extend along parallel lines of tension and give bone its tensile strength and some flexibility. In addition, part of the organic matrix is a homogeneous ground substance composed of protein polysaccharides, particularly proteoglycan, which binds between collagen fibers and surrounds the bone cell types. These proteins may serve to transfer mechanical information within the matrix to bone cells.

The cellular component of the organic matrix includes predominantly osteoblasts, osteocytes, and osteoclasts. Osteoblasts, formed from osteoprogenitor (mesenchymal) cells that line bone surfaces, produce the organic matrix (osteoid), which is subsequently mineralized to form new bone. Osteoblasts communicate with each other in a network of cell extensions to exchange minerals, nutrients, and stimulatory signals. When an osteoblast becomes engulfed in its own calcified matrix, it becomes a mature osteocyte within the bone, but is still connected with other cells via its extended cell processes. Osteoclasts are multinucleated cells that migrate to the bone surfaces in response to certain stimuli and are responsible for bone resorption and mobilization of minerals.

The second and largest component of bone is the inorganic material (mineral salts such as calcium and phosphate), which accounts for approximately 65% to 67% of bone weight. This mineral content, bound and embedded within the matrix mostly in crystals of calcium hydroxyapatite, gives bone its hard, rigid structural strength while also serving as the body's main reservoir for calcium and phosphorus.

Approximately 10% of bone weight is from water located within the organic matrix surrounding collagen fibers and ground substance and within the canals that carry nutrition to bone tissues.

Microscopically, the basic unit of bone is the osteon or the haversian system (Fig. 50.1). The haversian canal lies at the center of each osteon and contains blood vessels and nerve fibers. A concentric series of

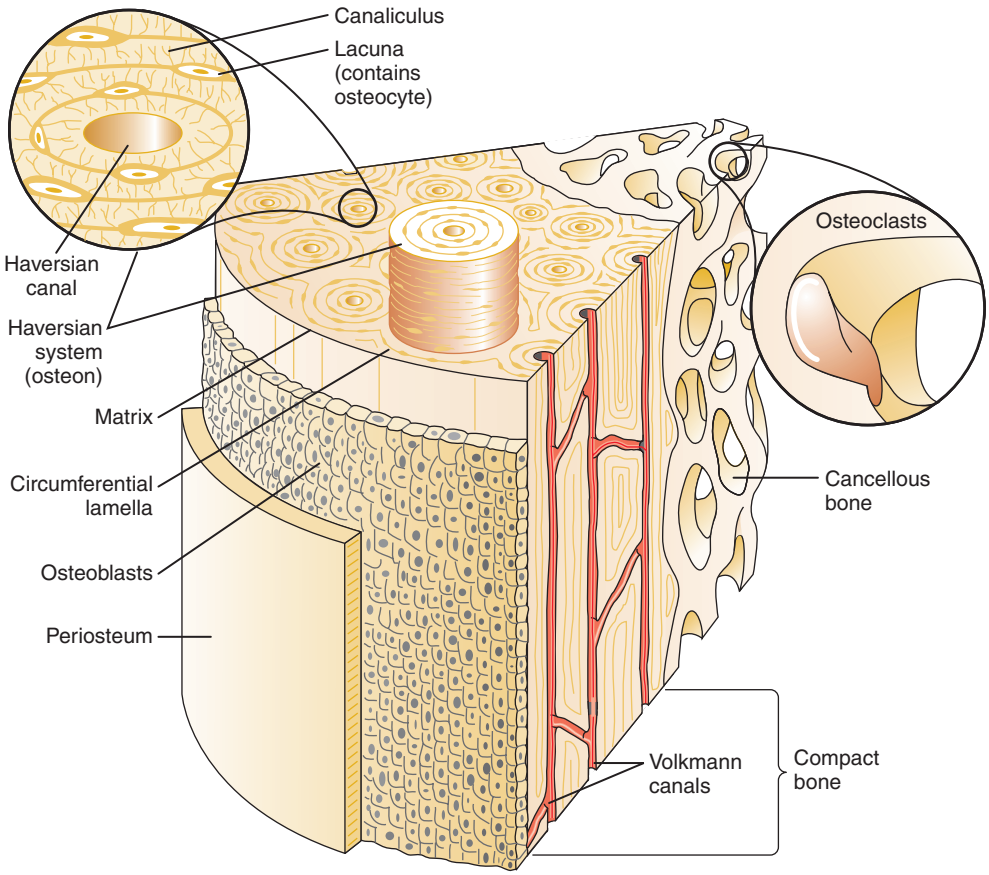


FIG 50.1 Microscopic anatomy of bone. The section has been enlarged to show the periosteum, osteoblasts, the haversian system, lacunae, and osteoclasts.

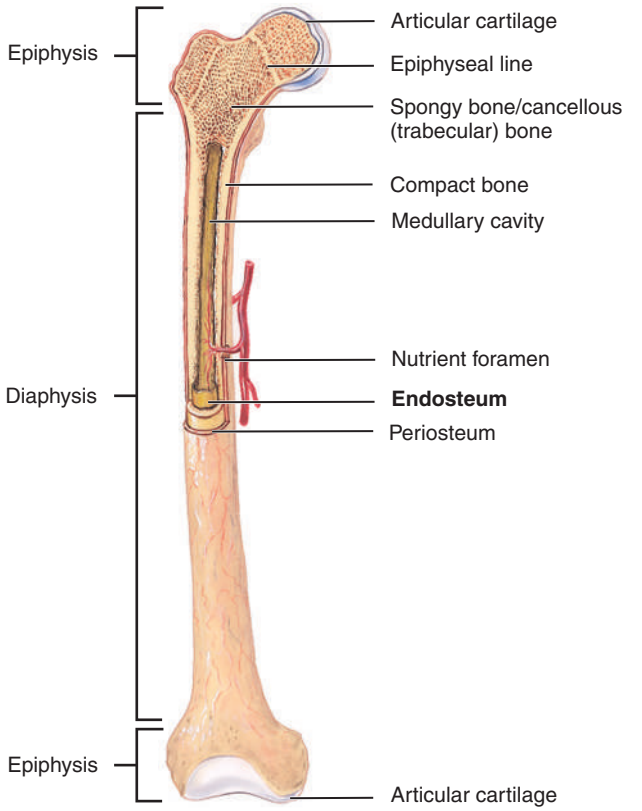


FIG 50.2 Structure and composition of a typical long bone. (From Applegate E: *The anatomy and physiology learning system,* ed 4, St Louis, 2011, Saunders.)

lamellae of mineralized matrix surrounds the central canal. Bordering the lamellae are small cavities (lacunae) that each contain a bone cell, the osteocyte. Many small channels, the canaliculi, connect adjacent lamellae with each other and eventually with the main haversian canal. This canal system allows nutrients from blood vessels in the haversian canal to reach osteocytes. Collagen fibers connect one lamella to another within the osteon and increase the mechanical strength of bone.

At the tissue level, bones are classified as two types: cancellous (also called *trabecular* or *spongy*) bone and compact or cortical bone (Fig. 50.2). Cancellous bone is a network of thin plates and rods called *trabeculae* found within the ends of long bones, vertebral bodies, and flat bones such as the pelvis. Trabeculae are laid down in response to stress and are shaped to accommodate loads placed on the bone. Cancellous bone is enclosed in a shell of compact bone. Compact bone is quite resistant to compression, is dense in structure, and is laid down in concentric layers. A tough fibrous membrane called the *periosteum* covers all bones. The periosteum is highly vascularized and provides nutrition for bone via Volkmann canals (see Fig. 50.1). An inner layer of the periosteum contains osteoblasts, which are responsible for bone growth and repair. The periosteum covers the entire bone except for the ends, which are covered by hyaline (articular) cartilage.

In longer bones, a central cavity (medullary cavity) is present (see Fig. 50.2). A thin membrane called the *endosteum* lines the inside of this cavity. In a child, the central cavity is filled with red marrow made of hematopoietic cells that produce red cells, white cells, and platelets. As a child ages into adulthood, the red marrow is largely replaced by fatty yellow marrow. In adults, red marrow is found predominantly in the central cavity of flat bones (pelvis, sternum.) Osteogenic cells are located in the endosteum.

Blood vessels are distributed through the haversian canals. Living cells in bone communicate with each other and the haversian system via threadlike processes.

Functional Properties

Growth and Ossification

Longitudinal bone growth involves a process called *endochondral ossification.* This process of growth is evident in embryonic development, fracture healing, and some bone tumor growth.

During embryonic development, for example, mesenchymal cells differentiate into chondrocytes, which produce a cartilaginous "model" of the bone, which is then mineralized starting at the center, the primary ossification center. As the primary center expands, secondary ossification centers form near the ends of long bones (epiphyses), which will then be sites of further growth of bone (at epiphyseal plates) even into teenage years (Fig. 50.3).

Flat bones (such as skull) grow by the circumferential growth process called *intramembranous ossification.* In this case vascularized fibrous tissue becomes directly mineralized as the mesenchymal cells present differentiate into bone-forming osteoblasts.

Continuous Growth

Interstitial growth is not possible within bone. Bone increases in length only through growth within a cartilage plate followed by endochondral ossification. The epiphyseal plate (see Fig. 50.3) allows for lengthening of the diaphysis (shaft) of a long bone and the metaphysis, which is the portion of widened bone between the diaphysis and the epiphysis (end of bone). The plate is a segment of proliferating cartilage located between the metaphysis and the epiphysis and is the site of continuous growth. The growth and thickening of cartilage cells of the plate move the epiphysis away from the metaphysis. Calcification and replacement of cartilage occur on the metaphyseal surface (endochondral ossification). Injuries to this growth plate in children may lead to limb length discrepancies due to bone overgrowth or skeletal arrest at these sites. In addition, inflammatory arthritis in children (particularly of the knee) can lead to increased blood flow to the epiphyseal plate, which also may accelerate growth or lead to premature plate closure. This also can lead to leg length abnormalities.

The function of the epiphyseal plate in the growth process may be illustrated by examining the specific zones of the plate (Fig. 50.4) and determining how they contribute to the growth process. The zone of resting cartilage maintains adherence of the plate to the epiphysis. Immature chondrocytes and vessels penetrate this first zone from the epiphysis and nourish the plate. The zone of young proliferating cartilage demonstrates the most active cartilage cell growth. The zone of maturing cartilage contains the enlarged and mature cartilage cells as they migrate toward the metaphysis. The final zone is the zone of calcifying cartilage, which is a very thin line of chondrocytes and the weakest segment of the epiphyseal plate. These chondrocytes are no longer active because of calcification of the matrix.

Bone is also deposited quite actively on the metaphyseal side of the plate. With the addition of new bone, the metaphysis becomes longer.

Osteoblasts in the inner layer of the periosteum are responsible for growth in the width of bones via intramembranous ossification. Resorption of bone by osteoclasts causes the medullary cavity to enlarge, causing additional widening of bone.

Hormones influence bone growth. Inadequate secretion of thyroxine by the thyroid gland or insufficient growth hormone secretion from the pituitary gland results in dwarfism. Oversecretion of growth hormone results in giantism. Sex hormones, such as estradiol, are produced in higher amounts during and after puberty and can cause more rapid maturation and fusion of the epiphyseal plates. These hormones may

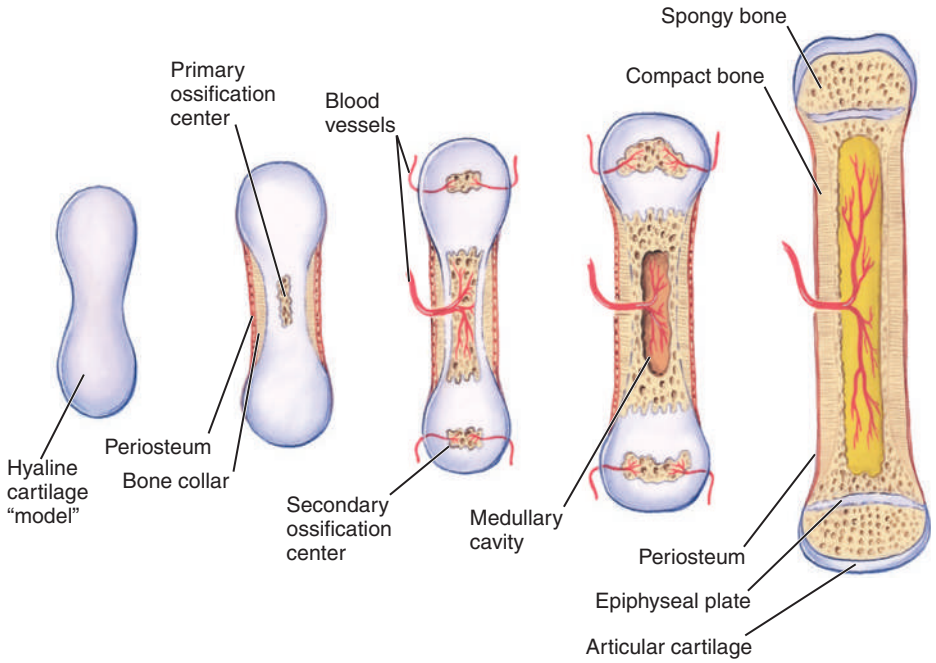


FIG 50.3 Events in endochondral ossification. (From Applegate E: *The anatomy and physiology learning system,* ed 4, St Louis, 2011, Saunders.)

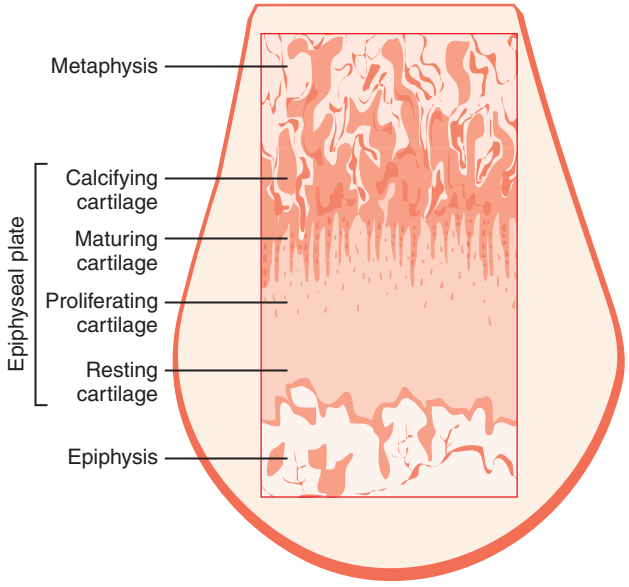


FIG 50.4 Zones of the epiphyseal plate.

limit the growth spurts of puberty, and early sexual maturity, especially in girls, can lead to shorter stature.

Bone Remodeling

Even in adulthood, bone is in a continuous state of turnover, a process that begins very early in skeletal development. This occurs in cycles of bone resorption and new bone formation called *remodeling*, with each complete cycle termed a *bone remodeling unit* (or *bone multicellular unit*). In a normal state, bone formation and resorption are closely coupled and generally balanced, serving to maintain the skeleton's peak strength by removal and repair of damaged areas, which contributes to osteocyte viability and affects the body's calcium homeostasis.

A bone remodeling unit occurs in a sequence of phases: stimulation and recruitment of osteoclasts to the area, resorption of bone, reversal and stimulation of osteoblasts, secretion of new matrix, mineralization of new bone, and completion of a resting phase. The stimulation of a remodeling cycle is thought to occur with osteoblastic production of factors that stimulate the recruitment and activity of osteoclasts (such as osteoclast differentiating factor, also called *receptor activator of nuclear factor κB ligand*). One such signal of the need for bone remodeling could come from mature osteocytes within the bone sensing bone deformation or damage. Osteocytes also secrete macrophage colony stimulating factor, which contributes to osteoclast differentiation and migration. Osteoclasts gather at the site on the bone surface and form large multinucleated cells that begin the process of removal of bone matrix and minerals. This resorptive phase creates a pit in the bone, which is next filled with osteoblasts that begin to fill the space with bone matrix (osteoid). When the bone pit is filled, the new bone matrix is mineralized to complete the cycle (Fig. 50.5).

Calcium Homeostasis

Almost all of the body's calcium supply is held within bone, but the small amount of calcium circulating in blood is essential for a wide variety of cellular functions, and therefore its concentration is tightly controlled. Much of this regulation of calcium balance in bone and blood depends on hormonal effects. Parathyroid hormone maintains serum calcium levels by increasing bone resorption as well as calcium reabsorption from renal tubules. Vitamin D metabolites can increase bone mineralization by increasing calcium absorption from the intestinal tract; however, in the setting of calcium deficiency, vitamin D can stimulate bone resorption to help maintain mineral supply in the blood. Calcitonin can act as an inhibitor of bone resorption, but likely only plays a minor role in adults. (See Chapter 51 for further discussion.)

Response to Injury, Stress, and Aging

The ability of bone to remodel after injury is important. Although remodeling of bone continues throughout life, death of the osteon or removal of calcium from bone requires that new bone be deposited to

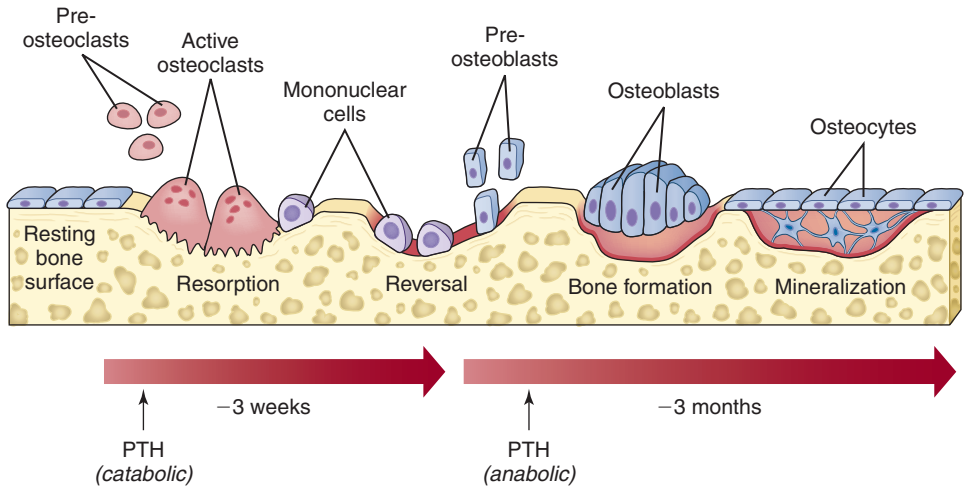


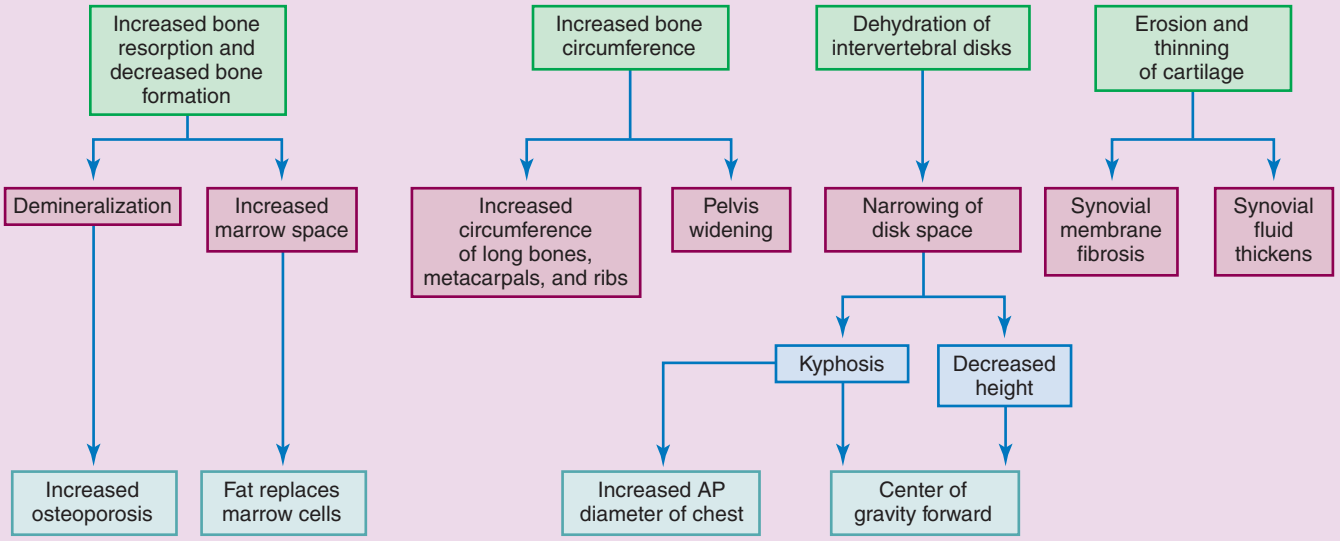
FIG 50.5 Bone remodeling cycle. First, osteoclasts begin bone resorption, creating the pit. Then osteoblasts enter the pit and produce bone matrix, which then calcifies. *PTH,* Parathyroid hormone. (From Garg AK: *Implant dentistry: a practical approach,* St Louis, 2010, Mosby.)

retain strength and function. Physical stresses lead to the realignment of bone trabecular systems and the deposition of additional bone at the site of increased stress. The response of bone to stress is summarized by Wolff's law, which states that bone is enhanced where it is needed and resorbed where it is not needed. If bone is immobilized or not subjected to mechanical stress, as occurs with prolonged bed rest, the activity of bone-resorbing cells increases. Without external forces (or loads), osteoclast activity is greater than osteoblast activity and bone mass decreases. It is probable that during bed rest, age-related bone loss might be temporarily accelerated and may result in a greater decline in bone mass over time. Patients becoming mobile after prolonged bed rest are at risk for fractures because of a combined loss of muscle and bone strength. With loss of muscle, gait becomes unsteady and patients are more prone to falls (see "Geriatric Considerations: Changes in the Skeletal System").

Internal fixation of a fracture may also cause decreased bone strength. With metal implants, mechanical stress is dispersed from bone and carried by the implant. Bone under the plate is resorbed, and "stress-relief"

GERIATRIC CONSIDERATIONS

Changes in the Skeletal System



With aging, bone resorption exceeds bone formation. There is a net loss of bone mass and bone protein matrix. The interior of the long and the flat bones is absorbed faster than that of other bones. Trabecular bone destruction is greater than cortical bone loss. Compared with aging men, postmenopausal women have a higher risk of bone loss and subsequent osteoporosis, although men can develop osteoporosis as well. Age-related bone loss begins at approximately 30 to 50 years of age in both men and women.

Studies have shown that elderly individuals express higher levels of certain markers associated with bone resorption, whereas bone formation markers are much more variable. One common cause of increased bone resorption is calcium

Continued

GERIATRIC CONSIDERATIONS

Changes in the Skeletal System—cont'd

and vitamin D deficiency, which causes more rapid mobilization of calcium from bone. A secondary hyperparathyroidism can also result. Decreased levels of estrogen in elderly women and men can contribute to age-related bone loss because osteoblasts have estrogen receptors and their ability to increase bone formation may be affected by the estrogen deficiency. An increase in the local production of cytokines that influence bone resorption may also occur with decreased estrogen levels. The end result is an imbalance between osteoblast and osteoclast function and progressive decline in bone mass (see further discussion on osteoporosis in Chapter 51).

Although interior bone is lost, the circumference of the bones increases because osteoblasts on the exterior bone beneath the periosteum continue bone formation. The long bones, metacarpals, and ribs become bigger in circumference, whereas the pelvis becomes wider and the skull thicker.

The intervertebral cartilage disks become dehydrated, with narrowing of the disk space leading to a decrease in height of 3 to 5 cm. Disk thinning along with compression of osteoporotic vertebrae of the spine can result in an increase in the thoracic curve, resulting in kyphosis and anterior scapular displacement. This change leads to an increase in the anteroposterior diameter of the chest. A decrease in the lordotic curve results in lumbar straightening and a decrease in lumbar flexibility. Greater flexion of the knees and hips is noted. The relationship between the pelvis and the femoral head and neck also changes.

Fissuring, erosion, and thinning of cartilage occur. With the loss of cartilage, the greater pressure that subchondral bone must withstand results in increased density and the formation of joint margin osteophytes. The synovial membrane undergoes fibrosis and the synovial fluid thickens.

Bone mass can also decrease prematurely with certain disease processes. For example, in osteoporosis, a metabolic bone disease, there is a severe general reduction in skeletal bone mass and thus a susceptibility to fractures. In short, bone resorption is more rapid than bone formation.

osteoporosis may occur. Care must be taken once implants are removed, and the bone must be protected until strength returns. Some implants are designed to compress fracture fragments to aid healing.

Fracture Healing

Bone may heal in one of two ways after a fracture. A periosteal or external callus forms in fractures managed by closed methods. The blood supply to surrounding soft tissue and motion at the fracture site contribute to healing. Medullary callus formation takes place with rigid immobilization at the fracture site. The process of bone turnover contributes to healing.

The five stages of fracture healing are (1) hematoma formation, 1 to 3 days; (2) fibrocartilage formation, 3 days to 2 weeks; (3) callus formation, 2 to 6 weeks; (4) ossification, 3 weeks to 6 months; and (5) consolidation/remodeling, 6 weeks to 1 year (Fig. 50.6). These five stages can be grouped into three phases: (1) inflammatory phase, (2) reparative phase (stages 2 to 4), and (3) remodeling phase.

Stage 1 begins when a hematoma forms at the fracture site. The size of the hematoma depends on the amount of damage at the fracture site. The hematoma offers some stability to fractured ends. Aseptic inflammation occurs at the fracture site.

Healing continues during *stage 2* with the formation of granular tissue containing blood vessels, fibroblasts, and osteoblasts. The hematoma provides the foundation for reparative tissue and bone healing. Vascular and mechanical factors such as motion and distraction of fragments influence stage.

Cartilaginous callus formation occurs during *stage 3* after the granulation tissue matures. There is formation of bone at the periosteal surfaces to unite the outer ends of the bone fragments to stabilize the fracture. If this stage is delayed or interrupted, the final stages cannot occur.

Stage 4, or ossification, occurs as the space in the bone is bridged and the fractured ends are united. The callus is slowly replaced by trabecular bone along the lines of stress, and unnecessary callus is reabsorbed.

During *stage 5,* consolidation and remodeling occur as the medullary canal is reestablished. Bone is resorbed and deposited along stress lines as bone reshapes to meet its mechanical requirements.

Fractures are usually considered healed when clinical healing is achieved. Clinical healing occurs when the fracture is stable and strong enough to resume its function, the fracture site is free of pain, no gross movement is seen across the fracture site, and radiographs show bone crossing the fracture site.

KEY POINTS

- Bone is capable of altering its shape and density in response to mechanical demands.
- The osteon is the basic unit of bone.
- Bone tissue may be dense and compact (cortical) or lighter and trabecular (cancellous).
- In long bones, the epiphyseal plate is the site of linear growth. Increases in bone width are mediated by osteocytes in the periosteum.
- Bone cells responsible for deposition are called *osteoblasts;* osteoclasts mediate bone resorption. The balanced coupling of bone resorption and new bone formation is called *remodeling.*
- Absence of bone stress because of immobility or altered weight bearing leads to demineralization.

STRUCTURE AND FUNCTION OF JOINTS

Coordinated movement is only possible because of joint, bone, and muscle structures. Joints permit complex, highly coordinated, and purposeful movements. A joint, also called an *articulation,* is a point of contact between bones. Functional articulations between bones in extremities such as the shoulder, elbow, hip, and knee contribute to controlled and graceful movement.

The type and configuration of a joint depend on the functional demands placed on that joint. As is the case with all aspects of the musculoskeletal system, structure determines function (Fig. 50.7). When considering the human joint, or articulation, it is also important to remember that once the articulation has developed, the configuration of the joint surface will determine the movement of the joint. Any aberrant joint movement has the potential to disrupt function and cause a breakdown in joint integrity.

Articulations can provide more than a single motion, such as flexion and extension. Flexion, extension, adduction, abduction, rotation, opposition, and circumduction may all be functional movements of a joint. The more complex the movements, the more complex is the joint structure.

Broadly speaking, articulations, or arthroses, in the human body may be divided into two categories based on the composition of the

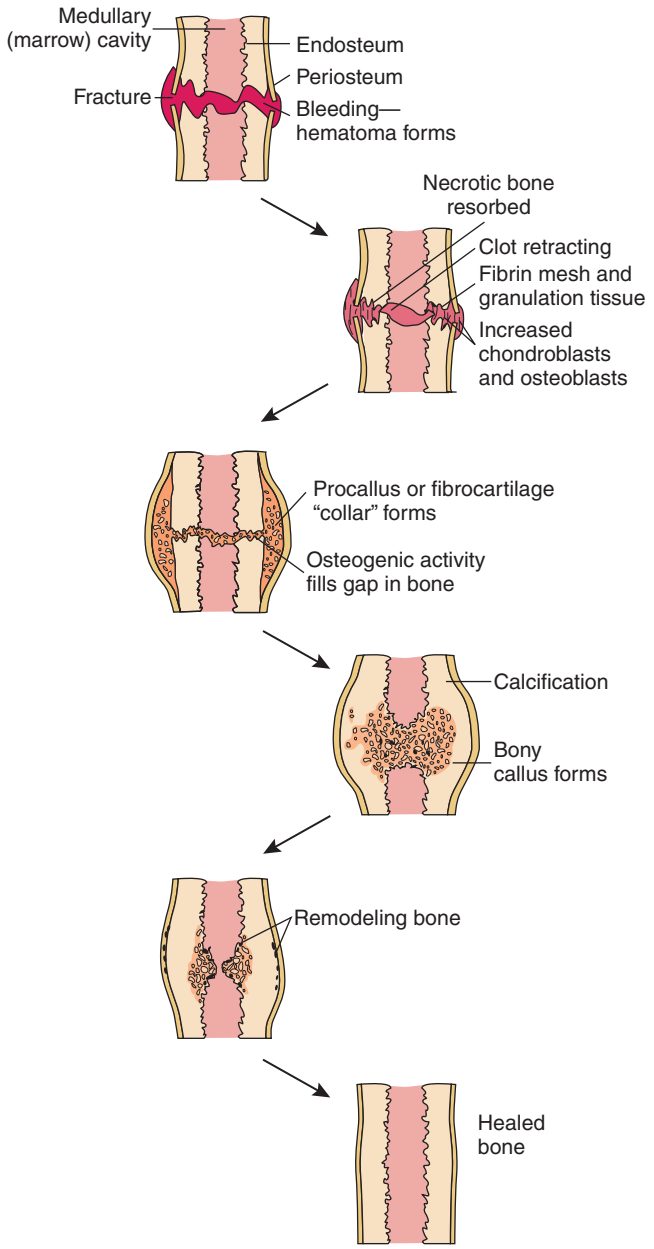


FIG 50.6 Healing of a fracture. (From Gould BE, Dyer R: *Pathophysiology for the health professions,* ed 4, St Louis, 2011, Saunders.)

joint and the method in which the joints unite the body components. The two categories are synarthroses, or fibrous and cartilaginous (nonsynovial) joints, and diarthroses, or synovial joints.

Synarthroses

Synarthroses have two subdivisions based on the type of connective tissue used to form the joint. Fibrous and cartilaginous tissues give these joints their names.

Fibrous Structure

In a fibrous joint, bones are united by fibrous tissue. Three types of fibrous joints are found in the human body: suture joints, gomphosis joints, and syndesmosis joints. A suture joint unites bones with a thin but dense layer of fibrous tissue. Interlocking bony ends overlap and increase stability. Suture joints are found only in the skull (Fig. 50.8). Fusion of the joint occurs later in life. This bony union is called a *synostosis.*

The joint that is found between a tooth and the mandible or maxilla is the only gomphosis joint in the human body. The best description of a gomphosis joint is that of a peg implanted into a hole. Fibrous tissue stabilizes the two bony structures and permits little movement.

A syndesmosis joint is a joint in which the two bony components are joined by a ligament or interosseous membrane. These joints normally allow slight movement and are quite functional. The interosseous membrane joining the fibula and the tibia is an example of a syndesmosis joint (Fig. 50.9).

Cartilaginous Structure

Bony segments connected by fibrocartilage or hyaline growth cartilage are classified as cartilaginous joints. Synchondrosis joints and symphysis joints are the two types of cartilaginous joints in the body.

In a synchondrosis joint, cartilage connects bony components. This joint allows bone growth while providing stability. This type of joint can be found at growth sites of the body. The first sternocostal joint is an example of a synchondrosis joint (Fig. 50.10). When bone growth is complete, these joints ossify and become unions (synostoses).

A symphysis joint connects bony segments by a fibrocartilaginous plate or disk, examples being the intervertebral joints of the spine and the symphysis pubis joint (Fig. 50.11), which joins the two pubic bones of the pelvis. This joint is a weight-bearing structure and is important in transmitting stress and providing stability. Only slight motion is permitted. Intervertebral joints comprise the predominant articulations in the spine. These joints are stabilized by the intervertebral disks, which are padlike structures between vertebral bones that help bind vertebrae together and act as shock absorbers between adjacent vertebrae. These disks allow slight movement between any two adjacent vertebral bodies. Disks contribute to the natural curves of the spine in the cervical and lower lumbar areas.

Each intervertebral disk consists of an outer annulus fibrosus, or outer fibrous layer, and a nucleus pulposus, or soft center. The annulus fibrosus consists of many layers of fibrous tissue and fibrocartilage that are strongly attached to the ends of the bodies adjoining the disks.

The nucleus pulposus is semigelatinous, containing a high percentage of water, and is located closer to the posterior edge of the disk. Because of its high water content, the intervertebral disk is prone to dehydration. Even when the vertebral column is not supporting the weight of the body, as in the supine position, intervertebral disks are maintained under pressure by ligaments connecting the arches.

Although the nucleus pulposus is not compressible, its softness allows it to change shape easily. As the vertebral column bends, the nucleus pulposus becomes wedge shaped, with the thin edge in the direction of bending. The annulus fibrosus on this side bulges out and on the opposite side is stretched by its attachment to the adjoining vertebrae.

Pain caused by the pressure of a protruded disk on a nerve root or spinal nerve leads to pain in the area innervated by compressed nerve fibers and is called *radicular pain.*

Diarthroses

Joints designed to allow mobility are classified as diarthroses, or synovial joints. These joints are covered with a fibrous joint capsule, and the joint interior is lined with a thin synovial lining layer. Movement in these joints is provided by contraction of the muscle–tendon unit, and control depends on the joint capsule and ligaments. Stability of the synovial joint is enhanced by additional soft tissue structures—the menisci, articular cartilage, and labra. Synovial fluid is produced by fibroblast-like cells of the synovial lining layer and is secreted into mobile

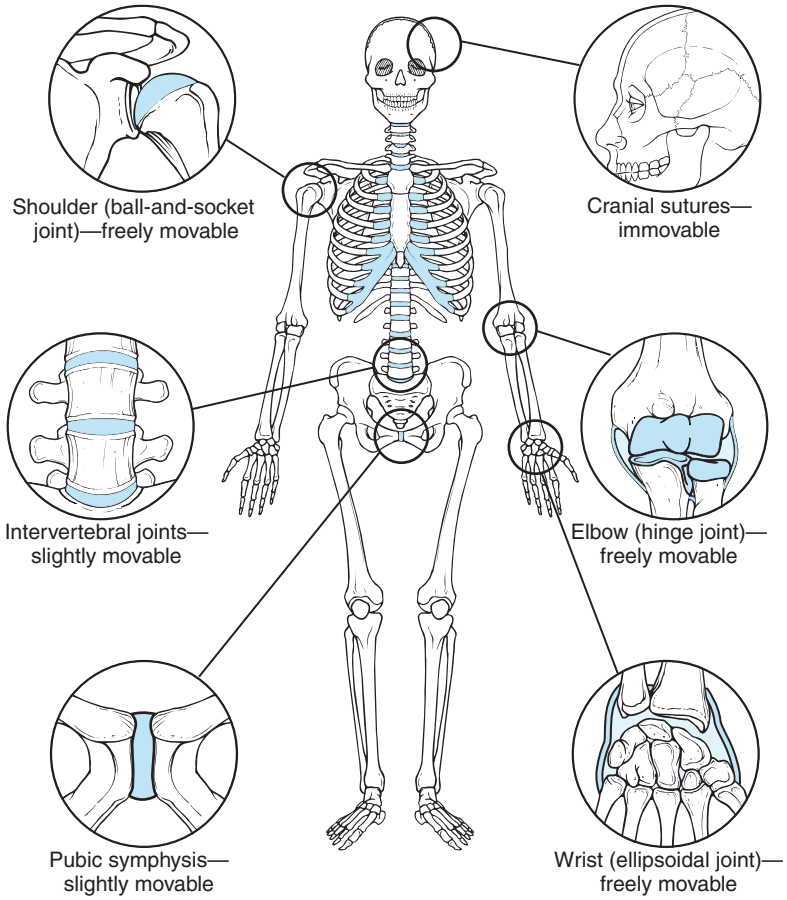


FIG 50.7 Examples of types of joints. (From Frazier MS, Drzymkowski JW: *Essentials of human diseases and conditions,* ed 5, Philadelphia, 2013, Saunders, p 302.)

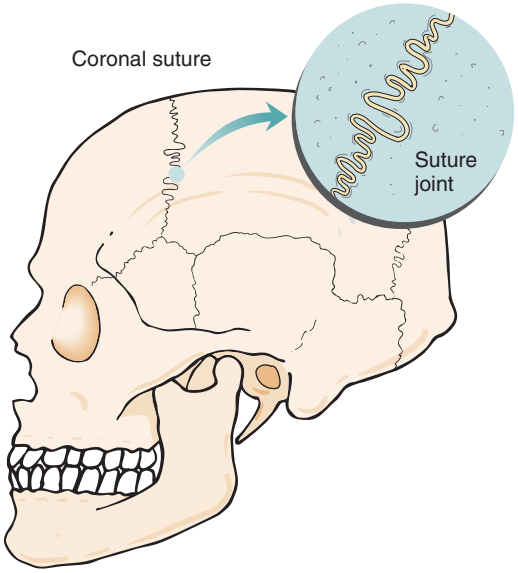


FIG 50.8 A suture joint is found only in the skull.

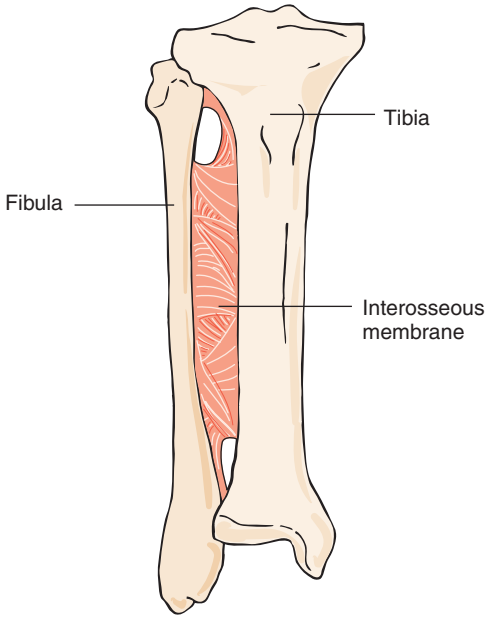


FIG 50.9 The interosseous membrane joining the fibula and the tibia is an example of a syndesmosis joint.

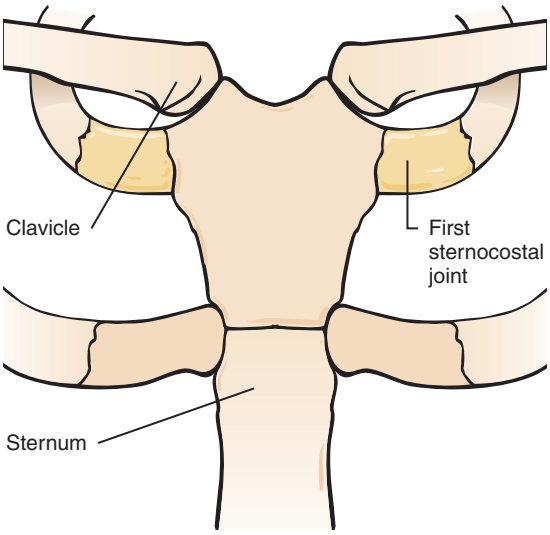


FIG 50.10 First sternocostal joint.

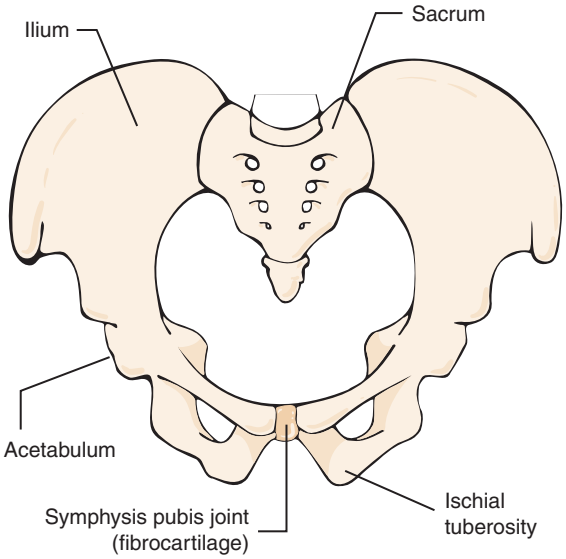


FIG 50.11 Symphysis pubis joint.

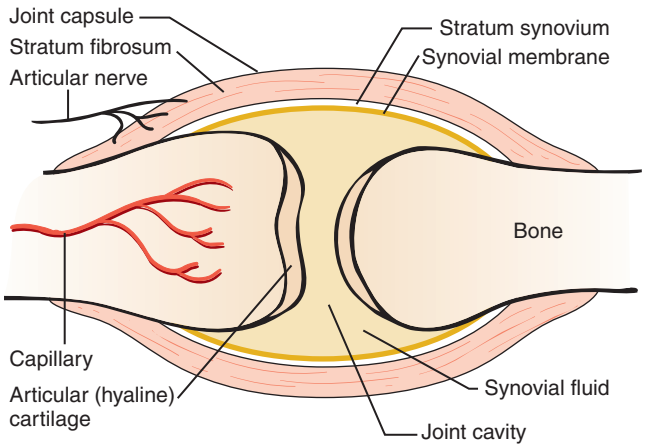


FIG 50.12 Typical synovial joint.

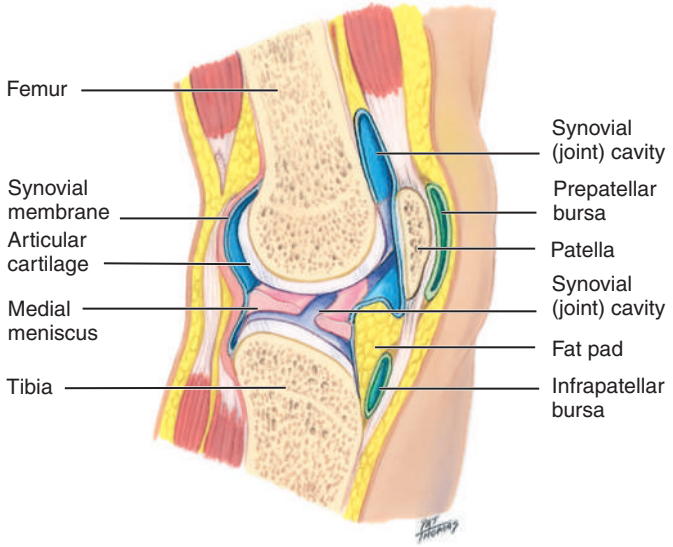


FIG 50.13 Schematic drawing of a typical diarthrodial (synovial) joint. (From Applegate E: *The anatomy and physiology learning system,* ed 4, St Louis, 2011, Saunders.)

joints to provide the lubrication necessary to reduce friction between articulating surfaces. In diarthrodial, or synovial, joints, the bony ends are free to move because no cartilaginous tissue connects the adjacent bony surfaces. The synovial joint connects adjacent bony surfaces through a joint capsule that surrounds the joint.

Synovial Structure

Features common to all synovial joints include (1) a fibrous joint capsule, (2) a joint cavity enclosed by a joint capsule, (3) a synovial membrane that lines the inner surface of the capsule, (4) lubricating synovial fluid that coats joint surfaces, and (5) articular cartilage, which covers the opposing joint surfaces (Fig. 50.12).

Many synovial joints also have accessory structures within the joint capsule. Ligaments, fat pads, disks, and menisci are a few of the structures situated in the capsule that are important to proper function of the joint. Ligaments and tendons keep joint surfaces together and aid in joint motion. Menisci, disks, and synovial fluid limit excessive compression of articulating surfaces.

The lateral and medial menisci of the knee are located on top of the tibia between the tibia and femur (Fig. 50.13). These semilunar fibrocartilaginous structures function as shock absorbers in the knee. In cross-section, these wedge-shaped cartilages are thinnest on the inner edge. Around the inner edge, the area of the synovial cavity between a femoral condyle and a meniscus is continuous with that between the meniscus and corresponding tibial condyle. On the outer edge, the cartilages are attached to both the synovial and the reticular capsule. The medial meniscus is firmly attached to the collateral ligament. The lateral meniscus has weak attachments to the lateral area of the capsule, from which it is in part separated by the tendon of the popliteal muscle. It is possible that because the medial cartilage is more firmly attached, it is torn more often than the lateral meniscus, which has no attachment to the fibular collateral ligament and is thus more mobile. Both menisci are anchored to the tibia via strong fibrous bands.

Menisci facilitate rotation at the knee by allowing better contact of the tibial surfaces with the femoral condyles. They function to evenly distribute load bearing on the tibial plateau. Menisci are often torn by

rotation of the femur when the knee is flexed. The torn portion of the meniscus locks the joint, with accompanying pain and edema in the knee. If torn, the menisci can be removed; however, weight-bearing areas on the femur and tibia may then decrease by almost 50%.

The joint capsule is composed of two layers of connective tissue. The outer layer is the fibrous membrane composed of collagenous tissue. It is dense and encapsulates the entire joint. This dense tissue is solidly attached to the periosteum of the adjacent bony components. The fibrous membrane is poorly vascularized and innervated by joint nerve receptors. Joint receptors are able to detect motion, compression, tension, vibration, proprioception, and pain. Pain can be caused by swelling and stretching of the capsule (as in arthritis or infection) or by injury to the ligaments (as in a strain). Articular cartilage has no nerve fibers. A general rule notes that a joint is innervated by the major nerves that cross it.

The inner layer, or synovial membrane (also called *synovial lining layer*), is highly vascularized and often only two or three cell layers thick. It is innervated by very small, slowly conducting nerve fibers that follow the course of blood vessels supplying the lining layer. Stimulation of these fibers may cause a diffuse burning or dull aching sensation. The specialized cells in the synovial membrane, called *synoviocytes,* are of two types: macrophage-like cells (10% to 20%) and fibroblast-like cells (the majority). The macrophage-like cells are capable of clearing waste material from the joint cavity via phagocytosis. The fibroblast-like synoviocytes function to synthesize collagen, extracellular matrix proteoglycans, and other components of synovial fluid.

Synovial fluid. Synovial fluid contains hyaluronic acid, a high-molecular-weight polysaccharide, and lubricin, a glycoprotein. Hyaluronic acid provides for viscosity and reduces friction between the capsule and joint surfaces. It also helps to maintain synovial fluid volume by slowing diffusion of water out of the joint space. Lubricin is an important lubricant of cartilage and articular surfaces. Synovial fluid resists shear loads, keeps surfaces lubricated to reduce friction, and provides nourishment for cartilage. Although synovial fluid is generally maintained at a constant volume, disease states, such as inflammatory arthritis or infection, can stimulate increased synovial fluid production by synoviocytes. When the accumulation of fluid outweighs its clearance, joint swelling results.

Range of Movement

Synovial joints can be divided into three main categories according to visible movement allowed at the joint: uniaxial, biaxial, and triaxial.

A uniaxial joint allows motion around a single axis of movement. Two types of uniaxial diarthrodial joints are hinge joints and pivot joints. A hinge, or ginglymus, joint permits flexion and extension; an example is the interphalangeal joint of the finger, the elbow, or the knee (Fig. 50.14A). A pivot, or trochoid, joint allows rotation as its single axis movement. The superior radioulnar joint of the elbow and the union between the first and second vertebrae are examples of a pivot joint (Fig. 50.14B).

A biaxial joint has two axes of movement and permits movement in two planes. Two kinds of biaxial joints are condyloid joints and saddle joints. The metacarpophalangeal joint of the hand is an example of a condyloid joint; it permits flexion and extension at one axis and adduction and abduction around another axis (Fig. 50.15, parts A and B, respectively). A saddle, or sellar, joint is a joint in which the surfaces are convex in one plane and concave in the other. The surfaces of a saddle joint fit together as a saddle fits a horse. The carpometacarpal joint of the thumb is a saddle joint; it permits both flexion-extension and adduction-abduction movements (Fig. 50.15C).

Triaxial joints permit movement around three axes so that motion can occur in three planes. A triaxial joint permits gliding movement between two bones and is exemplified by the carpal joints of the hand. The carpal joints may glide or rotate relative to the adjacent surfaces. A ball-and-socket joint is formed by a ball-like surface fitting into a concave socket. Ball-and-socket joints permit flexion-extension, adduction-abduction, and rotational movements. The hip and shoulder are examples of a ball-and-socket joint (Fig. 50.16).

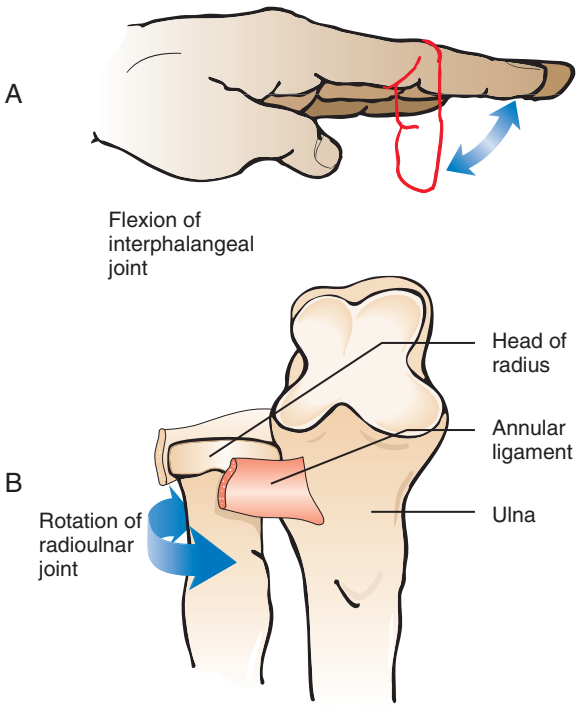


FIG 50.14 A hinge joint permits flexion and extension and is represented by the interphalangeal joint of the finger **(A).** A pivot joint allows rotation and is represented by the superior radioulnar joint of the elbow **(B).** Both the hinge joint and the pivot joint are considered uniaxial joints because they allow motion around a single axis.

KEY POINTS

- Joint configuration dictates possible motions of a joint. Types of joint movements include flexion, extension, adduction, abduction, and rotation. Joints that allow these types of movements are called *diarthroses* (synovial joints). The ends of bone in a synovial joint are held together by a joint capsule composed of two layers of connective tissue.
- Intervertebral disks are padlike structures that act as cushions between vertebrae. A strong annulus fibrosus surrounds a gelatinous, high-water-content nucleus pulposus that can herniate and press on spinal nerves.
- The lateral and medial menisci in the knee serve as shock absorbers between the femur and tibia. The medial meniscus has strong attachments to the collateral ligaments, whereas the lateral meniscus has weak attachments to the lateral area of the joint capsule. Thus because of its strong attachment, the medial meniscus is more likely to be torn than the lateral meniscus.
- The joint capsule is composed of two layers of connective tissue: an outer fibrous membrane and an inner synovial membrane.
- Synovial fluid provides nourishment and lubrication for cartilage. It becomes more viscous with slow movement and low temperatures and less viscous with fast joint movement and high temperatures.

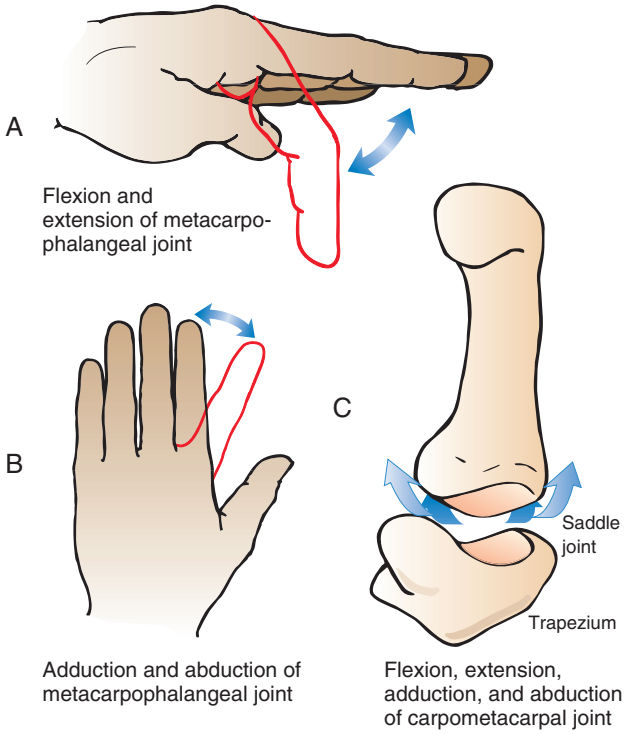


FIG 50.15 A condyloid joint permits flexion and extension at one axis and adduction and abduction around another axis; it is represented by the metacarpophalangeal joint of the hand **(A** and **B).** Because of its convex and concave surfaces, a saddle joint allows for flexion and extension, as well as adduction and abduction; it is represented by the carpometacarpal joint of the thumb **(C).** Both the condyloid joint and the saddle joint are considered biaxial joints because they have two axes of movement and permit movement in two planes.

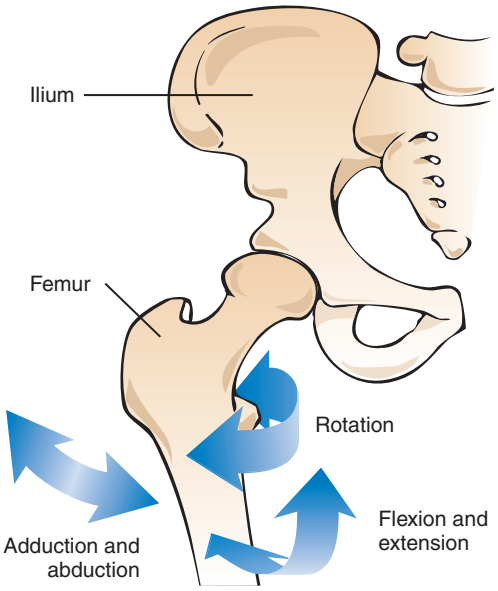


FIG 50.16 A ball-and-socket joint permits flexion and extension, adduction and abduction, and rotation; it is represented by the hip joint. A ball-and-socket joint is considered a triaxial joint because it permits movement around three axes; motion can occur in three planes.

- Synovial joints are classified according to the visible movements that they allow:
  - Uniaxial: Movement in one plane only (e.g., distal hinge joints of the fingers)
  - Biaxial: Movement in two planes (e.g., thumb saddle joint)
  - Triaxial: Movement in three planes (e.g., ball-and-socket hip joint)
- Some bones are held together by joints that allow little or no movement. These joints are called *synarthroses* (nonsynovial joints). Examples include sutures between skull bones, tooth–jawbone joints, and the symphysis pubis joint.

STRUCTURE AND FUNCTION OF ARTICULAR CARTILAGE

Articular cartilage appears smooth, shiny, and white on gross inspection. It is a specialized tissue designed to withstand stress imposed by the movement of bony structures. Articular (hyaline) cartilage covers the ends of bone. It functions to distribute joint loads over a wide area, to decrease the stress of prolonged compression from contacting joint surfaces, and to allow movement of joint surfaces with minimal friction and deterioration. Articular cartilage is devoid of blood vessels, lymph channels, and nerves. If a mechanical defect is present, however, this avascular structure can cause major disruption of joint movement.

Composition

Cartilage is hydrophilic in nature, with more than 70% of it being primarily water with some inorganic salts, proteins, glycoproteins, and lipids making up the interstitial fluid.

Chondrocytes comprise the main cell type within hyaline cartilage. These cells are sparsely distributed and serve to manufacture the organic components of the extracellular matrix. This extracellular matrix is composed of an intricate network of collagen fibrils, elastin, fibrillin, and other macromolecules. The collagen fibrillary network gives cartilage its flexibility and tensile strength.

Surrounding and interlocking this network is the proteoglycan component (especially aggregan), which provides cartilage with some stiffness but also allows the potential to be compressed by absorbing or extruding water with pressure. Articular cartilage is primarily avascular and is very limited in its ability to regenerate and repair itself.

Functional Properties

Articular cartilage has a biomechanical function. It spreads loads applied to articulating bone ends over a large area to decrease contact stress and limit wear and friction in the joint during movement.

Collagen fibers in articular cartilage are highly structured to provide stability (Fig. 50.17). The most important mechanical properties of collagen fibers are strength and tensile stiffness. By themselves, collagen fibrils tolerate tension but not compression.

To improve tolerance to compression, cartilage proteoglycan works with hyaluronate to form proteoglycan aggregates. This proteoglycan aggregation fosters immobilization of the proteoglycans within the collagen meshwork, which adds structural rigidity and better compression tolerance to the extracellular matrix.

The importance of proteoglycans and interaction with collagen does not end with an increase in tolerance to compression. Proteoglycans also associate with collagen as a bonding agent to stabilize cross-links between collagen fibers. By maintaining ordered structure and the mechanical properties of collagen fibers, proteoglycans assist in increasing strength.

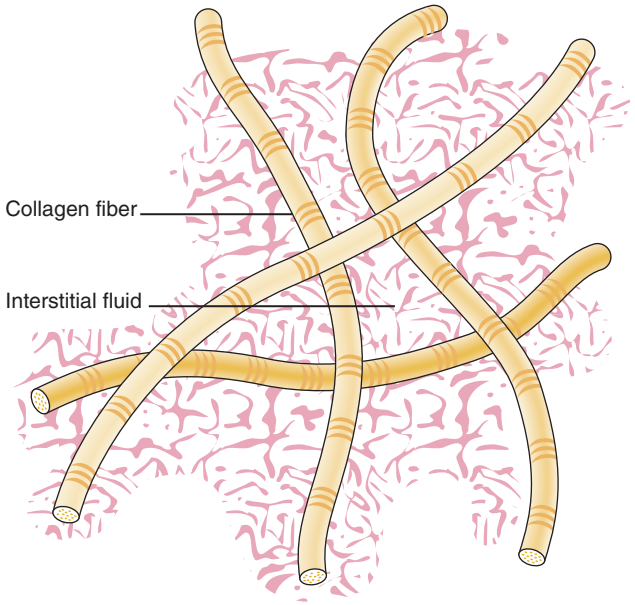


FIG 50.17 Collagen fiber in articular cartilage.

Articular cartilage requires a sophisticated lubrication process to ensure a decrease in friction between joint surfaces. Without correct lubrication by synovial fluid, articular cartilage will begin to break down as a result of mechanical action of the joint.

Joints are lubricated by two methods. One method is by the mechanics of joint physiology. A lubricating coating is formed between the joint surfaces when a weight-bearing force or load is applied to the cartilage and fluid is abstracted from the matrix. The movement of fluid under pressure acts as a self-lubricating mechanism. When the load is removed, liquid from the matrix is reabsorbed by the cartilage. The second method is assisted by glycoproteins covering cartilage and providing a lubricated surface. Lubrication of cartilage from a combination of these two methods decreases friction in the joint. Weight bearing and joint motion are essential for healthy cartilage. Cartilage will atrophy if joints are not used because cells cannot be nourished by the synovial fluid.

Response to Injury, Stress, and Aging

Articular cartilage can experience wear. This gradual degeneration can occur by the removal of material from solid surfaces by biomechanical action. Articular cartilage may begin to wear through two primary mechanisms: interfacial wear and fatigue wear. Interfacial wear results from the interaction of weight-bearing surfaces by either adhesive or abrasive action. Interfacial wear occurs when joint surfaces come into direct contact as a result of insufficient lubricating film. The nonlubricated surfaces are quite abrasive to each other, and joint surfaces may deteriorate. Fatigue wear results from repeated deformation secondary to weight bearing. Fatigue wear occurs as a result of the accumulation of microscopic injuries from repeated stress.

Because of the changes in the nature of glycoproteins with aging, cartilage becomes less able to retain water. This "drying out" effect can change the biomechanics of cartilage and lead to increased stress fractures or cracks in the collagen network. Over time and with joint wear, microcracks can accumulate and fragments of cartilage can detach into the joint space, creating "loose bodies." The resulting cartilage surface is rough and irregular and subject to further mechanical wear and degeneration. This process can form the basis of osteoarthritis or degenerative joint disease. The effects of aging on the skeletal system are described in "Geriatric Considerations: Changes in the Skeletal System."

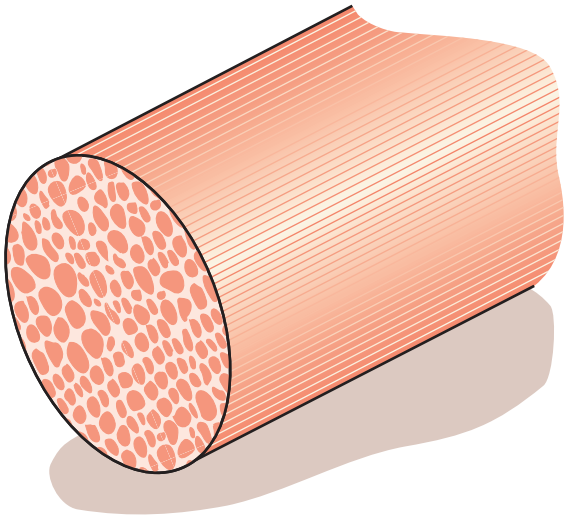


FIG 50.18 Parallel bundles of collagen fiber in tendons.

KEY POINTS

- The ends of bones are covered with articular cartilage, which helps distribute mechanical loads placed on the joint and minimize friction and wear.
- An important component of articular cartilage is collagen, which provides strength and tensile stiffness. A second component, proteoglycan, increases compression tolerance.
- Articular cartilage is avascular and relies on synovial fluid for nutrition and waste removal.
- Synovial fluid lubricates articular surfaces to reduce friction and minimize wear.
- Articular cartilage has limited capacity for repair and regeneration.
- Interfacial joint wear occurs because of insufficient lubrication.
- Fatigue joint wear occurs because of repetitive stress injuries.
- Sudden imposition of excessive stress may also cause trauma to the joint matrix.

STRUCTURE AND FUNCTION OF TENDONS AND LIGAMENTS

Approximately 200 bones in the human skeleton are connected by joints that provide movement and dynamic stability. Ligaments, tendons, and joint capsules provide joint stability but not movement because they are not contractile structures. Without joint stability, no movement of the limbs would be possible.

Ligaments and joint capsules connect bone to bone, provide mechanical stability to joints, and guide joint motion. Tendons, through attachment to a contractile structure (muscle) and a rigid object (bone), assist in the generation of movement. Injuries to ligaments and tendons are common, so an understanding of their function and properties is important.

Composition

Tendons and ligaments are dense connective tissue in which collagen fibers are positioned in generally parallel alignment (Fig. 50.18). The arrangement of fibers provides greater tensile strength to these tissues. It has been noted that although most collagenous fibers of a tendon are aligned in the same direction, they are not solely parallel. They intertwine to form small bundles, which again intertwine to form the

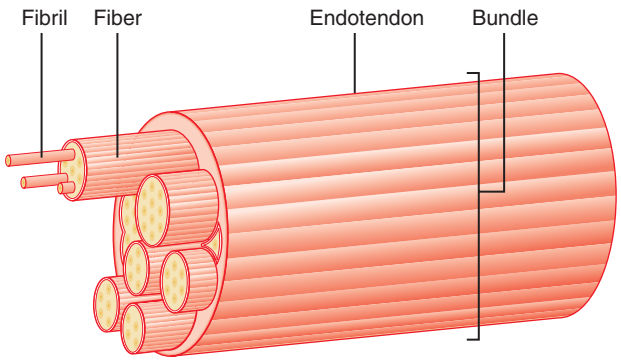


FIG 50.19 Schematic representation of a tendon.

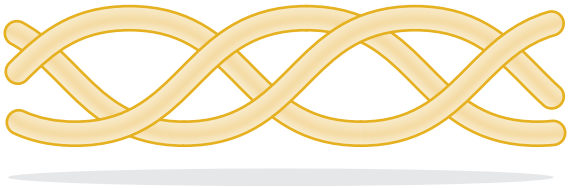


FIG 50.20 Triple-helix formation of collagen molecules.

larger parallel bundles that give tendons their unique appearance. As tendons near the bony attachments, larger tendon bundles also intertwine with each other. As a result, pull of any part of the muscle, instead of being limited to a tendon bundle, is spread widely through the tendon. Collagen fibers of the tendon nearest the bone blend into fibrocartilage and then become mineralized, merging into bone and forming a firm attachment, called an *enthesis*.

Ligaments are similar in appearance to tendons, but they unite bone to bone rather than muscle to bone. Most ligaments are composed of dense collagenous tissue, whereas a few consist of almost pure elastic tissue.

Fig. 50.19 shows a schematic representation of a tendon. Tendon and ligament tissue is composed of few cells (fibroblasts) and large amounts of extracellular matrix. Approximately 20% of the total tissue is fibroblastic and 80% of the structure consists of extracellular matrix. Of the matrix, 70% is water and 30% is solid material. The solids consist of collagen (75%), ground substance, and small amounts of elastin. Compared with ligaments, tendons contain more collagen.

Collagen molecules are in a triple-helix formation (Fig. 50.20), with hydrogen-bonded water bridges or cross-links providing molecular stability. Cross-links give strength to tissue and increase tolerance to mechanical stress.

The protein elastin is found in tendons and ligaments. Elastin provides for some elasticity or extensibility. With the exception of the ligamentum flavum, the majority of tendons and ligaments contain very little elastin, and minimal stretch is allowed. Unlike these stiffer tendons and ligaments, the ligamentum flavum connects laminae of adjacent vertebrae and provides stretch and stability to the spine. The ratio of elastin to collagen fibers in the ligamentum flavum is 2:1.

Ground substance in ligaments and tendons consists of a large amount of proteoglycans, as well as glycoproteins and plasma proteins. The proteoglycan aggregate binds extracellular water in the matrix and acts to stabilize collagen fibers and strengthen ligaments and tendons.

Functional Properties

Tendons and ligaments are quite interesting relative to function. Tendons are extremely strong but can angulate around bony prominences. This capability enables the pull of muscle to change direction and thus improve mechanical leverage. The smooth movement of tendons across bony prominences is facilitated by the presence of bursae. A bursa is a closed sac lined with mesenchymal cells, and is located where one tissue must glide over another. Ligaments are supple and flexible but at the same time rigid. Ligaments stabilize the joint because of their rigidity but allow mechanically correct movement of the joint because of their suppleness.

The strength of a tendon or ligament is determined by the number and quality of cross-links within collagen molecules. As a child matures into a young adult, the increase in the number and quality of cross-links contributes to an increase in tendon and ligament strength.

Response to Injury, Stress, and Aging

With disuse of muscle, ligaments and tendons lose elasticity and resiliency, which may cause them to be more prone to injury. With aging, the tensile strength and stiffness of ligaments and tendons decrease as the proliferative and synthetic activity of fibroblasts declines.

Tolerance to stress is also compromised during pregnancy and the postpartum period. During pregnancy, a laxity of tendons and ligaments is noted with a subsequent increased potential for injury. Estrogens relax various pelvic ligaments during pregnancy, and the sacroiliac joint and symphysis pubis become elastic. These alterations allow easier passage of the baby through the birth canal.

Similar to bone, ligaments and tendons respond to mechanical demands placed on them. Increased stress causes these structures to become stronger and tolerate higher mechanical loads. With a decrease in stress, ligaments and tendons become less taut, weaker, and potentially more susceptible to injury. Immobilization may also decrease the tensile strength of ligaments.

KEY POINTS

- Ligaments and joint capsules connect bones to bones and provide stability to joints.
- Tendons attach bones to muscles to allow movement.
- Tendons and ligaments are composed of dense connective tissue formed by fibroblasts.
- Collagen and elastin are the primary protein components in tendons and ligaments. Most tendons and ligaments have little elastin, which makes them strong but not very compliant. An exception is ligaments that connect adjacent vertebrae, which have more elastin than collagen.
- Tendons are composed of many very fine fibers, each of which originates in endomysium.
- Ligament and tendon strength is determined by the quantity and quality of collagen cross-links.
- Maximal strength is achieved in young adulthood; pregnancy and aging reduce collagen strength.
- Ligaments and tendons respond to increased functional demand by increasing strength. Disuse results in weakened structures.

STRUCTURE AND FUNCTION OF SKELETAL MUSCLE

Approximately 40% of the total body weight is composed of skeletal muscle. Nearly another 10% is smooth and cardiac muscle. Although many of the same principles of contraction apply to these various muscle types, skeletal (striated) muscle will be the focus here. Skeletal muscle not only enables bones to move at the joint, but also provides strength, stability, and protection of the skeleton by distributing loads and absorbing shock.

Composition

The structural unit of skeletal muscle is the muscle fiber (Fig. 50.21). A skeletal muscle is composed of thousands of muscle fibers. Each fiber is a single muscle cell enclosed in a membrane called the *sarcolemma*. Muscle fiber cells are narrow but can be very long and are grouped together in bundles called *fasciculi*. Individual muscles are composed of many fasciculi. The sarcolemma of an individual muscle fiber is surrounded by connective tissue called the *endomysium*. Connective tissue surrounding the fasciculi is called the *perimysium*. Connective tissue surrounding the entire muscle is called the *epimysium*. The epimysium tracks continuously with the endomysium and the perimysium (Fig. 50.22A). Tendons are attached to bones by Sharpey fibers, which are continuous with the perimysium.

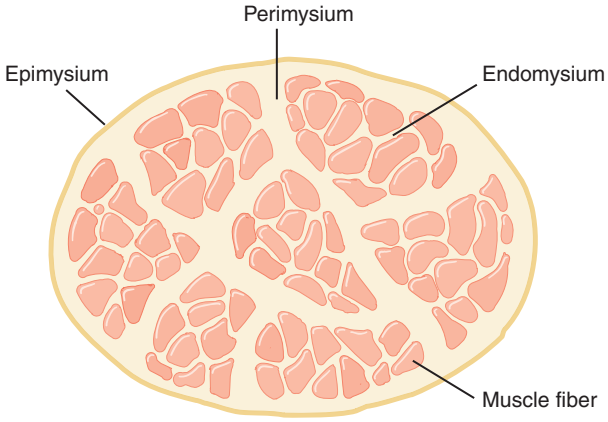


FIG 50.21 Cross-section of skeletal muscle.

The arrangement of fasciculi varies among muscles and can present a specific visual effect of the muscle (Fig. 50.22B). Fasciculi that lie parallel to each other are often found in muscles that function to generate larger range-of-motion joints. Muscles designated as strap or spiral have fibers situated in parallel arrangements. Fibers situated in an oblique pattern relative to the long axis of the muscle are called *unipennate, bipennate*, or *multipennate* muscles. Pennate (Latin for "feather") muscles usually contain a large number of muscle fibers and can transmit a large amount of force to the muscle tendon. Examples of pennate muscles include the gastrocnemius (a bipennate muscle), the deltoid (a multipennate muscle), and the flexor pollicis longus found in the forearm and serving the thumb (a unipennate muscle).

The cytoplasm of the muscle fiber is called the *sarcoplasm*. Structures composing the sarcoplasm include ribosomes, glycogen, and mitochondria, which are required for cell metabolism. Muscle contraction is accomplished by protein filaments of the contractile apparatus.

Contractile Apparatus

Microscopic inspection of a skeletal muscle cell reveals a typical pattern of banding called *striation*. This striated appearance is due to an organized structure of proteins (myofibrils) of the contractile apparatus (Fig. 50.22C). The contractile proteins are long and narrow, called *filaments*, and include two types: myosin and actin. Myosin filaments are larger and are referred to as *thick filaments*. Thin filaments are actually composed of three different types of proteins bundled together, including actin as the primary constituent of thin filament, with smaller amounts of the proteins tropomyosin and troponin bound to it.

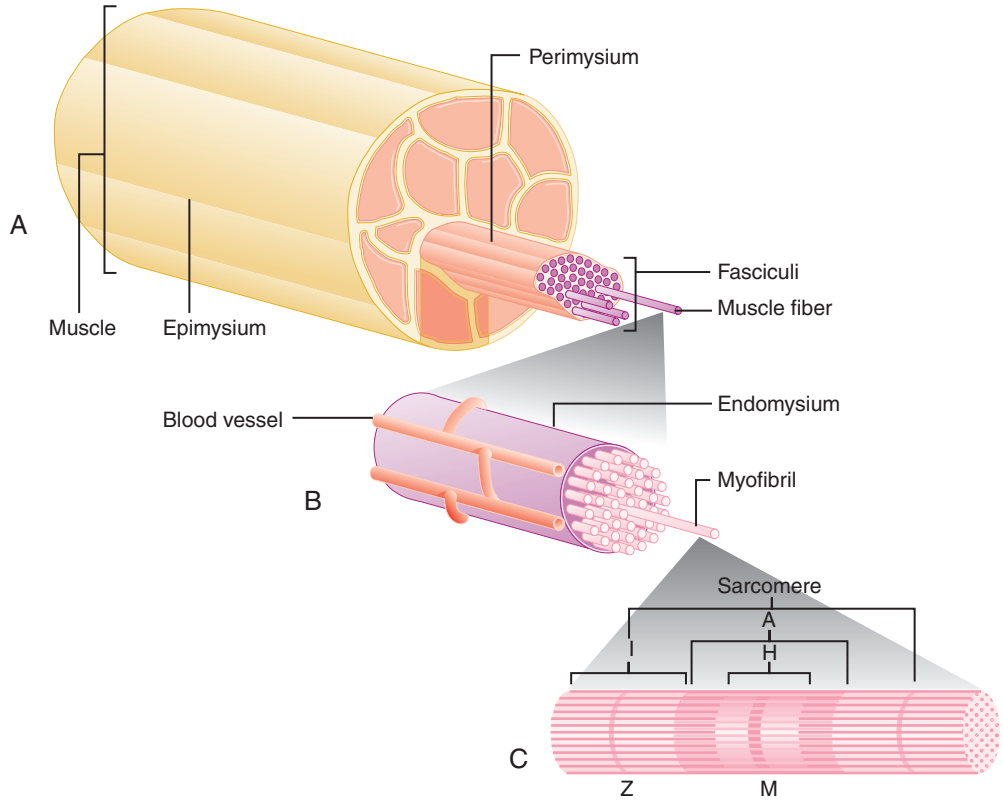


FIG 50.22 Structure of muscle fiber. **A,** The epimysium extends continuously with the endomysium and the perimysium. **B,** The arrangement of fasciculi varies among muscles. **C,** The banding pattern apparent on microscopic inspection of a muscle cell results from the organized structure of the proteins (myofibrils) of the contractile apparatus.

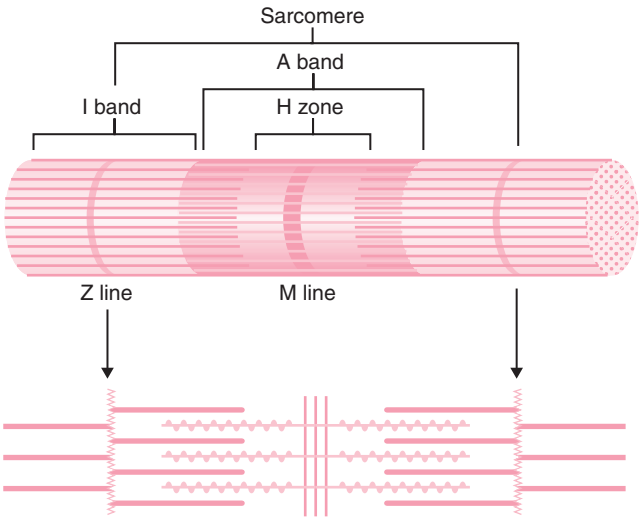


FIG 50.23 Thick and thin filaments are organized into contractile units called *sarcomeres*.

Thick and thin filaments are specifically arranged in contractile units called *sarcomeres* (Fig. 50.23). Sarcomeres are defined by dark bands called *Z lines* that lie perpendicular to actin and myosin filaments. A sarcomere extends from one Z line to the next. Thin actin filaments are attached to Z lines and extend from them toward the center of the sarcomere. The I bands (isotropic) are light in color and correspond to the position of thin actin filaments extending in both directions from the Z line. Thick myosin filaments lie parallel to and between the thin filaments. Each myosin filament is actually surrounded by a hexagonal arrangement of six thin filaments. The dark A band corresponds to an area where actin and myosin filaments overlap. An M line marks the center of the A band and the midpoint of myosin filaments. One other zone, the H zone, corresponds to a region occupied solely by myosin filaments with no actin filament overlap. An efficient, synchronized interaction of the thick myosin and thin actin filaments generates the contraction, a process that is enhanced by this precise arrangement of contractile elements. (See Chapter 17 for a detailed description of contractile filament structure.)

KEY POINTS

- Muscles are composed of bundles of muscle fibers called *fasciculi*.
- A single muscle fiber is one elongated muscle cell packed with contractile proteins and cytoplasmic organelles.
- Connective tissue encases each fasciculus (endomysium) and the muscle as a whole (perimysium).
- Tendons that attach muscle to bone are continuous with the perimysium.
- The arrangement of fibers within a muscle may be parallel or oblique.
- A parallel arrangement occurs in muscles with greater range of motion.
- Oblique patterns occur in muscles with large force potential.
- Skeletal muscle is striated because of an orderly arrangement of contractile proteins in muscle cells.
- Myosin is the primary component of the thick filament. Thin filaments are composed mainly of actin, with smaller amounts of the regulatory proteins troponin and tropomyosin.

MECHANICS OF MUSCLE CONTRACTION

To accomplish the powerful shortening, or contraction, of a muscle fiber, several processes are necessary. Contraction allows muscle tissue to pull on bones, and thus body movement is possible. The molecular basis of muscle contraction is described by the sliding filament, or cross-bridge, theory.

Sliding Filament Theory

The sliding filament, or cross-bridge, theory of muscle contraction is suggested by the anatomic configuration of the sarcomere. Muscle shortening is accomplished by increasing the amount of overlap of actin and myosin filaments. The Z lines at the ends of the sarcomere move closer together as interdigitating actin and myosin filaments slide past one another. Myosin head groups grip binding sites on actin filaments and pull the thin filaments toward the sarcomere's center. Each time a myosin head binds an actin bead, it forms a cross-bridge. Flexible myosin heads move in a ratchetlike manner to tug on actin filaments. Myosin heads bend back and forth, binding and pulling on actin filaments in a steplike fashion. Actin filaments are prevented from slipping back to their original position because some myosin–actin bonds are forming while others are breaking. Formation and subsequent breaking of each actin–myosin cross-bridge requires one molecule of adenosine triphosphate (ATP). Consequently, tremendous quantities of ATP are hydrolyzed with each muscle contraction.

The three energy-producing processes for ATP production are (1) the ATP–phosphocreatine system in which energy for resynthesis of ATP is derived from one compound, phosphocreatine; (2) anaerobic glycolysis, which generates lactic acid but provides some ATP from the partial degradation of glucose or glycogen without oxygen; and (3) the aerobic system, which uses oxygen and has two parts: part A, in which oxidation of carbohydrates is completed, and part B, in which fatty acids and some amino acids are oxidized.

ATP is the immediate source of energy for muscle contraction. Glucose, obtained from glycogen in the muscles and liver, is the primary source of energy for muscle contraction. When enough oxygen is present, glucose is oxidized to carbon dioxide and water. The energy released is partly used to form more ATP. Some energy is wasted in heat. When enough oxygen cannot be supplied via the respiratory and vascular systems, as during intense exercise, glucose is converted to lactic acid. The smaller amount of energy liberated by this anaerobic reaction contributes to the formation of additional ATP. Lactic acid is basically toxic to muscle, and oxygen is needed to remove it, so the muscle is said to have accumulated an oxygen debt. Resting muscle receiving enough oxygen uses the oxygen to reform glucose and glycogen from lactic acid and oxidize the lactic acid to carbon dioxide and water.

Role of Calcium

Muscle contraction depends on an adequate amount of calcium ions in the cytoplasm of the muscle cell. In the absence of free intracellular calcium, no muscle contraction will take place even though myosin head groups have high affinity for actin-binding sites. This phenomenon can be explained in the following way. Myosin heads are prevented from binding to actin by tropomyosin proteins, which lie on top of actin-binding sites. The position of tropomyosin protein is controlled by troponin. When calcium is absent, troponin induces tropomyosin to cover the actin-binding sites. When calcium is present, troponin allows tropomyosin to move and uncover the binding sites (Fig. 50.24). Cross-bridge formation immediately ensues because myosin heads have a high affinity for these sites in the relaxed state.

Electromechanical Coupling

The nerve impulse that a muscle fiber receives to begin contraction is transmitted through the α-motor neuron (Fig. 50.25). The neuron's cell body is located in the anterior horn of the spinal cord. The axon extends from the cell body to the muscle and divides into many small

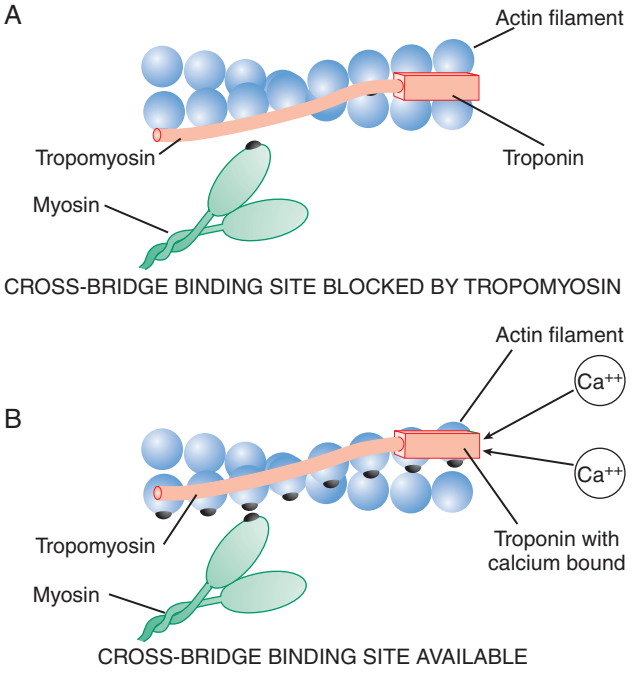


FIG 50.24 The proteins troponin and tropomyosin regulate the ability of actin and myosin to form cross-bridges. **A,** In the absence of calcium, tropomyosin covers the binding sites on actin and inhibits cross-bridge formation. **B,** In the presence of calcium, troponin induces the tropomyosin to uncover the actin-binding sites and allows cross-bridge formation.

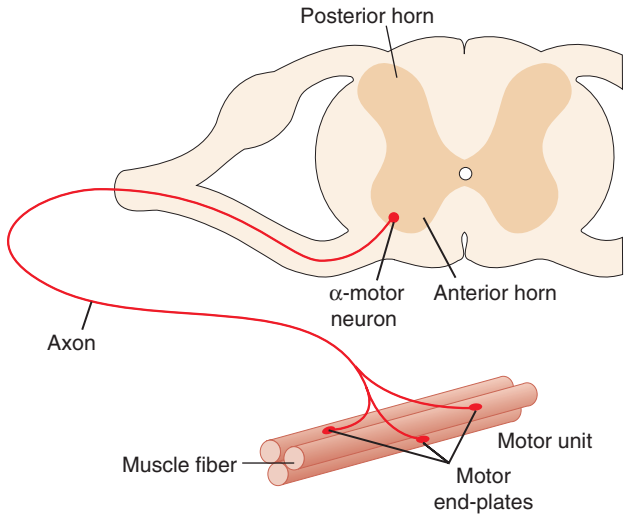


FIG 50.25 Relationship of the α-motor neuron to the motor unit of muscle. The nerve impulse that a muscle fiber receives to begin contraction is transmitted through the α-motor neuron. All muscle fibers innervated by a single motor neuron are part of one motor unit.

branches. Each branch ends in a structure called a *motor end-plate*. The end-plate is positioned near the sarcolemma of a single muscle fiber. All muscle fibers innervated by a single motor neuron are part of one motor unit (see Fig. 50.25).

After the nerve impulse is transmitted from the cell body, it passes along the axon to the motor end-plate. Acetylcholine is released into the neuromuscular synapse and diffuses across to bind with receptors on the skeletal muscle cell. More than enough acetylcholine is released with a single action potential to ensure depolarization of the muscle cell. Acetylcholine binding opens channels in the membrane that allow sodium ions (Na^+) to flow into the cell. Depolarization of the motor end-plate area to threshold then opens voltage-gated channels and produces an action potential. An enzyme (acetylcholinesterase) in the synapse quickly degrades acetylcholine to stop receptor activation. The sarcolemma is depolarized, and an action potential spreads along the surface of the sarcolemma and into the interior of the fiber through the transverse tubules (T tubules). The sarcoplasmic reticulum, a calcium-storing structure, fills the space between myofibrils and forms sacs. The sacs, the terminal cisternae, are positioned close to the T tubules. When the action potential passes down the T tubules, free calcium from the terminal cisternae is released into the myofibrils. Release of the calcium ions stimulates the actin–myosin cross-bridge, thereby causing muscle tension. After depolarization, or when the sarcolemma becomes electrically stable, calcium ions rebind in the sarcoplasmic reticulum and the muscle fiber relaxes.

The motor unit is the functional unit of skeletal muscle and consists of the α-motor neuron and all of the muscle fibers that it innervates. When stimulated, all of the muscle fibers innervated by a motor unit will respond as one. This response is called the all-or-none response, which means that the motor unit will contract to its maximum or it will not contract at all. The size of the contraction of the muscle depends on the number of motor units recruited. The greater the demand placed on the muscle or the more stimuli provided, the greater the number of motor units firing. Fibers of each motor unit are not in contact with each other but are dispersed throughout muscle and intermixed with other fibers. If a single motor unit is stimulated, a large section of muscle visibly contracts. If additional motor units of the nerve are stimulated, the muscle can contract with greater force. *Recruitment* is the term used for calling in more motor units in response to an increase in stimulation of the motor nerves.

Types of Muscle Contraction

Electromyography is used to evaluate muscle contraction. With electromyography, aspects of the contractile process, such as time relationships between the beginning of electrical activity and the actual contraction of the muscle, may be studied. The mechanical response of a muscle to electrical stimulation causes movement in the joint, control of joint motion, or joint stabilization.

Twitch Contraction

The fundamental unit of recordable muscle activity on electromyography is the muscle twitch. A twitch is the mechanical response to a single stimulus of a motor unit. After stimulation, a latency period follows before tension in muscle fibers begins to increase. This latency period is the time required for elastic structures to tighten to prepare for the development of tension. The time from initial tension development to peak tension is called the *contraction time*. The period between peak tension and zero tension is the relaxation time.

An action potential in a muscle lasts only a fraction of a second. It is possible for a series of action potentials to be initiated before completion of the first twitch. The mechanical response to repetitive stimuli is known as *summation*. The period before a second stimulus can induce a twitch during the latency period of the first muscle twitch so no additional response of the muscle occurs is termed the *refractory period*.

The frequency of motor unit stimulation is quite variable. The greater the frequency of stimulation, the greater the tension produced in the muscle. A muscle may achieve higher levels of work when it shortens immediately after being stretched. The elastic components of muscle do not entirely account for this phenomenon. Some energy must be stored in the contractile component of muscle. If the stimulus is so

great that the ability of the muscle to increase tension is exceeded, the muscle is said to be *in tetanus*. In this situation, the speed of stimulation is faster than the contraction–relaxation time of the muscle. Little relaxation occurs before the next contraction.

The variable grade of contraction demonstrated by muscles is important. Repetitive twitching of all recruited motor units develops as a summation of contractions of the muscle, which is responsible for the smooth movements of skeletal muscle.

Based on the mechanical activity that they exhibit, muscles can be divided into two groups: slow twitch (type I, red) and fast twitch (type II, white). Slow-twitch fibers contract and relax more slowly. They support high levels of oxidative metabolism instead of using glycolytic processes to produce energy. Continual energy is provided by large amounts of myoglobin, which potentiates the action of stored oxygen. Slow-twitch fibers are modified for either prolonged or continuous muscle activity, such as in the case of endurance marathon running and maintaining posture. These muscles have a high content of myoglobin.

Fast-twitch muscle fibers depend on energy released from the glycolytic process. Type II fibers are used for fast muscle contractions such as in sprinting, eye blinking, or jumping. Because of the lack of myoglobin in their fibers, they are white. Fast-twitch fibers fatigue more easily than slow-twitch fibers.

Concentric, Eccentric, and Isometric Contractions

Contraction of a muscle exerts a force that causes a torque, or turning, effect on the joint involved. When muscle force generates sufficient tension to overcome the resistance of a limb, the muscle will shorten and joint movement occurs. This shortening contraction is called a *concentric contraction*. Lifting a cup of water to one's mouth is an example of concentric contraction of the biceps muscle. If the load is greater than the amount of tension that the muscle is able to generate, the muscle will lengthen even though it is contracting. A lengthening contraction is termed an *eccentric contraction*. Walking down stairs is an example of an eccentric contraction of the quadriceps muscles. A third type of contraction is an *isometric contraction*. No movement occurs, and the muscle maintains its specific length. Holding a weight in the hand with the elbow flexed is an example of an isometric contraction.

Combined actions of concentric, eccentric, and isometric contractions provide the body with the ability to control movement and function in the environment. Walking, eating, and lifting require the interaction of various types of contractions to provide for smooth and coordinated activity. In a rehabilitative situation, it is interesting to note that isometric contractions generate greater tension than concentric contractions. Eccentric contractions may generate more tension than isometric contractions. When using strengthening programs in rehabilitation settings, a working knowledge of strengthening and tolerance of traumatized tissue is imperative to ensure safe reconditioning.

Mechanical Principles

The amount of tension that a muscle can generate is dictated by a number of mechanical concepts or principles of relationship. These principles include the length–tension relationship, the load–velocity relationship, the force–time relationship, and the effects of muscle temperature change and muscle fatigue. Muscle tension, fatigue, and prestretching are other important factors in muscle force production. A brief review follows.

Length–Tension Relationship

Maximal tension is produced when muscle is at its usual resting length because this position allows actin and myosin filaments to overlap and provide the maximal number of cross-bridges between filaments. At a short resting length, little tension or muscle shortening is possible inasmuch as myosin filaments abut the Z line. If muscle fiber is held at lengths beyond the resting length, tension decreases because actin and myosin do not overlap and therefore no active tension is present.

Load–Velocity Relationship

The velocity of shortening of a muscle contracting concentrically is inversely related to the weight of the applied load. The lower the weight, the higher the velocity of contraction. The greater the weight, the slower the contraction of the muscle. An isometric contraction occurs when the load equals the amount of force that the muscle exerts. If the load exceeds the force generated by the muscle, an eccentric contraction occurs. Greater load leads to faster eccentric lengthening.

Force–Time Relationship

The longer the time of contraction, the greater the force that the muscle can generate until the muscle reaches its point of maximal tension. An increase in the duration of force allows higher levels of tension to be produced by the contractile structures.

Effects of Temperature Change

Conduction velocity across the sarcolemma increases with a rise in muscle temperature. Temperature elevation increases the enzymatic activity of muscle metabolism and the elasticity of collagen in elastic components. For example, the warm-up exercises performed by athletes increases muscle temperature as a result of the increased blood flow and heat generated by metabolism. Both of these changes increase the amount of force that a muscle can produce.

Effects of Fatigue

The availability of ATP determines muscles' ability to contract and relax. Prolonged activity of muscles can be sustained only when the muscle has an adequate supply of nutrients and oxygen to synthesize ATP. If the activity is of sufficient intensity to deplete ATP faster than it can be replaced, muscle tension will gradually weaken and at some point drop to zero. When muscle returns to its original state, creatine phosphate, a major storage form of energy in muscle, must be resynthesized and glycogen stores replaced. This revitalization process requires energy, so the muscle will continue to consume oxygen at high rates even after termination of activity. Heavy, rapid breathing continues after a period of strenuous exercise to provide adequate oxygen for ATP synthesis. This oxygen is also essential for removing lactic acid from muscle. Resting muscle uses oxygen to reform glucose and glycogen from lactic acid and to oxidize the lactic acid to carbon dioxide and water.

Response to Movement and Exercise

Early mobilization may prevent muscle atrophy after surgery or injury. With early motion, muscle fibers position themselves in a more parallel alignment as opposed to the fibers in an immobilized individual. With movement, capillarization occurs more rapidly and tensile strength improves more quickly. With immobilization, the cross-sectional area of muscles decreases and oxidative enzyme activity is reduced. Early mobility prevents atrophy. Afferent impulses from the muscle spindles are increased, thus improving the stimulation of some muscle fibers.

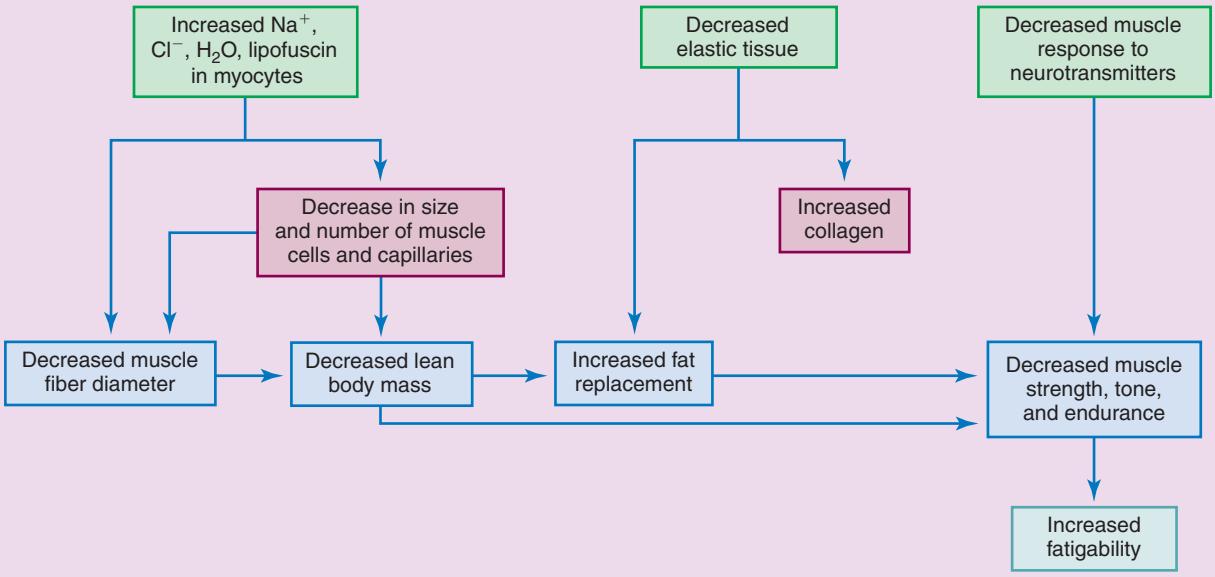
Physical training and conditioning increase the cross-sectional area of muscle fibers. An increase in area coincides with an increase in muscle bulk and strength. In addition, stretching exercises are effective in preventing injury and improving performance, as well as increasing muscle flexibility, maintaining and improving joint motion, and enhancing the elasticity and length of the musculoskeletal unit. The effects of aging on the muscular system are described in "Geriatric Considerations: Changes in the Muscular System."

KEY POINTS

- The fundamental unit of muscle contraction is the sarcomere.
- A sarcomere extends from one Z line to the next and consists of interdigitating thick and thin filaments.
- Muscle contraction occurs when myosin head regions bind to sites on the actin filament to form cross-bridges.
  - After binding, myosin tugs on the actin filament, which causes thick and thin filaments to overlap more.
  - Myosin then releases and proceeds to bind at another point farther along the actin filament. Each cross-bridge cycle requires one molecule of ATP.
- For contraction to occur, the cytoplasm must have sufficient calcium ions.
  - In the absence of calcium, tropomyosin covers binding sites on the actin filament and prevents cross-bridge formation.
  - Another regulatory protein, troponin, controls the position of tropomyosin.
  - When calcium is bound to troponin, tropomyosin is moved to expose binding sites on actin, and cross-bridge formation ensues.
- Calcium ions are stored in the sarcoplasmic reticulum and released into the cytoplasm when the muscle cell depolarizes during an action potential.
- A group of skeletal muscle cells innervated by a single motor neuron is called a *motor unit*. All of the cells in the unit contract simultaneously when the motor neuron depolarizes.
  - An action potential in the α-motor neuron releases acetylcholine at the motor end-plate. Acetylcholine binds to receptors on the muscle cell membrane and triggers an action potential in the cell. To generate more force in the muscle, a greater number of motor units can be activated, a process termed *recruitment*.
- Activation of a motor unit by a single action potential results in a brief twitch contraction.
- A train of action potentials in the motor neuron results in a sustained contraction, in which calcium is released into the cytoplasm faster than it is removed.
- Sustained contraction in response to repetitive stimulation is termed *summation*.
- Muscle contraction does not always result in muscle shortening.
  - Isometric contraction refers to contraction with no change in muscle length.
  - Eccentric contraction occurs when the muscle lengthens while contracting (because of a high load).
  - Muscle shortening with contraction is termed *concentric*.
  - Isometric contraction generates greater tension than concentric contraction; eccentric contraction may generate the highest amount of tension.
- The behavior of contracting muscle is governed by several mechanical principles:
  - Length–tension relationship: Up to a point, a greater resting length of the muscle generates a greater force of contraction. Optimal actin–myosin overlap occurs at about the usual resting muscle length.
  - Load–velocity relationship: The velocity of muscle shortening is inversely related to the weight of the applied load.
  - Force–time relationship: A longer contraction is associated with a greater force of contraction.
- Creatine phosphate is a storage form of energy that is quickly converted to ATP when cellular ATP levels fall.
- Fatigue results when energy and nutrient supply are insufficient.
- A higher rate of muscle oxygen consumption occurs during and for a period after muscle activity.
- Lack of muscle use (disuse) leads to a reduction in muscle mass and slowing of oxidative enzyme activity.
- Early activity after injury is associated with quicker recovery of tensile strength, less atrophy, and better circulation.

GERIATRIC CONSIDERATIONS

Changes in the Muscular System



With aging, the size and number of muscle cells decrease. The remaining muscle cells undergo atrophy, resulting in decreased muscle fiber diameter and reduced amount of elastic tissue. These changes result in reduced muscle mass. Within muscle cells, amounts of extracellular sodium, chloride, water, and lipofuscin pigment increase, with diminished intracellular potassium concentration. The loss of muscle protein may not be obvious because of increased collagen and fat replacement.

GERIATRIC CONSIDERATIONS

Changes in the Muscular System—cont'd

Because fewer capillaries are available to supply the muscles, removal of metabolites is decreased. Hormonal stimulation of muscle by testosterone, somatotropin, and thyrotropin is decreased in addition to reduced muscle uptake of glucose during exercise.

Muscle response to nervous system stimulation is decreased with reduced amounts of muscle norepinephrine. Muscles are also less responsive to neurotransmitters, including acetylcholine at the myoneural junction, and cholinesterase activity is decreased as well.

The muscle, neural, and hormonal changes of aging that affect the muscular system lead to a functional decrease in muscle strength of 30% to 50%, reduced muscle endurance, diminished muscle tone, and increased fatigability. Muscular decline rises with increasing age and usually occurs earlier in men; however, the extent of muscular system decline varies. An elderly individual with good nutritional balance and protein intake combined with adequate active exercise maintains muscle function and strength.

SUMMARY

The musculoskeletal system provides movement for the body. Alterations in function of the musculoskeletal system that decrease the efficiency of movement can often magnify the stress placed on uninvolved structures and increase the potential for degeneration, joint laxity, and pain.

A sound knowledge base of the anatomy, physiology, and mechanics of movement makes the diagnosis of aberrant movement and function easier and thus aids in planning the necessary care and education of patients. Too often, in attempts to provide relief for patients with musculoskeletal dysfunction, the tissue involved or the mechanism of activity causing the injury is not properly identified. Short-lived relief of pain may be provided through the use of analgesics. However, return to activity exacerbates pain and the restriction of movement present before medical care.

The familiarity with the musculoskeletal system also allows health care providers to identify basic tissues involved in injury or disease. Such awareness empowers the clinician to provide relief of pain as well as to address quality-of-life issues. Determination of the type of activity that creates the problem and identification of segments of the musculoskeletal system affected provide the basis to achieve positive long-lasting improvement. Weakness, instability, or decreased motion of structures involved in the dysfunction must be identified. Analysis of physical limitations can lead to patient education and referral to sources that can reduce the impact of physical limitations. In addition, use of recommended exercises, modification of living and working environments, and education of family members will help affected individuals achieve optimal motor function and prevent further injury.

RESOURCES

Barrett KE, Barman SM, Boitano S, Brooks HL, editors: *Ganong's review of medical physiology*, ed 25, New York, 2016, McGraw-Hill.

Beers MH, Berkow R, editors: *The Merck manual of geriatrics*, Rahway, NJ, 2000, Merck.

Firestein GS, Budd RC, Gabriel SE, et al: *Kelley's textbook of rheumatology*, ed 9, Philadelphia, 2013, WB Saunders.

Giangregorio L, Blimkie CJ: Skeletal adaptations to alterations in weight-bearing activity: a comparison of models of disuse osteoporosis. *Sports Med* 32(7):459–476, 2002.

Hochberg MC, Silman AJ, Smolen JS, et al: *Rheumatology*, ed 6, Philadelphia, 2015, Elsevier.

Klippel JH, Stone JH, Crofford LJ, White PH, editors: *Primer on the rheumatic diseases*, ed 13, New York, 2008, Springer.

Seeman E, Delmas PD: Bone quality—the material and structural basis of bone strength and fragility. *N Engl J Med* 354(21):2250–2261, 2006.

Wolff J: *Des gesetz der transformation der knochen*, Berlin, 1892, Hirschwold.

51

Alterations in Musculoskeletal Function: Trauma, Infection, and Disease

Carol L. Danning

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- How are the mechanisms of injury different between noncontractile and contractile soft tissue injuries?
- What are the key factors in the mechanisms of wound healing?
- What is the process and duration of normal bone healing after a fracture?
- What are the different complications that can occur after a fracture?
- What are the manifestations, dangers, and management of compartment syndrome?
- What are the clinical findings and management of bone infections?
- How are osteoporosis and osteomalacia or rickets similar, and how do they differ?
- What terminology is used to describe primary bone tumors?
- What are the cause and pathogenesis of muscular dystrophy and myasthenia gravis?
- What are the clinical features and treatment options in fibromyalgia syndrome?

CHAPTER OUTLINE

SOFT TISSUE INJURIES, 1021
Inert Soft Tissue Injuries, 1021
Ligament Injuries, 1021
Clinical Manifestations, 1021
Treatment, 1021
Joint Capsule Injuries, 1022
Adhesive Capsulitis, 1022
Internal Joint Derangement, 1023
Injuries to Fasciae and Bursae, 1023
Fasciae, 1023
Bursae, 1023
Injuries to Nerves, Nerve Roots, or Dura Mater, 1023
Contractile Soft Tissue Injuries, 1023
Injury to Tendons, 1023
Muscle and Tendon Strains, 1024
Blunt Trauma, 1024
Compartment Syndrome, 1024
Soft Tissue Healing After Trauma, 1024
Wound Repair, 1024
BONE INJURIES AND INFECTIONS, 1026
Bone and Joint Trauma, 1026
Types of Bone, 1026
Fracture, 1026
Types of Fracture, 1026
Extent of Fracture, 1027
Diagnosis of Fracture, 1027
Treatment of Fracture, 1028
Healing Process, 1028
Complications of Fractures, 1029
Dislocations and Subluxations, 1030
Infections of the Bone, 1031
Osteomyelitis, 1031
Etiology and Pathogenesis, 1031
Clinical Manifestations, 1031
Healing Complications, 1031
Treatment, 1032
Tuberculosis, 1032
Etiology and Pathogenesis, 1032
Clinical Manifestations, 1032
Risk Factors, 1032
Treatment, 1032
ALTERATIONS IN BONE STRUCTURE AND MASS, 1032
Bone Structure Disorders, 1032
Scoliosis, 1032
Etiology and Pathogenesis, 1032
Clinical Manifestations, 1032
Treatment, 1033
Metabolic Bone Diseases, 1033
Osteoporosis, 1033
Etiology and Pathogenesis, 1033
Clinical Manifestations, 1034
Treatment, 1035
Other Causes of Osteoporosis, 1035

Rickets and Osteomalacia, 1035
Clinical Manifestations, 1035
Treatment, 1035
Paget Disease, 1035
Etiology and Pathogenesis, 1035
Clinical Manifestations, 1036
Treatment, 1036
Bone Tumors, 1036
Benign Tumors, 1036
Osteochondroma, 1036
Chondroma, 1036
Osteoid Osteoma, 1037
Giant Cell Tumor, 1037
Malignant Bone Tumors, 1037
Osteosarcoma, 1037
Chondrosarcoma, 1037
Ewing Sarcoma, 1038
Multiple Myeloma, 1038
DISEASES OF SKELETAL MUSCLE, 1038
Idiopathic Inflammatory Myopathy, 1038
Polymyositis and Dermatomyositis, 1038
Clinical Manifestations, 1038
Treatment, 1038
Muscular Dystrophy, 1039
Duchenne Muscular Dystrophy, 1039
Etiology and Pathogenesis, 1039
Clinical Manifestations, 1039
Treatment, 1039
Becker Muscular Dystrophy, 1039
Etiology, Pathogenesis, and Clinical Manifestations, 1039
Facioscapulohumeral Muscular Dystrophy, 1039
Etiology and Pathogenesis, 1039
Myotonic Dystrophies, 1039
Clinical Manifestations, 1039
OTHER DISORDERS OF MUSCLE, 1039
Myasthenia Gravis, 1039
Treatment, 1039
Chronic Muscle Pain, 1040
Fibromyalgia Syndrome, 1040
Etiology and Pathogenesis, 1040
Clinical Manifestations, 1040
Treatment, 1040

A smoothly functioning musculoskeletal system facilitates a complete range of human actions, including walking, talking, running, breathing, and a myriad of voluntary physical activities. Any abnormality in the musculoskeletal system decreases the efficiency of movement and increases mechanical stress. The ballistic requirements of many sports and occupations, such as skiing and driving an automobile, have increased the potential for trauma. Diseases also disrupt the integrity of the musculoskeletal system. Infectious processes, genetic abnormalities, immune-mediated inflammatory disorders, and nutritional deficiencies may all affect movement.

The clinicians who work with patients experiencing dysfunctions of the musculoskeletal system must have a solid background in evaluation and management of such disorders. Without this preparation, interventions will not be sufficient to promote maximal functional return. This chapter discusses disorders particular to the musculoskeletal system as they affect the soft tissues, skeletal frame, and muscles.

SOFT TISSUE INJURIES

In addition to injuries of the bony skeleton, soft tissue may be traumatized. At times it is difficult to differentiate among the types of soft tissue. In an attempt to differentiate the exact site of a lesion, Cyriax described two types of soft tissue: contractile and inert. Contractile tissue is composed of structures involved in the contraction of muscle and includes not only the muscle belly but also the tendon and bony insertion, called an *enthesis*. Although not involved in a pure contraction, as is the muscle belly, the tendon and its insertion into bone are mechanically linked to tension generated by the muscle. Inert, or noncontractile, tissue possesses no ability to contract or relax. The inert soft tissues include joint capsules, ligaments, bursae, fasciae, dura mater, and nerve roots. Passive stretching provokes pain from inert tissue. Evaluation of inert tissue lesions requires identification of all structures involved: the capsule of a joint, a section of a ligament or a nerve, or the mechanical displacement of the meniscus.

INERT SOFT TISSUE INJURIES

Ligament Injuries

A ligament is a dense connective tissue with parallel-fibered collagenous tissues designed to connect bone to bone. Ligaments contribute to mechanical stability of the joint, guide motion, and prevent excessive motion. Injuries to ligaments occur when loading exceeds the physiologic range of motion. Microfailure precedes total failure of the ligament. With total failure of a ligament, damage to surrounding soft tissue occurs. Ligament injuries are classified by the extent of tear and may be described as mild, moderate, or severe.

Grade 1 (mild) = stretching injury without instability
Grade 2 (moderate) = severe injury with instability but some ligament fibers still intact
Grade 3 (severe) = complete disruption of ligament

Clinical Manifestations

A common site of ligament injury, particularly among athletes, is the anterior cruciate ligament (ACL) of the knee (Fig. 51.1). Symptoms may include a sudden "tearing" sensation or "popping" in the knee followed by pain with weight bearing and often acute swelling of the knee. Another common site for ligament injury is the anterior ankle (talofibular ligament).

Treatment

Treatment is geared primarily toward relief of symptoms, and recovery is usually complete. A moderate ligament injury is a definite tear in some component of the ligament with loss of strength. The fibers are not widely separated. Treatment is primarily protection of the ligament. With a severe ligament injury, the ligament is completely torn and no longer functions. Potentially the fragments are widely separated. Treatment may require surgical restoration of ligament continuity, when possible.

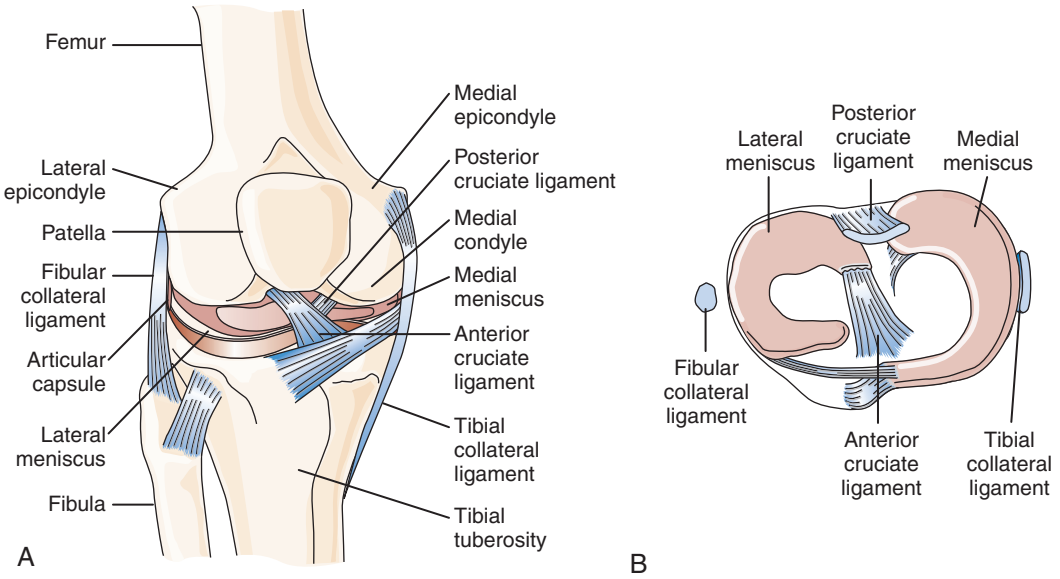


FIG 51.1 The principal structures of the interior of the knee joint. **A,** From the front. **B,** From above with the femur removed.

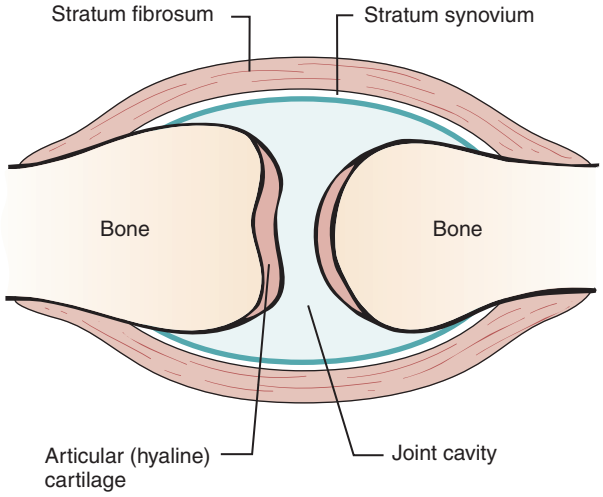


FIG 51.2 Joint capsule.

Joint Capsule Injuries

Another inert structure that is intimately involved in stabilization of a synovial joint is the joint capsule (Fig. 51.2). The joint capsule is composed of an inner layer and an outer layer. The inner layer, which lines the joint cavity, is called the *synovial membrane* (or *synovial lining layer*) and is highly vascularized. It synthesizes the hyaluronic acid component of synovial fluid, produces matrix collagen, and is essential for joint nutrition. The outer layer of the capsule is attached to the periosteum of the bones. The capsule is reinforced by ligaments and musculotendinous structures. The outer layer of the capsule is poorly vascularized but richly innervated by joint nerve receptors. Joint receptors are able to detect the rate and direction of motion, proprioception, compression and tension, vibration, and pain.

After an injury to the joint capsule, the ensuing increase in vascularity and development of fibrous tissue lead to a thickening of the capsule. An effusion (increase in synovial fluid) in the joint cavity may lead to stretching of the capsule and its associated ligaments. The joint capsule, like ligaments, provides joint stability. The capsule, however, has an interesting mechanical adaptation: capsular redundancy. An example of the importance of the redundancy has been identified by Hettinga: "The inferior medial portion of the shoulder joint capsule is a loose, redundant sac that becomes tense only when the shoulder is fully abducted or flexed. The posterior capsule of the knee is loose in flexion but so tight in extension that it becomes an important stabilizer."

Capsular redundancy provides for a stable joint at the end ranges of movement. Any injury or edema in the joint that causes scarring in the lax section of the capsule prevents full range of motion. Prolonged immobilization of a joint causes loss of mobility and extensibility of the capsule, with subsequent loss of motion. Immobilization of a joint causes an alteration in the flow of synovial fluid and contracture of the joint capsule and periarticular muscle. The altered flow of synovial fluid prevents fluid diffusion into and out of cartilage and causes compression and distention of cartilage. Nutrition of the joint components stagnates, leading to degenerative changes in the joint that become permanent. Contracture of the joint capsule and periarticular muscle results when fatty tissue proliferates in the joint space. This increase in the amount of connective tissue leads to adhesions that limit joint motion. The increase in connective tissue is due to failure to keep the latticework of tissue stretched open, which usually occurs by normal flexion and extension of muscle. The muscles bridging the immobilized joint also shorten.

Adhesive Capsulitis

An example of such a restriction is loss of function in the shoulder after even a minor injury, leading to a "frozen shoulder," also called *adhesive capsulitis.* With an injury to any component of the shoulder complex, inflammation occurs in the joint along with swelling and distention of the joint capsule. With prolonged immobilization, thickening of the capsule may ensue, possibly attributable to proliferation of fibroblasts and capsular contraction. Capsular tightness leads to a loss of movement and an increase in pain, especially at night. Excessive joint motion may cause a tearing of the capsule, similar to a ligamentous tear, and render the joint unstable. Conservative treatment of the frozen shoulder is usually recommended with intraarticular corticosteroid injections, gentle stretching and physical therapy, and antiinflammatory

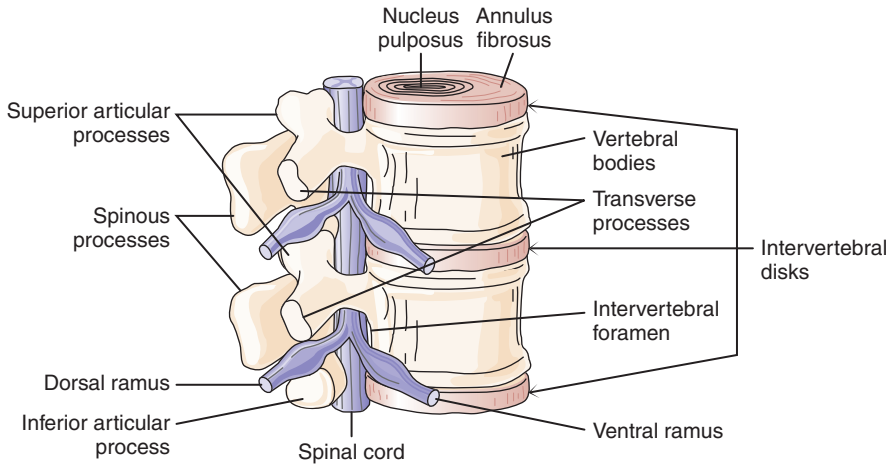


FIG 51.3 Lateral view (somewhat superior) of a segment of the lumbar part of the vertebral column.

medication. Prevention of adhesive capsulitis involves avoiding prolonged or excessive immobilization of the shoulder after minor injuries and performing early, gentle stretching.

Internal Joint Derangement

Internal joint derangement may be caused by injury to inert soft tissue structures. Meniscal tears at the knee, labrum tears at the glenohumeral joint, and disk tears in the temporomandibular joint all cause restrictions of the joint and may lead to soft tissue dysfunction in the form of weakness, loss of motion, or pain. Tears of the medial and lateral menisci in the knee are common causes of knee pain and one of the most common reasons for knee arthroscopy. The medial meniscus is more securely attached to the fibrous joint capsule and therefore is more prone to tearing than the more mobile lateral meniscus. Meniscal tears can be divided into two types: acute traumatic tears and chronic degenerative tears. Acute meniscal tears occur when the weight-bearing knee is rotated. Symptoms include acute pain, joint swelling, and sometimes a locking or popping sensation. In older individuals, meniscal tears are more likely to be chronic and degenerative resulting from arthritis. The pain is usually gradual in onset and often without antecedent injury.

Small meniscal tears may be treated conservatively with antiinflammatory medication, joint stabilization, and physical therapy. Other more significant tears may require surgical repair.

The knee joint is stabilized by four ligaments: the ACL, the posterior cruciate ligament (PCL), the medial collateral ligament, and the lateral collateral ligament. The ACL and PCL prevent anterior and posterior displacement of the tibia relative to the femur, respectively (see Fig. 51.1). These ligaments can be torn in varying degrees of severity. Injury to either of these ligaments can lead to some degree of instability in the knee joint. The collateral ligaments can also be injured with resultant pain and inflammation.

Injuries to Fasciae and Bursae

Fasciae and bursae may also be causes of pain and restriction of movement of the musculoskeletal system.

Fasciae

When connective tissues of the body are arranged in sheaths that envelop muscles, they are designated fasciae. Individual muscles are surrounded by a thin fascia called the *perimysium*. Trauma to fascia, as with any soft tissue, may cause edema and scarring. Restrictions in fascia movement cause a restriction in joint function.

Bursae

In many locations between muscles or between muscle or tendon and bone, connective tissue forms a pocket lined with synovium that contains fluid. These pockets are identified as bursae. A bursa is located in areas of high friction and is designed to dissipate some of the stress. With faulty mechanics of the joint, repetitive movement, or direct trauma, the bursal sac may become inflamed (bursitis) and painful. Bursitis, because of its strategic position at stress points of muscle function, can cause disruption of movement. An inflamed bursa may restrict movement of the joint and lead to restriction in capsular function or muscle dysfunction as a result of edema. Some of the more common sites of bursitis include the trochanteric bursa (lateral hip), subacromial bursa (shoulder), pes anserine bursa (medial knee), and olecranon bursa (elbow).

Injuries to Nerves, Nerve Roots, or Dura Mater

Trauma to any soft tissue may lead to adhesive constriction of the nerve, nerve root, or dura mater. Irritation or entrapment of a nerve causes pain that radiates along the structures innervated by that nerve. Pain, altered sensation (numbness and tingling), motor weakness, and diminished reflexes may result from trauma to these essential soft tissue components of the musculoskeletal system.

An example is trauma to vertebrae in the lumbosacral area with nerve root impingement. Components of the intervertebral disk can herniate and cause pressure on nerve roots (Fig. 51.3). The intervertebral disk is a shock absorber located between vertebrae. The center is a gelatinous-like material, the nucleus pulposus, which has a high water content. The nucleus pulposus is surrounded by the fibrous annulus fibrosus. Trauma to the back can cause unequal pressure on the disk leading to herniation. Common sites of disk problems are at L3 to L4, which affects the L4 nerve root; L4 to L5, which affects the L5 nerve root; and L5 to S1, which affects the S1 nerve root.

CONTRACTILE SOFT TISSUE INJURIES

Injury to Tendons

Injury to tendons occurs along a continuum from a minor strain, in which a few fibers of the tendon are torn, to a complete tear or rupture. The sheath in which a tendon slides may also be traumatized. Inflammation of the tendon within the sheath is called *tendinitis*. This inflammation may be due to infection, direct injury, or injury from repetitive motion. Tendons are injured when the stress placed on them is greater than the fibers can tolerate. Muscle tendons that are subjected to high tensile stress or compression are more prone to injury. Frequently

injured tendons include the following: extensor pollicis brevis and abductor pollicis longus of the thumb (de Quervain disease), rotator cuff of the shoulder, biceps brachii tendon, tendons of the patellar complex, quadriceps tendon, hamstring tendon, Achilles tendon, and posterior tibialis tendon.

Muscle and Tendon Strains

Muscle trauma compromises the contractile unit. As in the case of injury to a tendon, tears in a muscle may range from a minor tear to complete rupture. Most injuries to muscle are due to abnormal or sudden, unexpected muscle contraction. Muscle and tendon strains are often categorized by the severity of injury in a scale similar to that used for ligament injuries: mild, moderate, and severe, with the latter term applying in the setting of total rupture of the contractile structure.

Blunt Trauma

A soft tissue contusion or crush injury also compromises the contractile structure. Any blunt trauma that causes bleeding into the muscle belly may lead to an inability to contract the muscle. In about 9% to 20% of muscle contusions, hemorrhage in a muscle belly has the potential to coagulate and calcify. This abnormal calcification in a muscle results in a painful condition called *myositis ossificans*. Proximal muscle groups are most often affected, such as triceps, quadriceps, and thigh adductor muscles. The aberrant calcification prevents a normal and strong contraction of the muscle involved. Treatment of the initial injury includes rest, ice, and compression as well as antiinflammatory medication. Later, surgical removal of heterotopic calcification may be needed.

Compartment Syndrome

Compartment syndrome is due to trauma to soft tissue caused by the unyielding structure of inert tissue. Causes of compartment syndrome may be divided into three categories: decreased compartment size, increased compartment content, or externally applied pressure. With an injury, edema causes an increase in pressure within the compartment. Because volume is expanding in a confined area, pressure reduces capillary flow. Muscle and nerves become ischemic, with a resultant excruciating pain and tissue damage. For a more detailed discussion of compartment syndrome, see the "Complications of Fractures" section later in this chapter.

Soft Tissue Healing After Trauma

Trauma to soft tissue results in disruption of the circulatory and lymphatic systems. Hemorrhage, fluid loss, and cell death result. Blood vessels at the site of trauma constrict, which limits blood loss from the affected area. Norepinephrine, produced at the injury site, mediates this initial constriction response, which may last a few minutes. Serotonin (from mast cells of connective tissue) and platelets prolong vasoconstriction. These processes may also contribute to vasodilation in inflamed tissue.

Platelets adhere to collagen fibers and release serotonin and adenosine diphosphate, which cause platelets to adhere to the traumatized endothelial wall and form a platelet plug that temporarily decreases bleeding. Activated platelets also release many different cytokines and growth factors, which attract inflammatory cells to the site as well as trigger a blood-clotting cascade. Trauma to the endothelial surface also triggers release of enzymes that initiate clotting by converting prothrombin to thrombin, which converts fibrinogen to fibrin. The endothelial surfaces of small vessels are also compressed, thereby ensuring that vessels remain closed after vasoconstriction has ceased. At this point the release of histamine from mast cells, basophils, and platelets causes vasodilation and increased permeability of local venules. With the increase in permeability, serous fluid–containing cells and plasma proteins accumulate as edema in tissue spaces (Fig. 51.4). This edema fluid contains fibrinogen, which forms fibrin through an interaction with thrombin. Fibrin seals damaged lymphatics and confines the inflammatory reaction to an area immediately surrounding the injury.

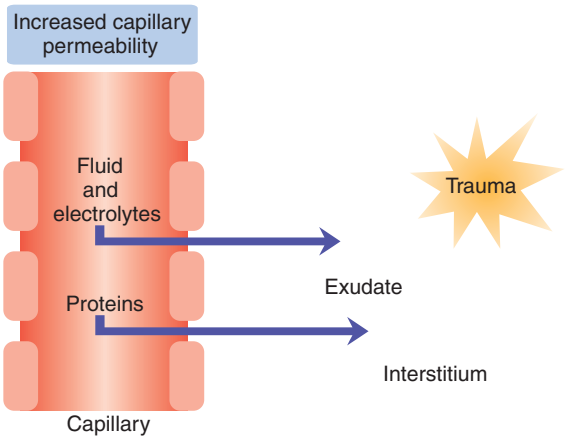


FIG 51.4 Edema formation. With trauma, increased capillary permeability and dilation cause leaking into tissue space. Initially clear, the exudate in the tissue space becomes more viscous with an increase in the amount of plasma protein.

Wound Repair

The inflammatory response prepares injured tissue to progress to the healing process of repair and reorganization. Figs. 51.5 and 51.6 provide a summary of the phases of wound repair. The acute response lasts about 2 weeks, and the subacute phase lasts another 2 weeks. An infiltrate of macrophages and fibroblasts is noted as the wound is cleared of foreign substances. A matrix of collagen, hyaluronic acid, and fibronectin develops. Lymphatics form in the matrix, prevent additional edema, and assist in preventing infection. This granulation tissue develops in the wound space. Macrophages have an important role in wound repair.

Next in the process of wound repair is reepithelialization of the wound surface. Epidermal cells migrate over established epidermal cells until the defect is closed. The formation of basement membrane follows. This membrane is first laid down at the wound periphery and then progresses to the center of the wound. A strong bond forms between epidermal cells and the newly formed basement membrane to complete reepithelialization (granulation tissue formation).

Wound tensile strength is a result of the deposition of collagen. Collagen production begins approximately 5 days after myofibroblast migration into the wound space. Hyaluronic acid, found in the extracellular matrix, assists glycosaminoglycans to stimulate fibroplasia. Myofibroblasts secrete an extracellular matrix, which induces cell migration and proliferation, and synthesize proteoglycans, which stimulate collagen formation and increase tissue resilience and tensile strength. By the end of the first month, tensile strength begins to increase, but several months are required to achieve the maximal level. Collagen reaches its maximal strength approximately 3 months after injury. Maximal tensile strength is only 70% to 80% of preinjury levels.

Revascularization (angiogenesis or growth of new blood vessels) must take place to ensure survival of the new tissue. Angiogenesis at the wound site begins early after the injury in response to local hypoxia from vasoconstriction at the site. Vascularization continues by the development of new circulatory networks in the wound and reattachment of existing vessels. The formation of extracellular matrix, the development of endothelial cells, and the presence of lactic acid and heparin are a few of the factors that stimulate revascularization.

Wound closure or contraction is the final phase of healing in soft tissue injuries. Contraction begins soon after injury and is completed

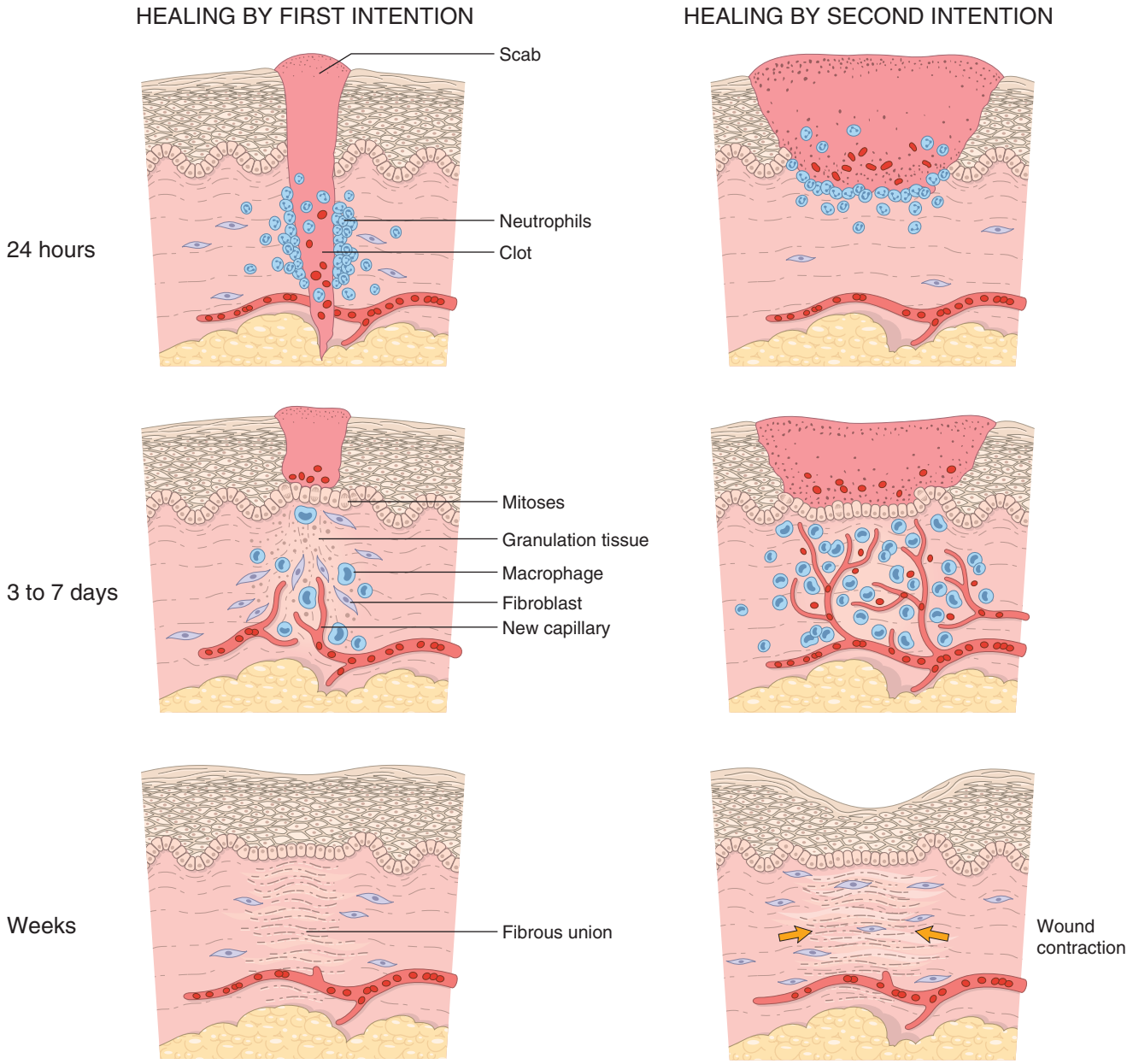


FIG 51.5 Steps in wound healing by first intention and second intention. In the latter, the resultant scar is much smaller than the original wound because of wound contraction.

in approximately 2 weeks. Myofibrils assist in wound closure. Interaction between extracellular matrix and granulation tissue results in a contractile unit called a *fibronexus*. Cytoplasmic actin binds to the fibronexus and draws tissue together to ensure a stable wound. As tension increases across the wound, collagen fibers are reoriented and collagen phagocytosis increases. Complete organization and concentration of collagen may require more than 40 weeks. In the case of rupture of soft tissue structures, surgical intervention may be necessary to ensure that the traumatized tissue is in close enough proximity to allow healing.

KEY POINTS

- Soft tissue injury refers to injuries of noncontractile elements (joint capsule, ligament, bursa, fascia, dura mater, and nerve root) and contractile elements (muscle and tendons).
- Ligament, tendon, and other soft tissue injuries are classified on the basis of the severity of fiber disruption.
- Passive stretching causes pain in noncontractile tissue injury, whereas active contraction is painful in contractile injury.
- Noncontractile tissue injuries generally cause altered range of motion around a joint as a result of pain, edema, adhesion, or fibrosis.
- Contractile tissue injuries are characterized by decreased muscle strength.
- Compartment syndrome is a dangerous complication of soft tissue injury that results from swelling of injured tissue within a restrictive fascia. Unless pressure is quickly reduced, compressed tissue may become ischemic and necrotic.
- Soft tissue injury results in local inflammation and initiates the process of wound healing. Strength of the injured tissue is improved by the deposition of collagen. Normalization of collagen may require more than 40 weeks.

BONE INJURIES AND INFECTIONS

BONE AND JOINT TRAUMA

The skeletal system is subject to alterations in function from mechanical stress and infection. The purposes of the skeletal system are to protect internal organs, contribute to mineral homeostasis, produce blood cells, and provide muscle attachment sites and thus facilitate body movement. Bone is one of the body's hardest structures, as well as one of its most dynamic and metabolically active tissues. Bone is vascular with a capacity for repair. It adapts to mechanical demands placed on it and alters its configuration in response to those mechanical stresses.

Types of Bone

Two basic forms of bone are present in the human body: cortical bone and cancellous bone. Cortical bone forms the cortex, or outer shell, of the bone. Cortical bone is designed to tolerate compression and shearing forces, but tension forces may exceed the tolerance of cortical bone. With bending, twisting, or straight tension, stress may exceed the bone's tolerance, and a fracture occurs on the convex side of the bend. Cancellous bone, which has a spongy or latticelike appearance, is found in the interior of bones. Unlike cortical bone, cancellous bone does not tolerate compression stress. Cancellous bone provides structural support to cortical bone and increases a bone's potential to withstand stress.

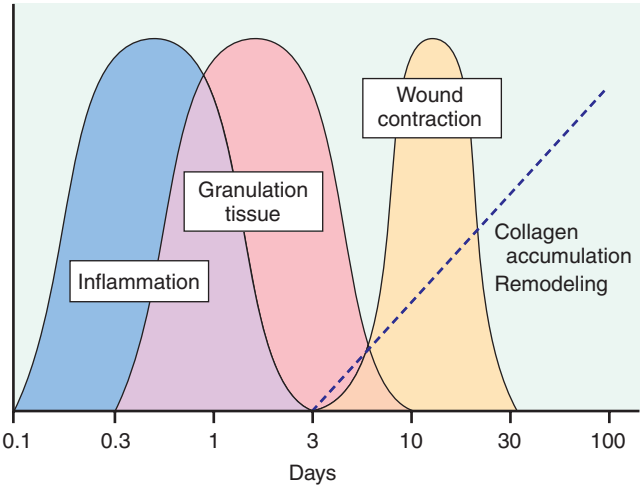


FIG 51.6 Orderly phases of wound healing. (Modified from Clark RA: Basics of cutaneous wound repair. In Goldsmith LA, editor: *Physiology, biochemistry and molecular biology of the skin,* ed 2, vol 1, New York, 1991, Oxford University Press, p 577. In Kumar V et al: *Robbins basic pathology,* ed 8, Philadelphia, 2007, Saunders, p 74.)

Fracture

A fracture is a break in continuity of a bone, an epiphyseal plate, or a cartilaginous joint surface. Trauma generating enough energy to fracture a bone also produces force sufficient to traumatize adjacent soft tissue. With that concept in mind, the remainder of this section will address injury to the bony component of the musculoskeletal system.

Types of Fracture

Fracture type reflects the type of tension stress placed on bone (Fig. 51.7). A transverse fracture occurs in a straight line at approximately a 90-degree angle to the longitudinal axis of the bone. Spiral fractures are the result of rotational forces and cause bone to separate in the form of an S around the bone. Longitudinal fractures split bone along its length. Oblique fractures result from a rotational force, but unlike spiral fractures, the break is along an oblique course (45-degree angle) and does not rotate around the entire bone. Comminuted fractures consist of more than one fracture line and more than two bone fragments. These fragments may be shattered or crushed. Comminuted fractures often present considerable treatment problems because of associated soft tissue damage and multiple bone fragments. An impacted fracture is caused by excessive force that telescopes or drives one fragment into another. A greenstick fracture is an incomplete break in the bone with the intact side of the cortex flexed. It is usually seen in children. A stress fracture is a failure of one cortical surface of the bone, often caused by repetitive activity such as running. Without proper treatment, a stress fracture can become a complete fracture with two distinct fragments. An avulsion fracture is the separation of a small fragment of bone at the site of attachment of a ligament or tendon.

In children, fracture through the shaft of a long bone can stimulate bone growth, possibly because of increased blood flow associated with fracture healing leading to increased nutrients to the growth plates. Particular attention to bone alignment or overlapping of fracture ends must be paid when managing a fracture in a child younger than 10 years of age because of bony overgrowth or skeletal arrest leading to limb length discrepancies. Of special concern are fractures at or near a joint line in children. This location of a fracture may suggest an epiphyseal growth plate fracture (Fig. 51.8). With epiphyseal injuries, the potential for disruption of growth of the long bones is present. Proper reduction and fixation are necessary to avoid growth disturbance in fractures through the growth plate. Crush injury to the epiphyseal plate commonly leads to premature growth cessation.

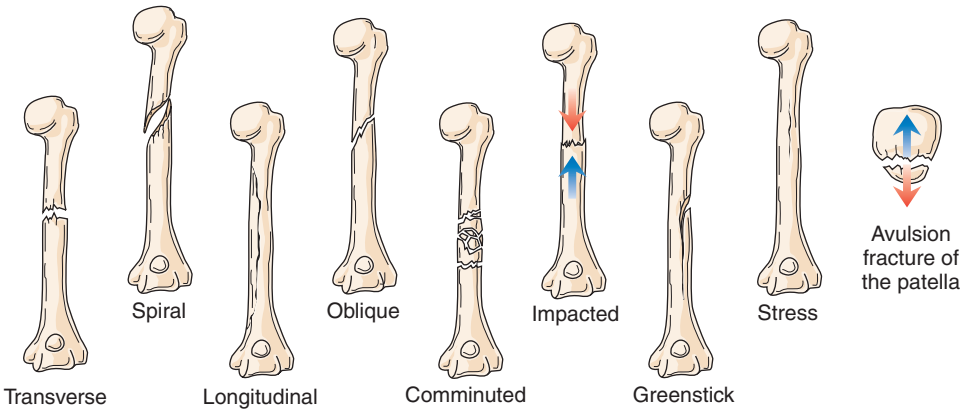


FIG 51.7 Types of fractures.

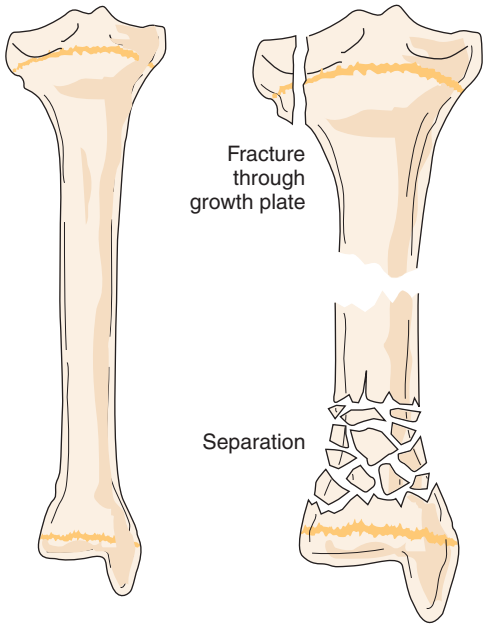


FIG 51.8 Epiphyseal injury. **A,** An uninjured long bone, showing the normal locations of growth plates in dark yellow. **B,** Two types of epiphyseal fractures.

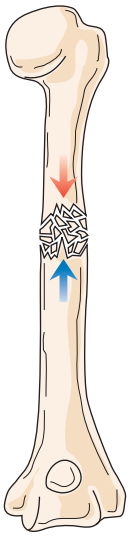


FIG 51.9 Compression fracture.

Cancellous bone does not tolerate compression stress; it buckles and then cracks. Therefore crush or compression fractures (Fig. 51.9) are consistent with cancellous bone trauma. In children, a compression injury to cancellous bone of the metaphysis of a long bone is identified as a buckle fracture, where bone buckles and eventually cracks. In adults, compression fractures are often found in a vertebral body of the spine, especially in older individuals with osteoporosis.

Extent of Fracture

Fractures can be classified according to extent and depth.A displaced fracture is one in which the ends of fracture fragments are separated. In a nondisplaced fracture, the fracture fragments remain in alignment and position. With a depressed fracture, the fragment is displaced below the level of the surface of the bone, usually in the skull. A complete fracture is one in which the fracture line disrupts bone continuity through

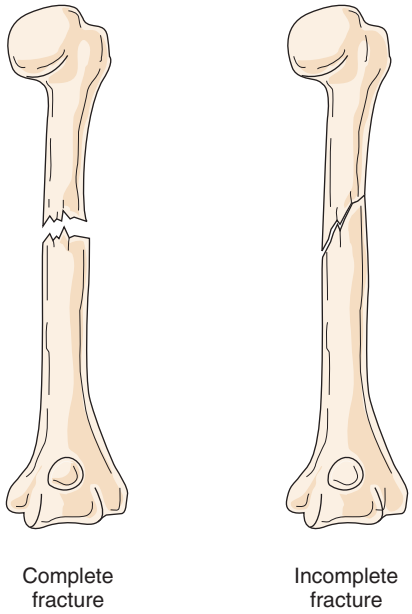


FIG 51.10 Comparison of complete and incomplete fractures.

the whole thickness of the bone, including the cortex (Fig. 51.10). In an incomplete fracture, the cortex of the bone buckles or cracks; however, bone continuity is not disrupted. Incomplete fractures tend to occur in the more flexible, growing bones of children. Fractures can also be classified as open (compound) or closed (simple) (Fig. 51.11). An open fracture occurs when bone is broken and an external wound leads to the fracture site. These fractures present an increased risk of infection and are therefore difficult to manage. A closed fracture is a fracture in which the fragments do not extend through mucous membrane or skin and skin is not broken.

In cases of open fractures, a wound classification system may be used that ranges from type I to type IIIC in increasing degree of severity. Type I is a wound smaller than 1 cm, moderately clean with minimal contamination. The fracture is a simple transverse or an oblique fracture with a bone spike piercing the skin. Soft tissue damage is minimal. Type II wounds are larger than 1 cm with moderate contamination. The fracture might be a moderate comminution or crush injury with moderate soft tissue damage. Type III wounds have a high degree of contamination. The fracture is severely comminuted and unstable. It is accompanied by much soft tissue damage involving muscle, skin, and neurovascular structures. Traumatic amputations would be classified here. With type IIIA wounds, soft tissue coverage of the fracture is sufficient. Segmental or severely comminuted fractures occur with these wounds. Type IIIB wounds with open fractures include extensive injury or loss of soft tissue, as well as periosteal stripping and bone exposure. Fractures are severely comminuted. Massive contamination is found with such wounds. Type IIIC wounds include any open fracture associated with arterial injury requiring repair, regardless of the extent of soft tissue injury.

Diagnosis of Fracture

In addition to a careful history and physical examination, diagnosing a fracture most often starts with plain radiographs of the skeletal area in question. It is essential to obtain views from at least two angles, preferably at 90 degrees from each other. Whenever possible, more than two views should be taken, which may increase the sensitivity of finding a fracture. Radiographs must be of optimal quality (proper exposure and angles) and include the entire bone or joint in question.

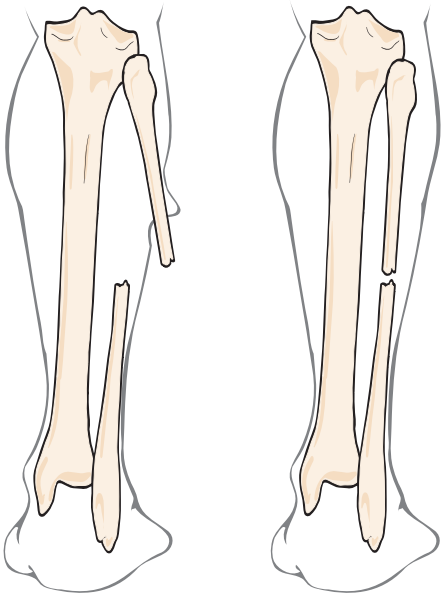


FIG 51.11 Comparison of open and closed fractures.

Occult nondisplaced fractures (including stress fractures) may not be seen on plain radiographs acutely. If a fracture is suspected but not easily seen, one may choose to immobilize the area as if a fracture was confirmed and then repeat the plain films in 1 to 2 weeks. Fine fractures not seen on acute films may become evident on later studies as the bone sclerosis can be seen at the site of healing. Alternatively, computed tomography (CT) or magnetic resonance imaging (MRI) can be very useful to reveal occult fractures or nondisplaced fractures in areas not easily seen on plain films (such as the scaphoid of the wrist or the femoral neck).

Treatment of Fracture

Initial management of a simple fracture may include icing, elevating, and immobilizing the affected limb as soon as possible to minimize soft tissue injury. Two main goals of treatment often include fracture reduction and immobilization. Fracture reduction means restoring the limb as close to the normal anatomic position as possible. This may need to be considered emergently if there are any concerns for neurovascular compromise or potential for ischemia to the distal limb. The purpose of immobilization is to maintain the proper alignment of the reduced fracture until adequate bone healing occurs.

Fracture immobilization may be accomplished through different means depending on the nature and severity of the fracture. For simple nondisplaced fractures, a hard cast may be the most useful. For injuries that involve considerable soft tissue damage, splinting may be preferable over a cast because of the risk of compartment syndrome (see "Complications of Fractures"). Functional bracing can sometimes be used to allow joint movement while still securing the bone (such as in some types of humeral fractures).

In some circumstances, surgical intervention to reduce and immobilize the injured bone may be preferred or even required, such as in cases of open fractures when debridement of the wound is needed or when multiple injuries are present. Intraarticular fractures often require surgical fixation to preserve joint motion, and fractures with severely displaced bone may require surgical reduction and fixation. In other instances, when a rapid return to mobility is desired, surgical fixation may have an advantage over nonsurgical treatment. Internal fixation may include plates and screws attached to bone or intramedullary nails placed within long bones. External fixation utilizes fixation devices applied temporarily until an injury or the patient is stabilized and a more permanent treatment initiated. This may be needed in the setting of multiple fractures, compromised blood supply, contaminated open fractures, or extensive surrounding tissue damage.

Further treatment options may also include bone grafting to bridge wider gaps in a displaced fracture or electrical bone stimulation, which utilizes negative current applied through the use of a bone stimulator to induce bone formation.

Other treatment considerations must include indications for antibiotic prophylaxis in the setting of open fractures, tetanus prophylaxis, pain medication, and eventual rehabilitation of the soft tissue structures surrounding the fracture.

Healing Process

When a fracture occurs, the continuity of both cortical and cancellous bone is usually compromised. The five stages of fracture healing are described in Chapter 50.

Healing in a cortical bone. At the time of fracture in a cortical bone, blood vessels in the haversian systems are torn. After a period of bleeding, clotting occurs at the fracture site and for a short distance on both sides of the fracture. Because of a lack of circulation, a small section of bone distal to the fracture site undergoes necrosis (Fig. 51.12). The avascular bone eventually is replaced by living bone through resorption and bone deposition. The majority of bleeding occurs from arteries in the periosteal sleeve.

The hematoma that forms becomes the medium for early stages of healing. Osteogenic cells, which develop from the periosteum, form the external and internal callus. If the periosteum is severely torn, healing cells must proliferate from the mesenchymal cells of surrounding soft tissue. During the early stages of repair, the amount of osteogenic tissue is extensive. Within the first few weeks, the thick mass of osteogenic tissue has formed a fracture callus.

During the initial stages of callus formation, no bone cells are present within the matrix. The callus is quite soft but becomes progressively firmer. With consolidation of the fracture callus, new bone formation begins. Initially, new bone forms at the edges of the periosteum, where the blood supply is more substantial. Where blood supply is sufficient, osteogenic cells differentiate into osteoblasts and primary woven bone. Near the fracture site, where the blood supply is less adequate, osteogenic cells initially differentiate into chondroblasts (cartilage).

As both the external callus (which unites cortical bone) and the internal callus (which unites cancellous bone) harden from the cartilage stage through ossification, the fracture site becomes firm and stable. No movement is detected by the medical evaluator or patient. At this point the fracture is clinically united. Although stable, cartilage and primary woven bone may be found intermixed at the site of healing.

With time, the primary callus is replaced by mature bone, and any excess callus is reabsorbed. This phase is the remodeling (last) stage of bone healing. When all immature bone cells have been replaced by mature lamellar bone, the fracture is said to be consolidated (radiographic union).

Healing in a cancellous bone. Cancellous fracture healing occurs mainly through development of an internal callus. The rich blood supply present in cancellous bone prevents necrosis of bone at the fracture site. If the fracture is nondisplaced, the healing process is much more rapid than that of cortical bone. Osteogenic cells in the trabeculae form the primary woven bone in the internal fracture hematoma (Fig. 51.13). The internal callus fills the open space of cancellous bone and crosses the fracture site. Woven bone develops and is eventually replaced

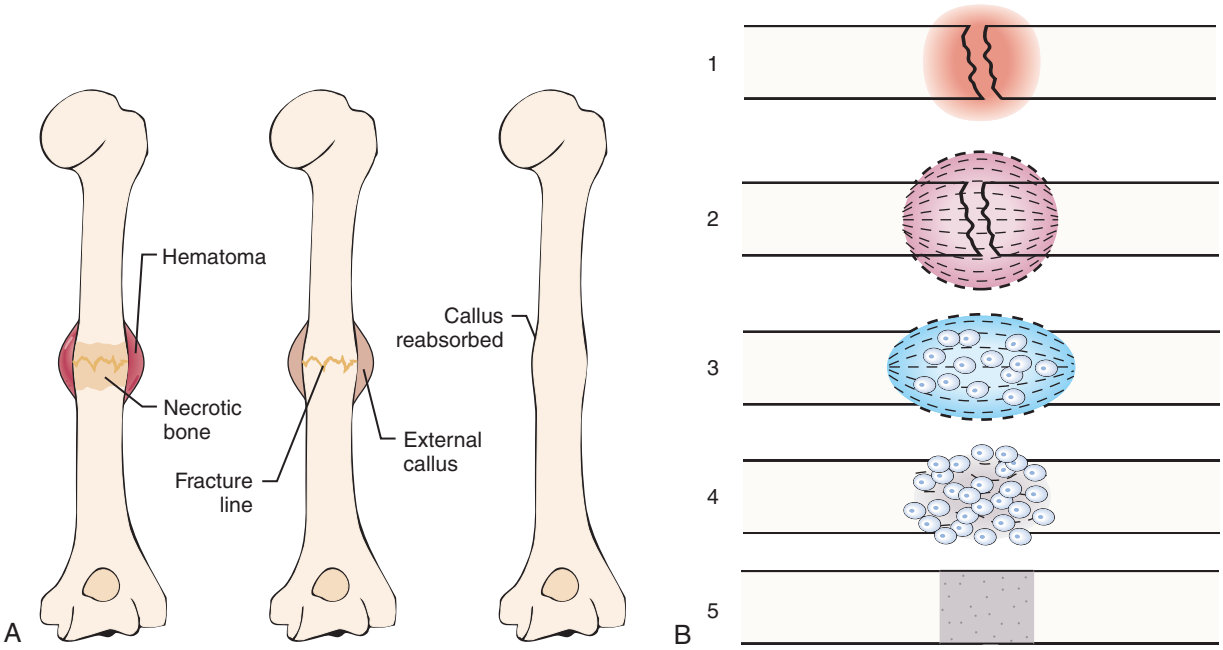


FIG 51.12 **A,** Stages of healing of cortical bone. **B,** Bone healing (schematic representation). *1,* Bleeding at broken ends of the bone with subsequent hematoma formation. *2,* Organization of hematoma into fibrous network. *3,* Invasion of osteoblasts, lengthening of collagen strands, and deposition of calcium. *4,* Callus formation: new bone is built as osteoclasts destroy dead bone. *5,* Remodeling is accomplished as excess callus is reabsorbed and trabecular bone is deposited. (B, From Lewis SM, et al: *Medical-surgical nursing,* ed 10, St Louis, 2017, Mosby.)

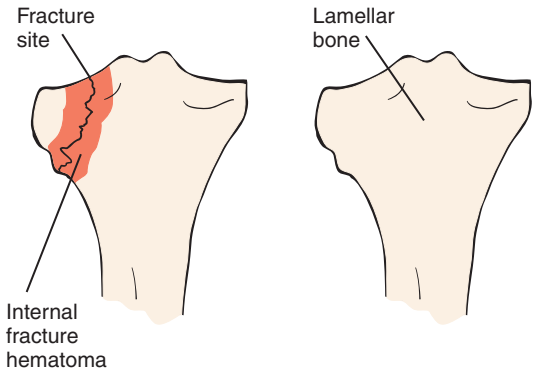


FIG 51.13 Healing of cancellous bone.

by lamellar bone. As noted earlier, cancellous bone is susceptible to compression forces, and the majority of injuries incurred are compression-type fractures. With a compression fracture, fragments of bone are impacted together, which provides a more suitable environment for healing of cancellous bone. Rapid union occurs because fracture fragments move in unison.

Complications of Fractures

Delayed healing. Fracture healing may not always progress smoothly without complications. Delayed union, malunion, and nonunion of the fracture are all complications that might occur. Delayed union is usually identified anywhere from 3 to 6 months after the fracture, when bone pain and tenderness are continuously increasing beyond the expected healing period for the wound type. Healing is slowed. The cause of delayed union might be either distraction of fracture fragments or the consequence of systemic causes such as infection. Bone healing can be delayed by additional factors such as smoking, malnutrition, use of corticosteroids, and poor vascular circulation to the area. Elderly patients or those with disease comorbidities, such as diabetes, coronary artery disease, peripheral vascular disease, or osteoporosis, are at particular risk.

Nonunion occurs when a fracture has not healed after 6 months. Failure to heal is due to poor blood supply and repetitive stress on the fracture site, and can be the result of interposition of muscle, tendon, or soft tissue between fracture pieces; prolonged or excessive traction; poor immobilization that allows motion at the fracture site; poor internal fixation; or wound infection after internal fixation.

Malunion results when unequal stresses of muscle pull and gravity lead to improper alignment of fracture fragments. It often happens in the case of fractures managed with cast immobilization after skeletal traction. Malunion may also occur if an ambulatory device is applied before the fracture is firm or if the extremity is subjected to weight bearing too early in the healing process. Primary features of malunion are external deformity and radiographic evidence of internal derangement. Prevention is accomplished by adequate reduction and immobilization of the fracture and adherence to specific activity and positioning restrictions.

Osteonecrosis. Osteonecrosis, also termed *avascular necrosis,* is a condition of compromised circulation to bone leading to ischemia and death of bone tissue. In the setting of a fracture, this can occur as a result of direct occlusion or injury of blood vessels because of fracture displacement or dislocations. Most common locations include the femoral head, proximal scaphoid, proximal humerus, and talar neck.

Osteomyelitis. Osteomyelitis, a severe bone infection, can occur after an open fracture as a result of direct invasion of bacteria into bone from a nearby contaminated soft tissue wound or direct penetration by a contaminated foreign body or contaminated surgical equipment (see "Infections of the Bone" for further discussion).

Compartment syndrome. Compartment syndrome is a result of the accumulation of pressure in a soft tissue compartment that is restricted by unyielding fasciae. This process can occur as a result of any type of severe soft tissue injury. Compartment syndrome may be classified as acute, chronic, or crush. In the case of severe soft tissue damage surrounding a fracture, the acute classification is of concern because it is the most severe form and often requires surgery urgently.

A compartment is a portion of the body where muscles, nerves, and blood vessels are enclosed within tissue such as fasciae. Compartment syndrome can be triggered by injury to the tissues surrounding bone with soft tissue inflammation, swelling, and, in some cases, hemorrhage into the area. This increase in edema causes an increase in pressure within the compartment attributable to the restriction of the surrounding compartment fasciae. Decreased blood flow from arterial damage can also occur, which leads to hypoxia of the cells of capillary walls. Capillary integrity is diminished, and colloid proteins and fluid escape into the extravascular tissues, causing further swelling and escalation of the intracompartmental pressures. If the tissue pressure exceeds the intravascular pressure, blood vessels will be collapsed, impeding blood flow and leading to further hypoxia and worsening edema. Intracompartmental pressures of 30 to 40 mm Hg can compromise microcirculation in muscle. The excessive compartment pressures lead to hypoxia, damage, and eventual necrosis of the soft tissue, especially muscles and nerves. Emergent decompression, such as by fasciotomy, is needed to preserve limb viability. Symptoms of compartment syndrome are pain out of proportion to the injury, paralysis, paresthesia, pallor, and pulselessness. Compartment syndrome is noted most often with injuries to the leg (anterior, deep posterior, superficial posterior, and lateral), forearm, upper arm (deltoid, biceps), hand (interosseus), and thigh (quadriceps). Compartment syndrome can also occur as a result of extrinsic compression, such as that of a cast on an injured, swollen limb.

Fat emboli syndrome. Fat emboli syndrome occurs when, after a fracture, fat particles are released from bone marrow (especially from the pelvis or long bones) into the bloodstream and lodge in the vasculature of the lungs. Often occurring within 24 to 72 hours of the trauma, fat emboli syndrome may be subclinical or may lead to respiratory failure and death. The patient may be noted to have shortness of breath, rapid breathing, hypoxemia, and a fine petechial rash (especially on the torso and neck area), as well as altered mental status or focal neurologic deficits. A plain chest x-ray may be normal, and often a spiral CT scan or ventilation-perfusion scan of the chest is more useful in diagnosis. Treatment is predominantly with ventilatory support. Early stabilization of the fracture reduces the risk of this complication.

Deep venous thrombosis and pulmonary embolism. Deep venous thrombosis and pulmonary embolism should be considered in a patient who develops chest pain, dyspnea, and hypoxemia more than 5 days after a fracture. This occurs when a thrombus forms in a distal extremity and clot fragments break loose to enter the circulation, thereby lodging in the lung circulation. Those with highest risk include patients with multiple traumatic injuries, patients with pelvic or long bone fractures who require more than 5 days immobilization in bed, obese patients, patients with a prior history of deep venous thrombosis, or patients with other risk factors for coagulation disorders. Diagnostic testing is similar to that of fat emboli syndrome, and treatment includes ventilatory support as well as anticoagulation therapy. Prevention of thrombosis often includes administration of anticoagulation prophylaxis, implementation of intermittent pneumatic compression devices, and early mobilization of the patient whenever possible.

Neurovascular injury. Neurovascular injury after a fracture may be due to either the fracture or the treatment for the fracture. Neurovascular damage occurring at the time of fracture may be the result of any of the following: the force causing the fracture, fracture fragments, hemorrhage, joint dislocation, or the body position assumed after trauma. Neurovascular damage related to treatment may be due to moving or splinting the fracture, manipulation at the time of reduction of the fracture, application of stabilizing devices such as a cast or splint, or the presence of hemorrhage or edema.

Dislocations and Subluxations

Two additional mechanical alterations in the musculoskeletal system are dislocations and subluxations. A dislocation is displacement of a bone from its normal position to the extent that articulating surfaces completely lose contact. A subluxation is displacement of a bone from its normal joint position to the extent that articulating surfaces partially lose contact. A dislocation or subluxation can occur when forces cause one aspect of the joint complex to move beyond its normal anatomic limit. A considerable amount of tissue damage occurs in dislocation and subluxation, including possible ligament tear or rupture. With any dislocation, especially first-time dislocation, evaluation for a fracture is necessary. Although almost any joint may dislocate, some are more prone to dislocation than others. Joints most commonly dislocated are small joints of the fingers, the patella, and the shoulder. Symptoms of dislocation are pain, alteration in the normal contour of the joint, change in extremity length, and loss of normal mobility. Treatment must include consideration of local soft tissue trauma and healing.

KEY POINTS

- Bones are subject to different types of fracture, depending on the type of tension stress imposed.
- Fractures can be classified according to the orientation of the break as transverse, longitudinal, oblique, or spiral.
- A comminuted fracture consists of more than one fracture line and more than two bone fragments.
- A greenstick fracture is an incomplete break.
- Fractures are classified as open or compound when the skin is penetrated and as closed or simple when the skin is not broken.
- Healing of fractured cancellous bone occurs more quickly than healing of cortical bone.
- Treatment goals of fractures are to minimize soft tissue injury, maintain proper alignment until adequate bone healing can occur, and prevent complications as much as possible.
- Trauma causes hematoma formation, followed by callus formation; the callus is initially soft and cartilaginous; then it progressively ossifies to become firm and stable.
- Radiographically apparent union occurs when the callus has been completely replaced by mature bone.
- Delayed union, nonunion, and malunion are potential complications of a fracture that does not heal in a normal time period and with proper alignment.
- Compartment syndrome after a fracture can result when a buildup of pressure occurs in a soft tissue compartment because of edema and inflammation within restrictive fascia layers. It is usually a surgical emergency.
- Symptoms of compartment syndrome include severe pain, pallor, paresthesias/paralysis, or pulselessness.
- Fat emboli syndrome and deep venous thrombosis with pulmonary emboli are two possible complications of a fracture that can lead to dyspnea, chest pain, hypoxemia, and respiratory failure.
- Complete separation of joint articulating surfaces is termed *dislocation.* Subluxation refers to partial separation. Soft tissue damage is the primary problem.

INFECTIONS OF THE BONE

Osteomyelitis

Osteomyelitis is a severe pyogenic infection of bone and local tissue that requires urgent treatment. Organisms may reach bone by one of three routes: (1) via the bloodstream (hematogenous osteomyelitis), (2) from adjacent soft tissue (contiguous focus), and (3) by direct introduction of the organism into the bone.

Etiology and Pathogenesis

Hematogenous osteomyelitis, in which the infectious agent may be introduced by blood from infection elsewhere in the body, is the most common type of osteomyelitis. It occurs most often in children younger than 16 years (mean age of 6 years old) and elderly adults, intravenous (IV) drug users, and patients with indwelling central lines. In children as well as infants, the long bones, which are rich in red marrow, are most commonly affected; and the infection usually begins acutely in the metaphyseal region of the bone. Bloodborne bacteria reach the marrow space via the nutrient artery, or after blunt trauma a hematoma develops; thus a pathway for the organism to reach the bone is present (Fig. 51.14).

Clinical Manifestations

In children, acute hematogenous osteomyelitis manifests as a high fever and pain at the site of bone involvement. The infection may remain localized if it becomes enclosed by fibrotic tissue reaction, a condition referred to as a *Brodie abscess.* Muscle spasms, redness, and swelling are common, and the child may refuse to move the limb. In adults, hematogenous osteomyelitis is more difficult to detect. Symptoms are vague and may include fever, malaise, anorexia, night sweats, and weight loss. Pain at rest is common. The diagnosis may be supported by radiographic signs of bone destruction. The most common causative organism is *Staphylococcus aureus* followed by *Streptococcus pneumoniae,* with gram-negative bacterial infections increasing in frequency. In children between 2 months and 3 years of age, *Haemophilus influenzae* can also be a cause, although this is quite rare since the development of a vaccination.

Osteomyelitis secondary to a contiguous focus of soft tissue infection can occur after burns, sinus disease, trauma, malignant tumor necrosis, periodontal infection, or an infected pressure ulcer. Again, *S. aureus* is the most common pathogen; however, some infections are polymicrobial and include gram-negative and anaerobic agents.

Direct invasion of the organism into bone can occur as a result of open fractures; penetrating wounds; surgical contamination; or insertion of surgical instrumentation such as prostheses, metal plates, or screws. Infections caused by surgical instrumentation can act as a focus for bacterial reproduction.

During the acute stage of osteomyelitis, bacteria remain in bone and proliferate where the circulation is not optimal. Before puberty the bacteria grow in the metaphyseal sinusoidal vein, which leads to infection of the metaphysis near the growth plate. The loose attachment of overlying periosteum permits exudate to accumulate in the subperiosteal area. Uncontrolled infection can disrupt the cortex and lead to joint infection or septic arthritis, which can cause osteoarthritis to develop later in life. In infants, medullary infection can reach the epiphysis and joint surfaces via capillaries crossing the growth plate, and stunted growth and angular deformities can result. The growth plate in children is avascular, so infection is limited at this site. The inflammatory reaction leads to pus formation, edema, and vascular congestion. Pus collects and is confined within bone, thus increasing pressure and adding to vascular occlusion, ischemia, and, finally, necrosis of bone. Volkmann and haversian canals allow a route for release of purulent material and thus spread of bacteria. Blood and therefore antibiotics cannot reach bone tissue when vascular system pressure equals arteriolar pressure. As a result, the course and virulence of the osteomyelitis are affected. Even after meticulous treatment, the organism can reappear years later in a context of trauma or immunosuppression.

Healing Complications

If osteomyelitis is not managed or if the treatment is not sufficient, the resulting necrotic bone can separate from healthy bone into dead segments called *sequestra.* A sequestrum is then a medium for the continued bacterial proliferation described as chronic osteomyelitis. Sequestra can enlarge and extrude through bone into soft tissue, where

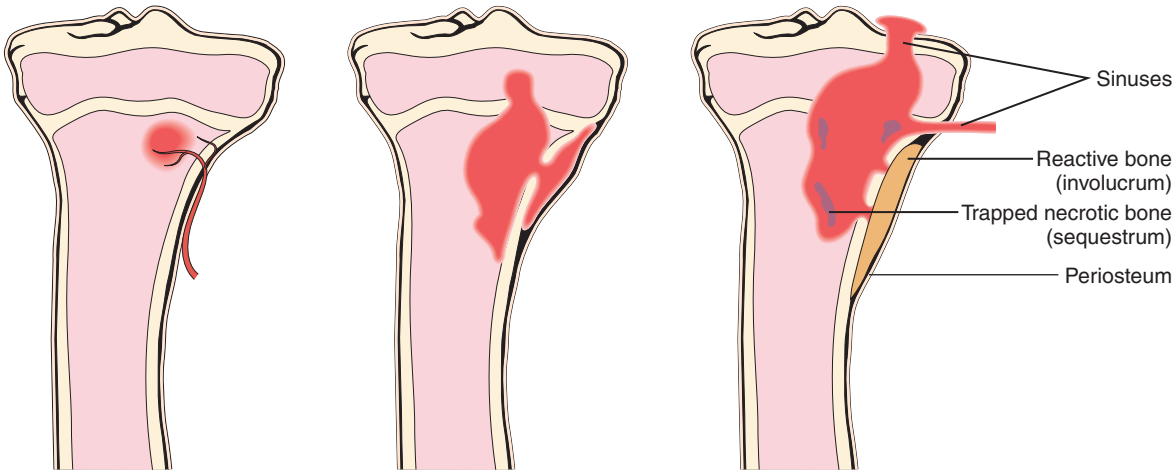


FIG 51.14 Osteomyelitis. The bacteria reach the metaphysis through the nutrient artery. Bacterial growth results in bone destruction and formation of an abscess. From the abscess cavity, the pus spreads between the trabeculae into the medulla, through the cartilage into the joint, and through the haversian canals of the compact bones to the outside. These sinuses traversing the bone persist for a long time and heal slowly. The pus destroys the bone and sequesters parts of it in the abscess cavity. Reactive new bone is formed around the focus of inflammation. (From Damjanov I: *Pathology for the health professions,* ed 4, Philadelphia, 2012, Saunders.)

it is possible that they might revascularize and resolve as a result of the body's defense mechanisms.

Osteoblasts may try to heal infected bone by isolating the dead tissue and forming an involucrum (a layer of new bone around old bone). Involucrum formation prevents successful effects of antibiotics and phagocytosis and leads to chronic infection.

Any type of osteomyelitis may become chronic, especially if the treatment was inadequate during the acute phase. It may be manifested months or years after assumed cure, especially after acute hematogenous disease. Drainage via a sinus tract to the skin can occur.

Treatment

Treatment usually includes 4 to 6 weeks of parenteral antibiotic therapy for acute osteomyelitis, although in children a typical course may be about 3 weeks in some cases. A shorter period of parenteral therapy followed by oral antibiotics can be effective if the infection is under control, the patient is afebrile, and a therapeutic blood level of the antibiotic can be maintained. Antibiotic choice is based on culture and sensitivity results.

If acute osteomyelitis is complicated by an abscess or extensive necrosis, the involved area is debrided and antibiotic therapy is instituted. After debridement, dead space is usually filled with packing, bone grafts, muscle pedicles, or skin grafts. In osteomyelitis associated with peripheral vascular disease, amputation is performed if antibiotic therapy is unsuccessful.

Tuberculosis

Etiology and Pathogenesis

Bone and joint tuberculosis (TB) is an extrapulmonary form of TB that occurs after lymphohematogenous or sometimes contiguous spread from a primary lung lesion. It is estimated to occur in about 1% to 5% of patients with pulmonary TB worldwide. As the incidence of TB increases in the United States, one can expect to see more cases of skeletal TB. Persons with skeletal TB may have a history of pulmonary TB, drug abuse, crowded and poor living conditions, diseases that depress the immune system, and immigration to the United States after 1991. Isolated musculoskeletal TB is not communicable to others unless an open wound exists. *Mycobacterium tuberculosis,* the organism responsible for the destruction of bone and joint, is transmitted via the airborne route. Initially, infectious droplets are inhaled and infect lungs; then *M. tuberculosis* spreads hematogenously from lungs or lymphatic drainage to bone. The bacterium may lie dormant for a long time before it is detected.

Clinical Manifestations

One of the more common sites of *M. tuberculosis*–infected bone (33% of skeletal TB cases) is in the vertebral column, particularly the lower thoracic and lumbar spine (called *Pott disease*). Infection often begins at the anterior portion of the vertebral body and then spreads further into the bone, causing bony destruction, anterior wedging, and collapse. On x-ray, the appearance is of a lytic lesion in the bone without local sclerotic (new bone formation) reaction. In about 50% of cases, the infection spreads to adjacent disks, and paraspinal fluid may accumulate as a "cold abscess" that can be seen on CT or MRI of the spine. Diagnosis may require culture of a tissue biopsy. Symptoms may include local pain, low-grade fever, and possible neurologic symptoms (weakness of lower extremities) attributable to local inflammation of nerve tissue and its surroundings as well as impingement.

Other common sites of skeletal TB or tuberculous arthritis include weight-bearing joints such as the hips, knees, and ankles, although any joint or bone could be involved.

Risk Factors

Persons most at risk for TB are those at extremes of age or individuals who are immunosuppressed or undernourished. Children are at higher risk for skeletal TB because of extreme vascularity. In AIDS, knowledge of the patient's human immunodeficiency virus status is critical to optimize the therapeutic plan.

Treatment

Treatment for skeletal TB requires long-term combination antibiotic therapy. Agents such as isoniazid, rifampin, pyrazinamide, ethambutol, and others are used in combinations for at least 6 to 9 months, with longer courses needed for certain drug combinations. Therapeutic response can be complicated by development of drug resistance. Surgical intervention may be indicated in cases of spinal TB when severe deformities or neurologic deficits are seen.

KEY POINTS

- Bone infections may be from bloodborne organisms or direct traumatic infection.
- With osteomyelitis, organisms reach the bone by one of three routes: bloodstream (hematogenous osteomyelitis), adjacent soft tissue (contiguous focus), and direct introduction of the organism into the bone.
- *S. aureus* and *S. pneumoniae* are the two most common organisms to cause bone infections in adults.
- Skeletal TB occurs when the TB infection spreads hematogenously or via the lung lymphatic drainage to bone and is most common in the vertebral bones, hips, and knees.
- Antibiotic therapy can help reduce the progression of bone infection and is the first-line treatment; however, abscess formation and chronic infection may occur, requiring surgical intervention.

ALTERATIONS IN BONE STRUCTURE AND MASS

BONE STRUCTURE DISORDERS

Scoliosis

Etiology and Pathogenesis

Scoliosis is a lateral curvature of the spine from the normal vertical line by 10 degrees or more that results in an S- or a C-shaped spinal column with vertebral rotation. Scoliosis can be a consequence of numerous congenital, connective tissue, and neuromuscular disorders. The majority of scoliosis cases are classified as idiopathic with a prevalence in the general U.S. population of about 2% to 4% with equal rates in men and women. The frequency is greater in children of women with scoliosis.

Clinical Manifestations

Scoliosis may be described as either nonstructural or structural. Nonstructural scoliosis resolves when the patient bends to the affected side. No vertebral rotation or bony deformity of the vertebrae is present, and the condition is not progressive. When the patient bends laterally, the spine usually appears symmetric. The scoliotic curve will disappear on forward flexion. Nonstructural scoliosis may be related to postural problems, nerve root irritation, inflammation, or compensation caused by leg length discrepancy or contracture in the cervical spine.

Structural scoliosis is a lateral curve of the spine that fails to correct itself on forced bending against the curvature and has vertebral rotation. Congenital scoliosis is present at birth, likely resulting from structural anomalies during fetal spine development. Infantile scoliosis is rare and

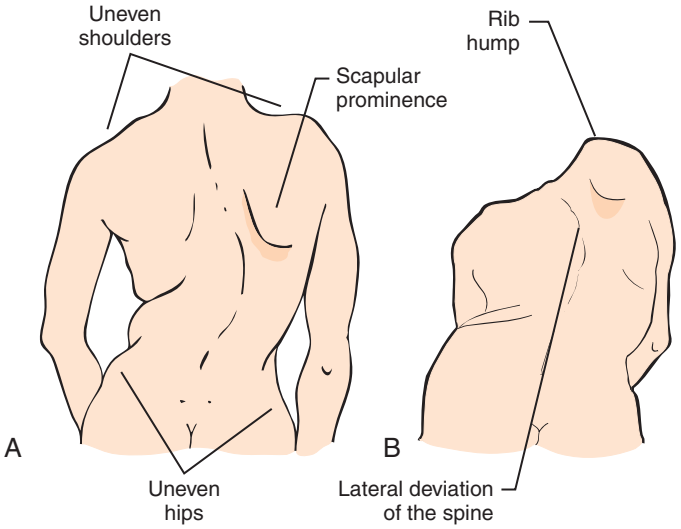


FIG 51.15 Structural scoliosis. **A,** The patient is standing erect, demonstrating the asymmetry of shoulder height as well as hip and scapular differences. **B,** The patient is bending forward at the waist, further emphasizing the spinal deviation and asymmetry of the shoulders and upper rib cage.

is first evident within the first 3 years of life. This form is likely genetic and leads to a lateral thoracic curve, which can sometimes resolve spontaneously later in life. Adolescent idiopathic scoliosis is the most common form and presents after 10 years of age. Mild scoliosis (less than 15 to 20 degrees) or scoliosis that does not progress (beyond 30 degrees) with skeletal growth may not require treatment. Progressive scoliosis that exceeds 45 to 50 degrees by skeletal maturity will likely worsen into adulthood. This type of scoliosis is more serious and involves deformity of the vertebrae and asymmetric changes in hip, shoulder, and rib cage positions. The patient lacks normal flexibility, and side bending becomes asymmetric. Severe structural scoliosis may require intensive therapy or surgical intervention to halt progression and correct deformities.

Scoliosis is detected by typical asymmetric changes (Fig. 51.15), including (1) uneven shoulders or hips, (2) shoulder or scapular prominence, (3) rib or chest hump when bending over, and (4) a C- or S-shaped spine. A scoliometer can be used to assess the angle of trunk rotation when the patient is bent in a forward flexion position. Scoliosis is usually diagnosed after puberty because of a tendency for the curve to be accentuated during periods of rapid skeletal growth. The diagnosis is confirmed by radiographic examination of the spine. The degree of curvature is determined from radiographs and is classified as right or left, depending on the direction of convexity. Surgery is indicated for curvatures of 40 to 50 degrees or greater, or in cases of significant progression of scoliosis in spite of bracing and other conservative therapies.

In addition to body image disturbances, scoliosis predisposes a patient to a number of physiologic problems. Respiratory difficulties from restricted expansion of the lungs may occur. Severe forms may be associated with significant pain. Gastrointestinal dysfunction can result from compression of abdominal organs. If uncorrected, scoliosis may progressively worsen with age due to increased upper body weight and gravitational forces exacerbating the vertebral deformity. Over time, significant degenerative changes to the intervertebral disks can occur.

Treatment

Treatment for structural scoliosis is aimed at correcting spinal malalignment. Nonsurgical measures include primarily braces and exercises. Bracing applies constant pressure to the spinal convexity to straighten the curve. Braces must be worn for prolonged periods each day to be effective. Compliance is a major difficulty because the braces are stiff and uncomfortable. Spinal muscle strengthening should accompany brace therapy because trunk musculature loses tone after prolonged bracing. Surgical intervention includes spinal realignment, fusion, and bracing with internal appliances, and most surgical procedures require prolonged body immobilization postoperatively.

Conditioning exercises to strengthen muscles and correct posture are used to treat nonstructural, or postural, scoliosis.

KEY POINTS

- Scoliosis is a lateral deformity of the spinal column that is detected from asymmetry of the shoulders, hips, and chest wall.
- Scoliosis may be described as structural or nonstructural.
- Severe scoliosis can compromise lung expansion and lead to a restrictive respiratory disorder.

METABOLIC BONE DISEASES

Osteoporosis

Etiology and Pathogenesis

Osteoporosis, the most common metabolic bone disease, affects an estimated more than 10 million people over 50 years of age in the United States, with incidence expected to increase over the next decade. Osteoporotic fractures, especially of the hip and spine, are associated with increased mortality, with an estimated 740,000 deaths per year worldwide as a result of hip fractures. Osteoporosis occurs when the rate of bone resorption is greater than that of bone formation, osteoblastic and osteoclastic balance is disrupted, and the levels of mineral and protein matrix components are decreased (see discussion on bone remodeling in Chapter 50). The number of trabeculae is decreased, and the width and mass of bone are reduced, which leads to fragile bone and thus fractures (Fig. 51.16). Cancellous bone is lost faster than cortical bone, with fractures happening earlier in cancellous bone (vertebrae) than in cortical bone (femoral neck).

A current definition of osteoporosis is based on gradations of bone mineral density (BMD), which is most commonly measured by dual-energy x-ray absorptiometry (DXA). This yields a T score, which is the number of standard deviations the patient's BMD measurement is greater than or less than the young normal mean BMD value. A Z score is the number of standard deviations greater than or less than the mean BMD value of age-matched controls. The World Health Organization (WHO) defines osteoporosis as bone mineral density (BMD) greater than or equal to 2.5 standard deviations below the mean peak BMD of young normal women (a T score less than −2.5). Osteopenia, which also carries an increased risk of fracture, is defined as a T score between −1.0 and −2.5. Because hip fractures are not uncommon in women with BMD in the osteopenic range, other risk factors for fractures and progression to osteoporosis must be considered. The WHO has developed a computer-based Fracture Risk Assessment tool (FRAX) that calculates a 10-year probability of hip fracture based on multiple variables, including age, body mass index, tobacco use, alcohol use, corticosteroid use, history of rheumatoid arthritis, and personal and parental history of fractures.

The specific cause of osteoporosis is not known. However, the rate of bone loss is influenced greatly by age, genetics, estrogen level, and risk factors. A family history of osteoporosis is a major risk factor. The normal bone loss that occurs with aging is accelerated during

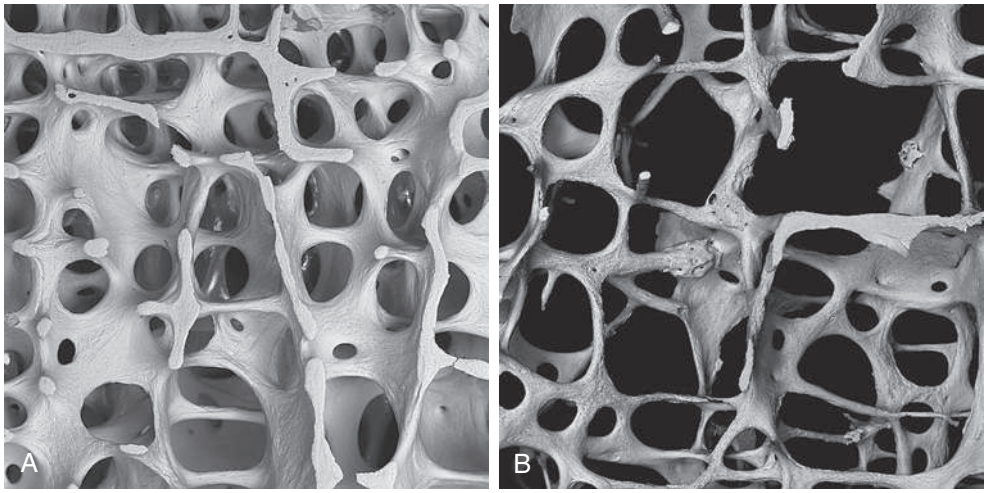


FIG 51.16 Micrographs of osteoporosis. On the left is normal bone. On the right is osteoporotic bone with decreased density and loss of trabeculae. (From Fillet HM: *Brocklehurst's textbook of geriatric medicine and gerontology,* ed 8, Philadelphia, 2017, Saunders. Courtesy of Professor A. Boyde, Department of Anatomy and Developmental Biology, University College, London.)

menopause, with the most rapid phase of loss occurring in the first 5 years because of the sudden decrease in estrogen concentration. The exact mechanisms of estrogen action are unclear, but estrogen derivatives may influence osteoblast activity and the production of local cytokines and growth factors that modulate the balance of bone resorption and formation. Estrogen deficiency increases the risk of osteoporosis by stimulating bone resorption over formation. Other risk factors include small frame, Caucasian or Asian race, early surgically induced menopause, high doses of thyroid hormone supplementation, use of corticosteroid drugs (such as prednisone), a diet low in sources of calcium and vitamin D, physical inactivity, and smoking or increased alcohol intake. Patients with chronic renal disease often have abnormal parathyroid function, as well as altered calcium and vitamin D metabolism, which can lead to a decline in bone mass. Furthermore, chronic inflammatory diseases, such as rheumatoid arthritis or systemic lupus erythematosus, can be associated with increased risk of osteoporosis, even independent of corticosteroid use (Box 51.1).

Clinical Manifestations

Evaluation of the patient must include an assessment of risk factors because osteoporosis is most often asymptomatic until a fracture occurs. On physical examination of someone with long-standing disease, the patient may have a Colles fracture of the wrist, femoral or hip fractures, or vertebral compression fractures. Obvious kyphosis of the thoracic spine (dowager's hump) may be present. The patient often has shortened stature, muscle wasting or spasms of back muscles, and difficulty bending over. The patient may complain of impaired breathing (because of deformities of the spine and rib cage) and poor dentition. Screening bone density measurement is also generally accomplished using DXA, assessing BMD of the hip (trochanter, femoral neck, and total hip), spine (L1 to L4 vertebral bodies), and sometimes the distal radius. Laboratory tests may show normal levels of urinary and serum calcium, phosphorus, and alkaline phosphatase but elevated serum osteocalcin levels. Other markers of bone turnover, such as cross-linked telopeptides and type I procollagen propeptides, can also be measured and are useful in the assessment and management of metabolic bone disease. Radiographic and CT findings may show diffuse radiolucency of bones, sparse transverse trabeculae, normal vertical trabeculae, indistinct articular cortices, wedge-shaped thoracic vertebrae, biconcave lumbar vertebral bodies, and possibly old or new compression fractures.

BOX 51.1 Risk Factors for Osteoporosis

Common Risk Factors
Female gender
Ethnicity (Caucasian or Asian)
Family history
Increased age
Low calcium intake (<400 mg/day)
Low body weight
Smoking
Alcoholism
Prolonged immobilization

Comorbid Diseases Increasing Risk
Rheumatoid arthritis
Cushing syndrome
Hyperthyroidism and thyroid replacement therapy
Hyperparathyroidism
Anorexia nervosa or exercise-induced amenorrhea
Chronic obstructive pulmonary disease
Ankylosing spondylitis
Celiac disease
Hypogonadism
Type 1 diabetes

Genetic Disorders Increasing Risk
Osteogenesis imperfecta
Menkes syndrome
Ehlers–Danlos syndrome
Marfan syndrome
Homocystinuria

Medications Increasing Risk
Corticosteroids
Thyroid hormone supplements
Heparin
Antiepileptic agents
Gonadotropin-releasing hormone agents
Aromatase inhibitors
Cytotoxic/immunosuppressive drugs

Treatment

Treatment varies depending on the cause. Moderate, regular exercise such as walking or riding a stationary bicycle is valuable in prevention as well as treatment of osteoporosis. Physical therapy exercises for individuals who are immobilized or paralyzed are helpful as well.

Calcium and vitamin D. Both calcium and vitamin D are necessary for maintenance of bone mass. Calcium is needed as a constituent of bone, and vitamin D is essential for increasing intestinal absorption of calcium and for calcium uptake into bone. In vitamin D deficiency states, calcium absorption can decrease from a normal rate of 30% to 40% to just 10% to 15%. Vitamin D must be metabolized to its active form, 1,25-dihydroxyvitamin D, which is also regulated by the kidneys. Although sun exposure can increase vitamin D production in skin, this source is often inadequate, even in warm climates. Recommendations for calcium supplementation vary, but most experts recommend 1000 to 1500 mg daily—preferring calcium citrate, which does not depend on an acidic environment in the gut for absorption. Vitamin D deficiency can be detected by measurement of blood levels, with a goal of greater than 20 mg/mL. Many believe a level above 30 mg/mL not only increases BMD but may also decrease risk of falls by a stimulatory effect on muscle function. The recommended doses of daily vitamin D_3 are 400 to 1000 IU; also, in a deficient state 2000 to 4000 IU daily appears to be safe.

Antiresorptive agents. The most often used antiresorptive agents are the bisphosphonates (alendronate, risedronate, ibandronate, and zoledronate), which increase BMD and decrease fracture risk by inhibiting bone resorption by osteoclasts while osteoblast-related bone formation continues. These agents bind to hydroxyapatite in the bone, thus blocking the enzymes needed for osteoclast cell structure stability during resorption. Questions about long-term use (>5 years) of bisphosphonates have arisen with the concern that the longevity of bisphosphonate binding into the bone may decrease bone's ability to repair microtrauma and be associated with atypical fractures. One such complication is osteonecrosis of the jaw, which, although very rare, has been reported particularly in patients receiving IV bisphosphonates for the treatment of bone cancer. Another rare complication under study is the atypical femoral fracture, which occurs below the level of the trochanter and also may be at slightly increased risk in patients taking bisphosphonates for prolonged periods. Ultimately, the risk of a fracture attributable to untreated osteoporosis is 7- to 10-fold higher than the risk of these rare complications. It is recommended that after 5 years of bisphosphonate therapy, a drug "holiday" of 2 years or more be considered with subsequent BMD testing being used to determine need for further treatment.

Other agents for the treatment of osteoporosis include teriparatide, a recombinant human parathyroid hormone, which is the only true anabolic agent. Parathyroid hormone, when given at intermittent doses, increases BMD by stimulating bone turnover with the rate of bone formation exceeding that of resorption. It is approved for daily subcutaneous administration, but its use is limited to 2 years because rat studies have shown an increase in risk for osteosarcomas. Denosumab is a monoclonal antibody that inhibits the receptor activator of nuclear factor κB ligand (RANKL). By binding and inhibiting RANKL, osteoclast activity is suppressed and bone resorption is decreased. This agent is given subcutaneously every 6 months.

Other Causes of Osteoporosis

Disuse osteoporosis may occur with prolonged bed rest, which leads to an increase in osteoclast activity and resorption of bone greater than osteoblast accumulation. Stress placed on bone as a result of weight bearing is necessary for osteoblast function. The stress of exercise stimulates new bone growth as a result of changes in electrical charges on the bone surface.

The loss of bone density after a period of reduced weight bearing may be restored upon return to normal activity, although recovery may not be complete. Osteoporosis may also occur when collagen formation is impaired in such conditions as scurvy (severe vitamin C deficiency), protein deficiency, or Cushing syndrome.

Rickets and Osteomalacia

Clinical Manifestations

Rickets and osteomalacia are characterized by deficits in mineralization of newly formed bone matrix either in the growing skeleton (rickets) or in the mature skeleton (osteomalacia) with resulting soft osteopenic bone. Deficiency of vitamin D prevents maintenance of normal levels of calcium and phosphorus. Children may have either vitamin D–resistant rickets or congenital hypophosphatasia. In rickets, cartilage in the growing epiphyses fails to calcify. Cartilage is not replaced by bone and continues to enlarge, leading to widening of epiphyseal plates and irregularity of the junction with the metaphyses. Bone is poorly calcified and less rigid. Kyphosis, genu valgum ("knock knee"), and genu varum ("bowleg") are common deformities as well as growth retardation. Delayed eruption of the teeth, enlargement of costochondral junctions, and decreased muscle tone can all be seen.

Osteomalacia is the adult counterpart of rickets with defects occurring after closure of epiphyseal plates. Osteomalacia is always due to an inadequate concentration of vitamin D, calcium, and/or phosphorus in the body as a result of any of the following factors: decreased intestinal absorption of vitamin D and calcium attributable to poor intake or malabsorption, poor vitamin D metabolism attributable to decreased sun exposure, renal disease (especially chronic renal failure or nephrotic syndrome), or a combination of these conditions. In the case of vitamin D deficiency, calcification fails to occur and the bone is soft. Patients may complain of bone pain and muscle weakness, and plain x-rays may show bowing of bones and "pseudofractures." All bones are affected, but weight-bearing structures may collapse and cause compression-type fractures.

Treatment

Treatment involves correction of the underlying deficiency with adequate intake of vitamin D supplementation (especially vitamin D_3). Adequate intake of calcium and phosphate is also indicated by dietary adjustments or supplements.

Paget Disease

Paget disease of bone (osteitis deformans) is a slowly progressive metabolic bone disease characterized by an initial phase of excessive bone resorption, mediated by osteoclasts, followed by excessive bone formation (Fig. 51.17). The end product is a disorganized mosaic of bone matrix composed of woven and lamellar bone at affected sites of the skeleton. This new bone is less compact, more vascular, and more fragile, which accounts for the deformities and fractures of Paget disease.

Etiology and Pathogenesis

The specific cause of Paget disease is unknown, but a genetic component is suggested. It affects bones of adults, rarely being seen on x-rays of patients under 40 years of age. It is prevalent in parts of northern Europe but rare in Africa or Asia. In the United States, abnormal pelvis films affected by Paget disease were found in about 2% of individuals 55 to 74 years old. Residents of northeastern states and Caucasian individuals were more frequently affected.

It has also been theorized that a viral infection may affect osteoclastic function, leading to aberrant bone remodeling, a theory based on mouse models of study. Confirmation in this research remains elusive. Changes

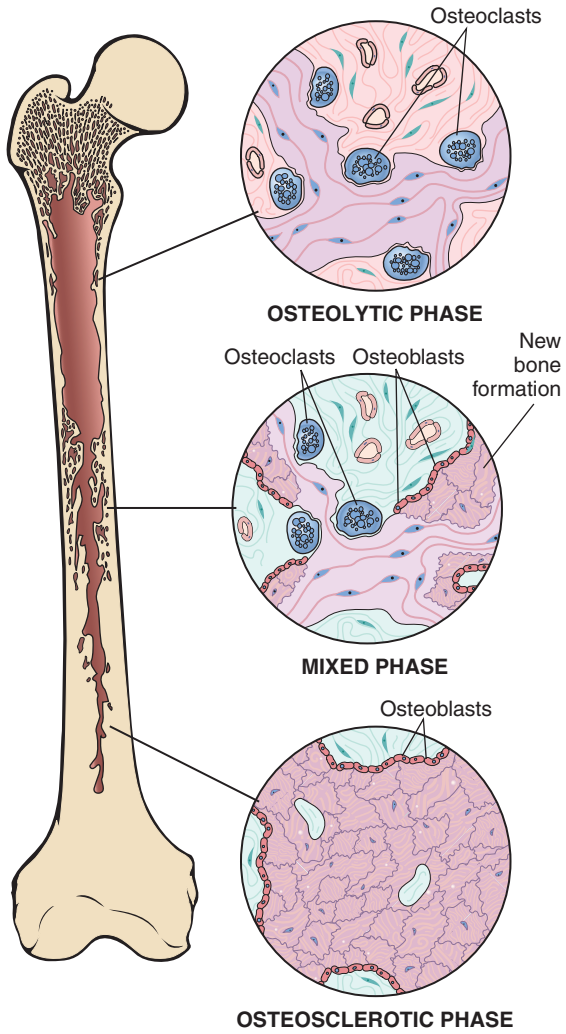


FIG 51.17 Diagrammatic representation of Paget disease of bone demonstrating the three phases in the evolution of the disease.

in certain cytokines produced in the local bone marrow environment may also influence the bone formation/resorption balance.

Clinical Manifestations

In the early stages the disease may not cause any symptoms; however, when pain develops, it can be severe and persistent bone pain. Fatigue and joint stiffness are also noted. In the initial stages of Paget disease, affected bones soften and tend to bend. As the disease progresses, irregular subperiosteal bone formation occurs and causes bone to become thick and hard. Thickening of cranial bones may cause compression of cranial nerves and result in vertigo, blindness, deafness (with or without tinnitus), headaches, and facial paralysis. Any bone can be affected, but the most common sites include the sacrum and spine (50%), femur (46%), skull (28%), and pelvis (22%). Other complications may include hypertension, arthritis, calcific periarthritis, and pain.

Treatment

During active stages of the disease, treatment focuses on reducing pain and preventing deformity and fracture, often with the use of calcitonin or bisphosphonates (such as alendronate, risedronate, or pamidronate, among others). These medications have been shown to decrease bone resorption, stabilize the fragile bone lesions, and reduce pain and the risk of fractures.

KEY POINTS

- Bone density is a product of the rate of bone resorption and bone deposition.
- Osteoporosis occurs when the rate of bone resorption is greater than that of bone formation. A reduction in bone mass predisposes to fractures.
- Osteoporosis in women is defined by the WHO as a bone mineral density ≥2.5 standard deviations below the mean peak BMD of young normal women.
- Hormone deficiencies (estrogen, androgen), poor calcium intake, and inadequate muscle use are common factors in the rate of bone loss.
- Treatment of osteoporosis should include adequate calcium and vitamin D supplementation along with a choice of antiresorptive agents (e.g., bisphosphonates), anabolic agents (e.g., teriparatide), or other inhibitors of bone resorption (e.g., denosumab) among others.
- Vitamin D deficiency is associated with rickets and osteomalacia, disorders characterized by soft, weak bones.
- Paget disease may be genetic. It has also been theorized that a viral infection may affect osteoclastic function, leading to aberrant bone remodeling. Painful deformities or bone fractures may result.

BONE TUMORS

Neoplasms occurring in the musculoskeletal system can be benign or malignant with a relative incidence ratio of benign to malignant lesions of 200 to 1. Benign tumors often are undiagnosed because they cause no pain. These lesions typically do not grow aggressively or metastasize; also, they do not tend to recur. Malignant neoplasms that originate in bone are referred to as *sarcomas.* These lesions can be very destructive, tend to regrow, and may metastasize. More common than sarcomas are metastatic lesions, which have spread to bone from a primary tumor elsewhere. Primary carcinomas that most commonly metastasize to bone are breast, prostate, lung, and kidney carcinomas. Other malignancies that can metastasize to bone include cancer of the thyroid, bladder, uterus, colon/rectum, and vagina. Common sites of bone metastases are the vertebral bodies, pelvis, proximal ends of the femur and humerus, and ribs. Metastases occur via direct spread within a body cavity or by hematogenous or lymphatic spread.

Thus although the majority of bone tumors are metastatic, a number of primary tumors of bone can be identified. A few representative types are discussed, including some benign and malignant tumors (Fig. 51.18).

Benign Tumors

Osteochondroma

Etiology, pathogenesis, and clinical manifestations. Osteochondroma is a common cartilage-forming benign tumor that is most often asymptomatic and may not be discovered until adulthood. Osteochondromas can be hereditary and are often found unintentionally.

The lesion arises from a growth plate defect that can become pedunculated or sessile. Bony projections on the external surface of the bone are capped with cartilage. Pressure on surrounding soft tissue may cause pain. These tumors are usually located on the metaphyses of long bones such as the proximal end of the tibia and the distal part of the femur, the shoulder, and the pelvis (see Fig. 51.18).

Chondroma

Chondroma or enchondroma is a cartilage-forming tumor in bone that can be located in the medullary cavity or in the subperiosteal layers of bone. It is believed to arise from remnants of epiphyseal cartilage. Chondromas develop most often in the small bones of the hands and feet but can be found in other areas. Tumor growth may erode the cortex of bone and expand the contour. Chondromas may be found incidentally, sometimes not until adulthood.

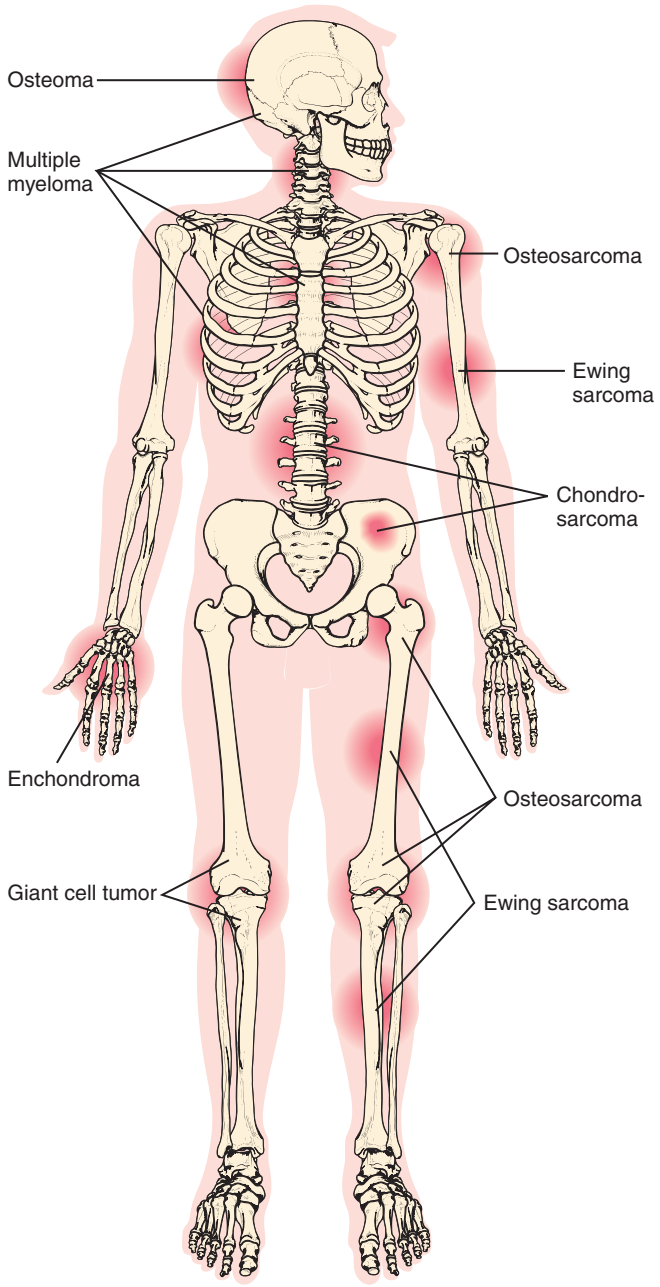


FIG 51.18 Schematic presentation of the most common sites of origin of bone tumors. Most often, osteosarcomas originate in the metaphyses of long bones, chondrosarcomas arise in the axial skeleton, Ewing sarcomas develop in the diaphyses of long bones, and giant cell tumors originate in the epiphyses of long bones. Osteomas occur most often in the skull, and enchondromas in the small bones of the hand. Multiple myelomas involve the calvaria, vertebrae, and ribs, but also other bones that contain hematopoietic bone marrow. (From Damjanov I: *Pathology for the health professions,* ed 4, Philadelphia, 2012, Saunders.)

Osteoid Osteoma

Osteoid osteoma is one of the more common types of benign bone-forming tumor and accounts for approximately 10% to 13% of symptomatic benign lesions. The patient often complains of persistent, dull pain, which is often worse at night and alleviated by aspirin or other nonsteroidal antiinflammatory drugs. This small lesion is often found in the cortex of the tibia and femur, but any bone may be involved. Radiographs show the lesion enclosed in a sclerotic shell. This tumor usually occurs in persons in their 20s.

Osteoblastoma is a form of osteoid osteoma that is larger and has a more lytic appearance. These are less common than osteoid osteomas and tend to occur in the spine (although they can also be seen in long bones). The lesion is benign but can be destructive with a richly vascularized mass of immature osteoid and microtrabeculae lined with plump osteoblasts. Multinucleated osteoclast-like cells are also present.

Giant Cell Tumor

Many different tumors can contain giant cells but are not true giant cell tumors. A giant cell tumor, or osteoclastoma, is benign but aggressive with richly vascularized tissue consisting of plump spindle-shaped cells and numerous giant cells. These lesions account for about 5% to 10% of all primary bone tumors. In some cases, giant cell tumors undergo transformation to sarcomas, a complication that is thought to be related to exposure to radiation therapy. In 1% to 2% of cases, the cells can metastasize without malignant transformation, in particular to the lung. Giant cell tumors commonly occur during the third decade of life and are slightly more common in females. The area of development includes the distal end of the femur, proximal part of the tibia, distal part of the radius, and proximal end of the humerus (see Fig. 51.18). Pain is the initial complaint.

Malignant Bone Tumors

Osteosarcoma

Etiology and pathogenesis. Osteosarcoma, an extremely malignant bone-forming tumor, is the most common primary malignant bone tumor (aside from multiple myeloma) and accounts for 20% of all primary malignant bone cancers. The majority of patients are adolescents and young adults 20 to 30 years of age, although cases can be seen in adults 60 to 70 years old. It develops in the metaphyseal region of long bones and is characterized by the formation of bone or osteoid by tumor cells. The most active epiphyseal growth areas—the distal end of the femur, proximal end of the tibia, fibula, and proximal end of the humerus—are common sites of involvement (see Fig. 51.18). Lesions can also be seen in flat bones of the pelvis, skull, scapula, ribs, or spine.

Clinical manifestations. Osteosarcoma grows rapidly and is quite destructive; destruction of the cortex of the metaphyseal region predisposes it to pathologic fracture. Metastasis to lungs can be noted early in disease development and portends a much poorer prognosis. Limb pain may occur very early in the disease and become consistent and progressively more intense. Joint function may be compromised as a result of the proximity of the metaphysis.

Treatment. Although radical amputation was the only treatment previously employed, conservative surgery and chemotherapy are currently providing positive results, with studies showing 5-year disease-free survival rates near 70%.

Chondrosarcoma

Pathogenesis. A primary chondrosarcoma is a malignant cartilage-forming tumor most often diagnosed in adults 30 to 60 years old with a higher predominance in women. These tumors usually develop slowly and have a higher cellularity and greater pleomorphism than a chondroma.

Secondary chondrosarcomas arise from benign lesions such as osteochondroma or multiple enchondromatosis, which undergo malignant transformation. These tumors develop in the pelvis and proximal ends of the femur and humerus in individuals between 20 and 40 years of age. Evidence of malignant transformation may include pain, an irregular border, or an increase in the proximal end of long bones after patient growth is complete.

Clinical manifestations. Because of the slow growth of the tumor, pain is not usually a prominent clinical symptom initially. Even with a slow rate of development, the tumor will eventually metastasize, typically to the lung. Chondrosarcomas tend to develop in the pelvic

and shoulder girdles, as well as the ribs and the proximal ends of long bones such as the femur (see Fig. 51.18).

Ewing Sarcoma

Pathogenesis. Ewing sarcoma is the third most common primary sarcoma of bone and is characterized as a rapidly growing, malignant, round cell tumor. This tumor most often develops in the bones of children and young adults between the ages of 5 and 25 years with higher frequency in males than females. Ewing tumor is composed of densely packed small cells with round nuclei. It arises in the medullary canal of bone and perforates the cortex of the shaft, producing a painful soft tissue mass (with a central lytic area) overlying the involved bone. Lesions can often be mistaken for a focus of osteomyelitis. The tumor favors pelvic bones followed by long tubular bones, such as the femur, tibia, humerus, or scapula, although any bone can be involved (see Fig. 51.18).

Clinical manifestations and treatment. Ewing sarcoma metastasizes quite early in its development to the lungs and other bones, with 15% to 25% of cases being metastatic at diagnosis. Because of the rapid rate of growth, pain is a dominant symptom that increases in severity. Ewing sarcoma is often confused with osteomyelitis because patients often appear systemically ill and may develop fever, anemia, leukocytosis, and an increased sedimentation rate. Treatment with local resection and chemotherapy of an isolated lesion can have a 5-year survival approaching 70%. In patients with metastatic disease at diagnosis, 5-year survival drops to 30% even with surgery and radiotherapy or chemotherapy.

Multiple Myeloma

Etiology and pathogenesis. Multiple myeloma, which is the most common primary tumor of bone, is a slowly growing bone marrow malignancy with neoplastic proliferation of a single clone of plasma cells. The annual incidence is approximately 4 per 100,000 and represents about 1% of all malignant cancers. It is usually a disease of elderly adults.

Clinical manifestations and treatment. Although multiple myeloma is not a sarcoma, its symptoms and radiographic findings are similar. On radiographs, evidence of bone destruction by a lytic, or bone-destroying, process and bone marrow involvement can be seen. Homogeneous immunoglobulin is also present in urine or serum. Because multiple myeloma is a slow-growing lesion, it takes a long time to become symptomatic. Bone pain is the most common symptom, particularly of the chest and back, and is related to excessive accumulation of abnormal plasma cells in the bone marrow (Chapter 11). Although it can affect any bone, multiple myeloma most commonly occurs in the thoracic and lumbar vertebrae. Patients experience hypercalcemia and pathologic fractures where bone has been destroyed. This disease can also cause kidney dysfunction, lung or pleural involvement, and neurologic symptoms attributable to nerve compression. Treatment often requires aggressive combination chemotherapy, although at times local radiation or use of bisphosphonates may be useful for refractory bone pain. Chemotherapy can induce a temporary remission in 50% to 70% of cases, but survival longer than 2 to 3 years is less than 10%.

KEY POINTS

- Primary tumors of the bone are not common. Osteochondroma, chondroma, osteoid osteoma, osteoblastoma, and giant cell tumors are benign primary bone tumors.
- Malignant bone tumors include osteosarcoma, chondrosarcoma, and Ewing sarcoma.
- Multiple myeloma is a slow-growing bone marrow malignancy in which plasma cells proliferate. This disease affects the kidneys and the immune and circulatory systems.

DISEASES OF SKELETAL MUSCLE

Skeletal muscle, the most abundant tissue in the human body, accounts for approximately 40% of total body weight. Skeletal muscle performs dynamic work (locomotion) and static work (posture). As with other tissue of the musculoskeletal system, muscle atrophies in response to disuse and immobilization, and hypertrophies when subjected to increased stress.

IDIOPATHIC INFLAMMATORY MYOPATHY

Polymyositis and Dermatomyositis

Polymyositis and dermatomyositis are idiopathic inflammatory myopathies. With these diseases there is focal or extensive degeneration of muscle fibers attributable to inflammatory infiltrates of lymphocytes, macrophages, and other inflammatory cells. In addition to producing proinflammatory cytokines and autoantibodies, infiltrating cells can have a cytotoxic effect on muscle cells. Eventual necrosis of muscle fibers followed by fatty replacement of muscle leads to significant weakness. Possible triggers of this immune-mediated inflammation include viruses, bacteria, parasitic organisms, neoplasms, drugs, vaccinations, and stress.

Proximal limb and neck weakness and associated muscle stiffness are clinical signs of these illnesses. Muscle pain is often mild. Most patients initially complain of hip and leg weakness and difficulty with climbing stairs and rising from a chair. Later in the progression of the disease, weakness in the arms prevents functional overhead activity. Anterior neck weakness makes lifting the head from the pillow very difficult. When classic skin changes occur with polymyositis, the disease is classified as dermatomyositis. Inflamed, injured skeletal muscle leaks several enzymes (creatine kinase, aldolase, aspartate aminotransferase) into the bloodstream and are found at high levels on laboratory testing.

Clinical Manifestations

During the physical examination, manual muscle testing reveals weakness in the proximal limb muscles. Contractures are not usually present, but they may develop later. Facial and ocular muscle weakness seldom occurs, distinguishing myositis from myasthenia gravis, although esophageal dysmotility can be seen. Diagnosis of myositis can be detected by electromyography and by the presence of inflammatory changes and muscle edema on MRI. Diagnosis is confirmed on skeletal muscle biopsy of involved areas.

In dermatomyositis, cutaneous manifestations may develop with muscle involvement. Common findings are flat-topped papules overlying the dorsal surface of the small joints of the hands (Gottron papules). A more common finding is development of an erythematous smooth or scaly patch over other joints such as the elbow, knees, or medial malleoli areas (Gottron sign). Other skin changes may include the heliotrope rash (violaceous edema of the eyelids) or a dry erythematous rash over the anterior chest wall (V sign) or the upper back and shoulders (shawl sign).

Cardiac involvement may occur but is often mild and asymptomatic. Dysrhythmias can occur, but development of congestive heart failure or pericarditis is less common. Weakness of the respiratory muscles can cause dyspnea, as can interstitial lung disease, which can be severe in rare cases.

Treatment

Initial therapy with corticosteroids is usually used to decrease muscle inflammation and preserve function. Immunosuppressive agents, such as methotrexate, azathioprine, or mycophenolate mofetil, may be helpful in severe cases. Physical therapy is also important and should include passive range-of-motion activities, followed by assisted and then active strengthening exercises.

MUSCULAR DYSTROPHY

Muscular dystrophy comprises a group of genetically determined myopathies characterized by progressive muscle weakness and degeneration as muscle tissue is replaced by fat and fibrous connective tissue. The dystrophies are classified by their pattern of inheritance, age of onset, and distribution of muscular weakness.

Duchenne Muscular Dystrophy

Etiology and Pathogenesis

Duchenne muscular dystrophy, the most common and most severe form of muscular dystrophy, is inherited as an X-linked trait and therefore afflicts only males. The incidence is 1 in 5000 male births. Because of a genetic mutation, muscle cells are deficient in the protein dystrophin, a deficiency that weakens the cell membrane and allows extracellular fluid to leak into the cell. Proteases and inflammatory processes are activated, leading to muscle fiber necrosis and muscle degeneration.

The disease begins at birth and is usually apparent by the age of 2 to 5 years, with initial involvement of the pelvic girdle and progression to the shoulder girdle.

Clinical Manifestations

The calf muscles of an individual with Duchenne muscular dystrophy are noticeably enlarged because of the infiltration of fat cells and degeneration of muscle fibers. Distal muscle involvement leads to frequent falling by the age of 5 or 6 years, and by age 12 to 14 years most children are confined to a wheelchair. Some muscles, such as those in the hands, face, jaw, pharynx, larynx, and eyes, are spared to the end. Survival expectancy is to late 20s or early 30s with cardiac failure or pulmonary infection as the usual cause of death.

Treatment

Treatment of muscular dystrophy is focused on appropriate education for the patient and family, preservation of physical function as long as possible, and prevention of contractures. In some cases, corticosteroid therapy may be useful to delay loss of muscle strength and prolong independent ambulation, but eventual steroid-induced osteoporosis is a risk. Medical management of the complications of cardiomyopathy and declining pulmonary function may also be needed. Immunosuppressive therapies have also been tried but with limited success.

Becker Muscular Dystrophy

Etiology, Pathogenesis, and Clinical Manifestations

Becker muscular dystrophy is a milder form of inherited muscle degeneration, somewhat less common than the Duchenne type. The genetic mutation leads to production of a reduced amount of an abnormal dystrophin protein and a slower muscular degeneration. Calf hypertrophy is still prominent and often painful with progressive loss of strength and ability to ambulate. The mean age of symptom onset is somewhat later (older than 5 years and even into adulthood) with patients requiring a wheelchair by the age of 30 years.

Facioscapulohumeral Muscular Dystrophy

Etiology and Pathogenesis

Facioscapulohumeral muscular dystrophy is an inherited autosomal-dominant trait that affects the muscles of the shoulder girdle and the face. It is rare, with an annual incidence of 1 in 15,000. The onset of disease can occur at any age, but it usually begins in the second decade. Facial muscles are involved early, with later involvement of scapular and upper arm musculature. Associated symptoms may include sensorineural hearing loss or retinal vascular abnormalities. It progresses slowly with periods of arrest and can ultimately involve more distal muscles of the upper and lower extremities. Both males and females are affected, and with supportive management, most live to a normal age.

Myotonic Dystrophies

The myotonic dystrophies are the second most common inherited muscle disease, and it has an autosomal-dominant pattern of inheritance. Prevalence is about 1 in 8000. There are two types of myotonic dystrophy, both of which are caused by chromosomal aberrations in the gene that encodes for dystrophia myotonica protein kinase. This leads to progressive muscle weakness and myotonia with delayed muscle relaxation after stimulation.

Clinical Manifestations

Type 1 myotonic muscular dystrophy (MMD1) most commonly affects facial muscles, distal limb muscles, and oropharyngeal and extraocular muscles. Distal limb weakness can develop slowly as well as muscle wasting. In the type 2 form (MMD2), the proximal limb muscles are more affected, sparing the face. Other complications in both types can include cardiac conduction defects, cataract formation, endocrine dysfunction, testicular atrophy, hypersomnolence, and sleep apnea. Treatment is largely supportive along with management of complications, although the antiarrhythmic drug mexiletine functions as a sodium channel blocker and has been shown in limited studies to improve muscle relaxation.

OTHER DISORDERS OF MUSCLE

MYASTHENIA GRAVIS

Myasthenia gravis is a chronic autoimmune disease affecting the neuromuscular function of voluntary muscles and characterized by profound muscle weakness and fatigability. Its peak onset in females occurs at 20 to 30 years of age, although onset later in life is becoming more widely recognized. Women are affected more often than men, with a prevalence of 15 per 100,000. Characteristically, painless muscle weakness begins with ocular and cranial muscles, and then proximal limb muscles can be involved. The weakness is exacerbated by repetitive muscle use and improves with rest. Severity of symptoms can vary week by week. During times of emotional stress, respiratory muscles may be involved.

In myasthenia gravis, acetylcholine receptor antibodies are produced that destroy or block acetylcholine receptors of the muscle end-plate of the neuromuscular junction. These antibodies impair the transmission of acetylcholine activity across the junction. The result is the muscle weakness and fatigability so prevalent in this disease.

Myasthenia crisis can be due to insufficient medication, emotional stress, trauma, infection, or surgery. A sudden increase in blood pressure and pulse rate is noted. Other symptoms include cyanosis from hypoxia, absent cough and gag reflexes, restlessness, increased secretions and lacrimation, diaphoresis, decreased urine output, bowel and bladder incontinence, dysarthria, and respiratory distress.

Cholinergic crisis is usually due to excessive medication. Patients experiencing such a crisis will have fasciculations, especially around the mouth; difficulty chewing, swallowing, and speaking; advancing muscle weakness approximately 1 hour after anticholinesterase medication; nausea and vomiting; cramps and diarrhea; increased secretions (salivary, perspiration, lacrimal, bronchial); headache; confusion; irritability and anxiety; syncope; and respiratory distress leading to respiratory arrest.

Treatment

Anticholinesterase inhibitors (e.g., pyridostigmine bromide, neostigmine) may be used to inhibit breakdown of acetylcholine in the neuromuscular

synapse. Increased synaptic acetylcholine enhances the activation of postsynaptic receptors and improves skeletal muscle contraction force. Because myasthenia gravis is an autoimmune disorder, corticosteroids, IV immunoglobulin, plasmapheresis, and immunosuppressive agents may be used to regulate the immune system. In severe cases, respiratory muscle fatigue may necessitate mechanical ventilation. Thymectomy is often recommended when patients fail to respond well to medications, particularly in patients less than 45 years old. Research in biologic therapies targeting B cells and complement proteins are ongoing.

CHRONIC MUSCLE PAIN

Fibromyalgia Syndrome

Etiology and Pathogenesis

The cause of fibromyalgia syndrome (FMS) is unknown. No laboratory abnormalities have been found, muscle biopsy findings are nonspecific, and patients are usually normal on psychological testing. The condition is not an inflammatory process but rather a "pain syndrome," with recent studies suggesting that changes in the central nervous system may lead to amplification of pain fiber impulses, a theory called *central sensitization.* This generalized increase in pain sensitivity may involve both ascending and descending neural pathways and a variety of neurotransmitters and neuropeptides.

FMS is characterized by chronic pain in muscles and surrounding structures often of months' or years' duration. Additional symptoms include fatigue, sleep dysfunction, headache, numbness and tingling (i.e., paresthesia), joint pain, memory and concentration difficulties, irritable bowel syndrome, depression, edema of the hands, and sensitivity to cold. Patients either may have no other musculoskeletal disease or may have rheumatoid arthritis, osteoarthritis, Lyme disease, or sleep apnea. FMS is characterized by a strong female preponderance, with an estimated prevalence in the United States of more than 45% of the general population.

Clinical Manifestations

Patients with FMS complain of widespread musculoskeletal pain, stiffness, and fatigability. Joint pain and swelling may be perceived by the patient, but the swelling, if present, is usually soft tissue "puffiness" and not a true inflammatory process. Complaints of muscle pain and weakness are expressed without objective demonstration. In addition to pain and fatigue, sleep disturbances are a common complaint. The examination of a patient with FMS is characterized by an excessive number of reported symptoms with minimal objective findings other than diffuse muscular tenderness. The diagnosis is made on a clinical basis as there are no diagnostic laboratory or radiographic findings. Proposed criteria for the diagnosis of FMS established by the American College of Rheumatology in 2010 include widespread pain and muscle tenderness for at least 3 months in combination with other somatic complaints such as irritable bowel symptoms, fatigue, headaches, nausea, and others. These newer criteria are less restrictive than the original 1990 diagnostic criteria, which focused on specific muscle tender points. Although the specific tender points are the most common and are included in the criteria for fibromyalgia, in actuality, nearly any muscle in the human body could be tender to palpation.

Treatment

FMS is a chronic pain condition that is not life threatening, nor does it lead to destruction of musculoskeletal tissues. Because the cause of FMS is unknown, treatment focuses on maintaining functionality and reducing symptoms. Patient education is important and may be associated with improved outcomes and better prognosis. An exercise regimen is essential and should include regular stretching; improvement in physical conditioning via low-impact aerobic exercise (biking, swimming, walking); and measures of pacing, muscle protection, and relaxation. Because pain and fatigue may be aggravated by stress and other psychological factors, counseling may be helpful.

Blinded, randomized, placebo-controlled studies of amitriptyline, cyclobenzaprine, zolpidem, and alprazolam administered at bedtime have indicated that all are effective FMS therapy. Treatment begins at the lowest possible doses and increases as tolerated, with the goal being to improve quality of sleep without drug side effects, such as daytime somnolence or excessive dry mouth. Other medications under investigation in FMS are meant to lower pain sensitivity. The selective serotonin reuptake inhibitors that have been studied in FMS include fluoxetine, sertraline, and citalopram. Although their efficacy as monotherapy is modest at best, they may prove beneficial in combination with other agents. Most recently, pregabalin (an anticonvulsant agent) has been shown to be effective in reducing pain and is the first drug approved by the Food and Drug Administration for the treatment of FMS. Additional medications that may have a role include venlafaxine, duloxetine, milnacipran (serotonin-norepinephrine reuptake inhibitors), and tramadol (an opioid-like analgesic), among many other agents being investigated. Moclobemide and pirlindole are monoamine oxidase inhibitors that have also been studied.

KEY POINTS

- Inflammatory myopathies, including polymyositis and dermatomyositis, are associated with immune-mediated muscle inflammation and destruction.
- Muscular dystrophy comprises a group of genetic disorders characterized by degeneration of skeletal muscle.
- Duchenne muscular dystrophy is inherited as an X-linked disorder and affects only males.
- Becker muscular dystrophy is a milder form of inherited muscle degeneration, somewhat less common than the Duchenne type. The genetic mutation leads to production of a reduced amount of an abnormal dystrophin protein and a slower muscular degeneration. Calf hypertrophy is still prominent and often painful, with progressive loss of strength and ability to ambulate.
- Facioscapulohumeral muscular dystrophy is an autosomal-dominant disorder in which degenerating muscle fibers are replaced by connective tissue such that muscles may increase in bulk even though muscle strength is lost.
- Myotonic dystrophies (type 1 and type 2) are inherited disorders that lead to progressive muscle weakness and myotonia with delayed muscle relaxation after stimulation.
- Myasthenia gravis is an autoimmune disorder characterized by progressive weakness as the muscles are used. Antibodies against acetylcholine receptors in the motor end-plate interrupt neuromuscular transmission.
- FMS is a poorly characterized chronic disorder associated with generalized pain, stiffness, sleep dysfunction, and fatigability.

SUMMARY

A solid working knowledge of the anatomy, physiology, and biomechanics of movement is extremely important when dealing with any type of alteration in the musculoskeletal system. With a grasp of the mechanics involved in function, the clinician is able to approach each aberration with an awareness of the time requirements for healing, stress tolerances, and expected management outcomes.

The injuries and diseases discussed in this chapter are a small representation of the many dysfunctions that may afflict the musculoskeletal system. An ability to determine the specific type of tissue involved (contractile or inert) allows the clinician to be cognizant of activities that would aggravate trauma, types of injury that require supportive devices, and injuries that respond to medical intervention.

An awareness of the tissue response to healing enhances the clinician's evaluative skills and provides a signal regarding when intervention has achieved the expected results within an appropriate timeframe. It is the responsibility of the practitioner to become knowledgeable about the variety of dysfunctions that occur. This knowledge base must continue to expand as technological advancements provide increasingly complex levels of information and new diagnostic tools become available.

RESOURCES

Ahn JM, El-Khoung GY: Occult fractures of extremities. *Radiol Clin North Am* 43(3):561–579, 2007.

Bennett JE, Dolin R, Blaser MJ, editors: *Mandell, Douglas & Bennett's principles and practice of infectious disease*, ed 8, Philadelphia, 2015, Saunders Elsevier.

Clauw DJ: Fibromyalgia: a clinical review. *JAMA* 311(15):55–1547, 2014.

Cyriax J: *Textbook of orthopedic medicine: diagnosis of soft tissue lesions*, ed 8, London, 1982, Bailliere Tindall.

Firestein GS, Budd RC, Gabriel SE, et al: *Kelley and Firestein's textbook of rheumatology*, ed 10, Philadelphia, 2017, Saunders.

Frassica FJ, Sponseller PD, Wilckens JH, editors: *5-minute orthopaedic consult*, ed 2, Philadelphia, 2007, Lippincott Williams & Wilkins.

Gardner DG, Shoback D, editors: *Greenspan's basic and clinical endocrinology*, ed 9, New York, 2011, McGraw-Hill.

Goldman L, Schafer AI, editors: *Goldman's Cecil medicine*, ed 25, Philadelphia, 2016, Elsevier.

Hettinga DL: Inflammatory response of synovial joint structures. In Gould J, editor: *Orthopedic and sports physical therapy*, ed 2, St Louis, 1990, Mosby, p 100.

Hochberg MC, Silman AJ, Smolen JS, et al: *Rheumatology*, ed 6, Philadelphia, 2015, Elsevier.

Klippel JH, Stone JH, Crofford LJ, White PH, editors: *Primer on the rheumatic diseases*, ed 13, New York, 2008, Springer.

Mencio GA, Swiontkowski MF, editors: *Green's Skeletal trauma in children*, ed 5, Philadelphia, 2015, Elsevier.

Miller MD, Hart JA, MacKnight JM: *Essential orthopedics*, Philadelphia, 2010, Elsevier.

Siegel R, Ward E, Brawley O, Jemal A: Cancer statistics, 2011: the impact of eliminating socioeconomic and racial disparities on premature cancer deaths. *CA Cancer J Clin* 61(4):212, 2011.

Silver DS: Calcium and vitamin D controversies. *Rheum Dis Clin North Am* 37(3):351–363, 2011.

Skinner HB, McMahon PJ, editors: *Current diagnosis and treatment in orthopedics*, ed 5, New York, 2014, McGraw-Hill.

Staud R: Abnormal pain modulation in patients with spatially distributed chronic pain: fibromyalgia. *Rheum Dis Clin North Am* 35(2):263–274, 2009.

Zhang J, Saag KG, Curtis JR: Long-term safety concerns of antiresorptive therapy. *Rheum Dis Clin North Am* 37(3):387–400, 2011.

52

Alterations in Musculoskeletal Function: Rheumatic Disorders

Carol L. Danning

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- How are osteoarthritis and rheumatoid arthritis differentiated on the basis of cause, clinical findings, and treatment?
- What are the similarities and differences among rheumatoid arthritis, systemic lupus erythematosus, and scleroderma?
- What are the infective organisms associated with joint inflammation and Lyme disease?
- What is the pathogenesis of gouty arthritis?
- How do the three subtypes of juvenile rheumatoid arthritis differ?

CHAPTER OUTLINE

Local Disorders of Joint Function, 1042
Osteoarthritis, 1042
Etiology and Pathogenesis, 1043
Clinical Manifestations, 1043
Treatment, 1043
Infectious Arthritis, 1045
Etiology and Pathogenesis, 1045
Clinical Manifestations, 1045
Treatment, 1045
Joint Prosthesis Infection, 1045
Lyme Disease, 1045
Clinical Manifestations and Treatment, 1046
Systemic Disorders of Joint Function, 1046
Immune-Mediated Disorders, 1046
Rheumatoid Arthritis, 1046
Systemic Lupus Erythematosus, 1050
Scleroderma, 1050
Ankylosing Spondylitis, 1051
Postinfectious Systemic Disorders, 1052
Reactive Arthritis (Reiter Syndrome), 1052
Acute Rheumatic Fever, 1052
Joint Dysfunction Secondary to Other Diseases, 1053
Psoriatic Arthritis, 1053
Etiology and Pathogenesis, 1053
Clinical Manifestations, 1053
Treatment, 1053
Enteropathic Arthritis, 1053
Clinical Manifestations, 1053
Treatment, 1053
Neuropathic Osteoarthropathy, 1054
Clinical Manifestations and Treatment, 1054
Hemophilic Arthropathy, 1054
Clinical Manifestations, 1054
Gout, 1054
Clinical Manifestations, 1054
Asymptomatic Hyperuricemia, 1054
Acute Gouty Arthritis, 1054
Intercritical Gout, 1055
Chronic Tophaceous Gout, 1055
Treatment, 1055
Adult-Onset Still Disease, 1055
Clinical Manifestations and Treatment, 1055
Pediatric Joint Disorders, 1056
Nonarticular Rheumatism, 1056
Hypermobility of Joints, 1056
Juvenile Idiopathic Arthritis, 1056
Clinical Manifestations, 1056
Treatment, 1056

Arthritis is the most common disabling musculoskeletal condition in the United States. The National Arthritis Foundation estimates that approximately 46 million people have arthritis, with the numbers increasing yearly as the population ages. More than 150 defined rheumatologic diseases have been identified. This chapter discusses the more common rheumatologic diseases.

LOCAL DISORDERS OF JOINT FUNCTION

Osteoarthritis

Osteoarthritis (OA; degenerative joint disease) is the most common arthritis worldwide. It is a progressive, noninflammatory disease of diarthrodial joints, especially those that are weight bearing. It is

characterized by a progressive loss of articular cartilage and by formation of thick subchondral bone and new bone at the joint margins. OA becomes more prevalent with increasing age. Individuals older than 70 years have the highest incidence. In postmenopausal women, the knees and hands are most frequently affected by OA. It is difficult to estimate the exact prevalence of OA because of difficulties associated with diagnosis, lack of longitudinal data, and problems in defining disease onset.

Etiology and Pathogenesis

The etiologic progression of OA varies widely. The age of onset can be variable as well as the rate of progression to more advanced stages of cartilage loss or osteophyte formation. Development of OA may be related to factors that increase the likelihood of abnormal "wear and tear" on joints such as obesity, joint trauma, and congenital disorders (e.g., hip dysplasia, joint laxity, leg length discrepancy). Other predisposing conditions include lifestyle factors and occupation (stress to joints), genetic predisposition, and hormonal status (postmenopausal).

Biomechanical, biochemical, inflammatory, and immunologic factors may all be involved in the development of OA (Fig. 52.1). An initial injury causes release of proteolytic and collagenolytic enzymes from chondrocytes. A breakdown of the matrix of proteoglycan and collagen occurs. The decreased hydration of cartilage that occurs with aging can increase the likelihood of wear and damage. Collagen fatigue and microfracture occur with the stress of weight bearing. The ability of the structure to absorb shock is decreased as a result of subcortical bone and cartilage microfractures. Breakdown of joint integrity overloads the capacity for repair, with resultant degenerative changes.

The structural deterioration of the cartilage involves fissuring, pitting, and erosion. The erosion can become so extensive that the articular surface denudes the full thickness of the cartilage. Osteophyte spur formation, subchondral bone sclerosis, and cyst formation are also examples of structural changes present in OA. Cartilage fragments may break off into joints and form "loose bodies" (Fig. 52.2). Joint effusions are common in advanced cases. Synovium becomes inflamed and secretes an increased amount of synovial fluid, which causes the joint to distend.

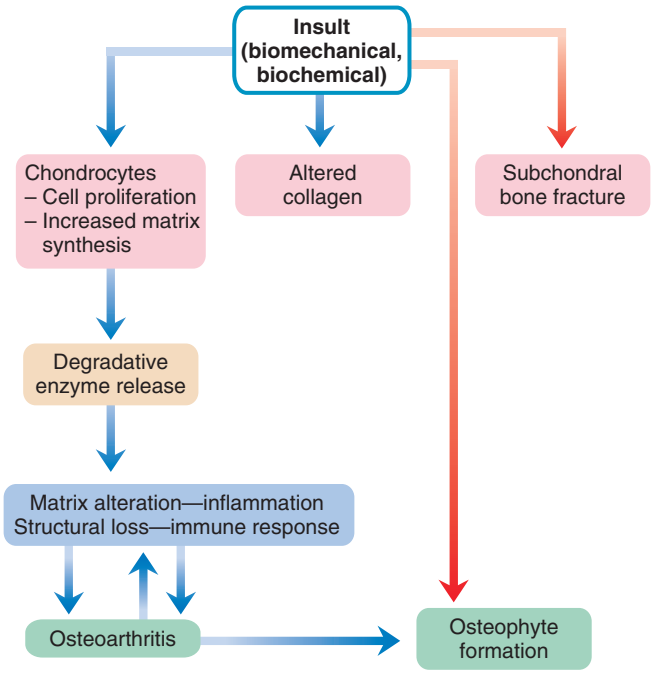


FIG 52.1 Pathogenesis of osteoarthritis.

Clinical Manifestations

Bony enlargement of joints, crepitus with movement, morning stiffness lasting less than 30 minutes (that improves with joint mobility), and pain with function are typical clinical manifestations of OA. These signs and symptoms are usually local. Although any joint may be affected, weight-bearing joints such as the hips and knees, cervical and lumbosacral joints, and interphalangeal joints are most frequently involved. Degenerative arthritis or OA may occur in an isolated joint, or multiple joints can be involved, especially in the hand. Mechanical dysfunction, anatomic anomalies, or trauma may cause breakdown of the joint surface. It is imperative to establish a differential diagnosis and to eliminate a systemic or medical problem. Although OA is localized, rheumatoid arthritis (RA) is a systemic autoimmune disorder that causes a highly inflammatory, symmetric, peripheral arthritis.

Radiologic abnormalities are normally consistent with clinical symptoms. Classic findings include bony proliferation at the joint margins (i.e., osteophytes or bone spurs), asymmetric narrowing of the joint space, and sclerosis of the subchondral bone. Later, malalignment of the joints and cyst formation in subchondral bone can also be seen. When significant synovial fluid accumulates in the joint, it is usually translucent, noninflammatory fluid containing less than 2000 white blood cells per cubic millimeter.

The most common deformity of the hands occurs in the distal interphalangeal (DIP) joints (Fig. 52.3). Enlargement is caused by bone spurs *(Heberden nodes)* that form on the dorsolateral and medial aspects of the joint. Similar enlargements in the proximal interphalangeal (PIP) joints are called *Bouchard nodes.* The knees and hips are also common locations for OA. Local pain over joint margins, tenderness, crepitus, and muscle atrophy are common findings. Loss of cartilage in medial or lateral compartments of the knee may lead to such structural changes as genu valgus or varus (Fig. 52.4).

Pain is relieved by rest during initial stages, although stiffness can be a complaint after prolonged sitting. Because cartilage does not contain nociceptors (pain receptors), pain originates from intraarticular and periarticular structures. Although an acute inflammatory response is often the result of a specific traumatic incident and may cause synovitis in the joint capsule, acute inflammation is not as commonly associated with OA. As breakdown in structure progresses, even light activity elicits discomfort, and pain at night is common.

Treatment

Initial treatment is designed to decrease stress on the joint and protect it from additional trauma. Nonpharmacologic interventions include weight reduction (for weight-bearing joints) and exercise or physical therapy to improve range of motion as well as muscle strength and conditioning. Orthotics can provide symptomatic relief through joint stabilization and support of deformities. These may include shoe insoles, heel lifts, patellar taping, and joint braces or splints. Assistive devices, such as a cane or walker, afford mechanical relief of weight-bearing stress as well as gait stabilization.

With pharmacologic intervention, acetaminophen is often the initial analgesic agent recommended for management of mild OA symptoms. Nonsteroidal antiinflammatory drug (NSAID) therapy decreases swelling and pain. Most NSAIDs are nonselective, but antiinflammatory agents that target the cyclooxygenase-2 (COX-2) enzyme have clinical benefit equal to traditional NSAIDs and may have potentially fewer gastrointestinal side effects. The only COX-2 inhibitor presently available in the United States is celecoxib. All NSAIDs (selective or nonselective) have the potential for renal toxicity and possibly even cardiovascular side effects. Topical NSAID preparations are being studied with variable efficacy but likely improved safety. A topical preparation of diclofenac is currently available in the United States. Clinical research on the efficacy

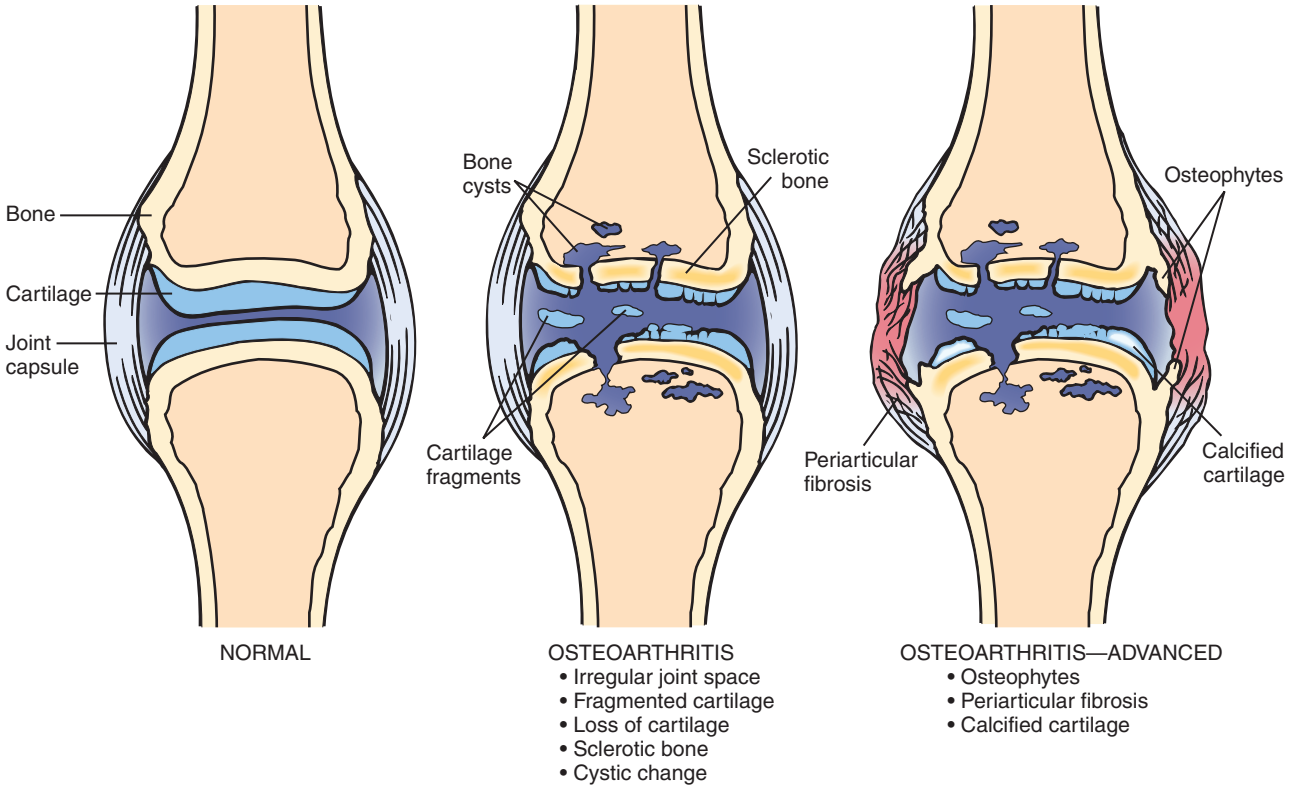


FIG 52.2 Schematic presentation of the pathologic changes in osteoarthritis. Fragmentation and loss of cartilage denude the subchondral bone, which undergoes sclerosis and cystic change. Osteophytes form on the lateral side and protrude into the adjacent soft tissues, causing irritation, inflammation, and fibrosis. (From Damjanov I: *Pathology for the health professions,* ed 4, Philadelphia, 2012, Saunders.)

FIG 52.3 Comparison of Heberden nodes with Bouchard nodes (seen in patients with osteoarthritis).

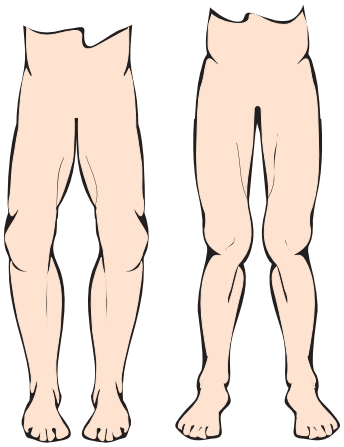


FIG 52.4 Genu varus and genu valgus.

of nutraceutical agents for OA, especially glucosamine and chondroitin sulfate, has yielded very mixed results. Use of these agents may, in some individuals, reduce joint pain or stiffness, but effects on joint tissue, such as disease modification or slowing of cartilage loss, have been unsubstantiated.

Viscosupplementation, the intraarticular injection of hyaluronan or its derivatives, may increase joint lubrication, reduce inflammation, and alleviate pain. The actual mechanism of action is not known. These agents are currently only available for use in knee OA. Intraarticular corticosteroid injections can provide temporary pain relief, but too-frequent usage (i.e., more than three or four injections per year in the same joint) may have eventual deleterious effects.

Surgical intervention may be necessary if the joint surface loses enough integrity to prevent joint function. OA is the most common reason for total hip and total knee replacement.

Infectious Arthritis

Infectious or septic arthritis may be defined as an invasion of the synovial membrane by bacteria or another pathogen, leading to a closed-space infection. A reported annual incidence of infection in a joint is 2 to 10 cases per 100,000, with much higher rates noted in patients with comorbid diseases such as RA, diabetes, OA, chronic kidney disease, and others. Bacterial joint infection is the diagnosis found in 8% to 27% of patients presenting with one or more acutely painful joints. The pathogen can invade the joint space via a hematogenous route, by extension of an adjacent infection, or from direct inoculation after trauma or an invasive procedure. Infection causes both synovium and cartilage to deteriorate. Joints with underlying disease or inflammation are more susceptible to infection because of increased vascularity and defective barrier effects of the synovial tissue.

Etiology and Pathogenesis

The basic cause of bone and cartilage destruction is the interaction of antigenic bacterial cell wall components, the toxic effects of bacteria, the destruction caused by the purulent inflammatory exudate, and the local immune-mediated synovial or cartilage response. If the bacterial infection is not managed, cartilage can be destroyed, and this can lead to eventual joint destruction.

In most adults and neonates, *Staphylococcus aureus* is the most common causative organism followed by *Streptococcus pyogenes* and *Streptococcus pneumoniae.* Also in neonates, gram-negative bacilli can be cultured, including *Kingella kingae,* which is an oral flora. Gram-negative bacilli can infect joints of the elderly, the immunocompromised, or intravenous (IV) drug users (especially with *Pseudomonas aeruginosa*). In young children, *Haemophilius influenzae* type B (HIB) was a frequent cause of infection, but this agent is now rare since the development of the HIB vaccine. *Neisseria gonorrheae* is a causative organism in some adults younger than 30 years of age.

Clinical Manifestations

The patient with septic arthritis presents with joint pain, fever, chills, and leukocytosis. Fever may range from mild to high fever with shaking chills. A warm, red, and very swollen joint with limited range of motion attributable to pain is symptomatic of any type of infectious arthritis. Usually only a single joint is involved, but polyarticular joint infections can occur in 10% to 20% of cases, most often in debilitated or immunosuppressed individuals. Synovial fluid analysis reveals a very high white cell count with a predominance of neutrophils (often >50,000 cells/mm^3), and diagnosis is established by recovery of bacteria from synovial fluid. Blood cultures are also used to provide a medical diagnosis.

A form of chronic septic arthritis can be seen when the infecting organism is a fungal or mycobacterial agent. These cases typically follow a slower, more indolent course with mild symptoms. Immunocompromised individuals may have higher risk of this.

Treatment

Treatment of a septic joint should include appropriate antibiotic therapy (often initially IV and then oral) with the average duration required being 4 to 6 weeks. This therapy is most effective when the bacteria can be isolated and identified from the synovial fluid and sensitivities can be used to determine the most effective antibiotic to be used. In addition, the infected joint usually requires repetitive drainage to facilitate bacterial clearance, decrease pain, and prevent loss of function. This can be accomplished by repeated joint aspiration, arthroscopy with tidal lavage, or open surgical drainage.

Joint Prosthesis Infection

Prosthetic joint infections are uncommon, occurring in 1% to 3% of joint replacement surgeries in the United States. When this infection occurs within a few weeks to months of the surgery, it is likely to be related to contamination at the time of surgery. Infections occurring more than 24 months postoperatively are related to hematogenous spread. Any bacteria can lead to infection in a prosthetic joint via the hematogenous route. *S. aureus* is still common, but *Staphylococcus epidermidis* can also result in prosthetic joint infection and is rarely seen in a native septic joint. Generally, a severe prosthetic joint infection requires removal of the prosthesis, replacing it with an antibiotic saturated spacer, followed by a rigorous course of IV antibiotic therapy, often for 6 weeks or longer. The prosthesis is replaced when cultures from the wound show no growth. For prosthetic infections that are diagnosed early, have a stable prosthesis, and no soft tissue damage, a process of debridement, antibiotics, and implant retention might be possible.

Lyme Disease

Lyme disease is a complex illness caused by the *Borrelia burgdorferi* tickborne spirochete and is prevalent in North America as well as parts of Europe and Asia. In the United States, Lyme disease is found most commonly in three regions: the Northeast (North Carolina to Maine), the Midwest (Wisconsin, Minnesota, and Michigan), and West Coast (mainly Northern California). It is commonly carried by the deer tick, although other species serve as vectors as well. The process by which the presence of the spirochete leads to the later chronic clinical symptoms of Lyme disease is unclear. Theories suggest that persistent antigenic fragments of the dead or inactivated organism trigger the inflammatory responses, or the initial infection stimulates an autoimmune inflammatory

response. Diagnosis is usually made by typical symptoms and positive serologic testing for immunoglobulin responses to the spirochete.

Clinical Manifestations and Treatment

Early disease symptoms occur most often in the summer months. The tick bite produces a red macule or papule that may expand to form an annular lesion within days to weeks of the exposure. The lesion may expand and become quite red. The lesion is warm to the touch but not painful and is often accompanied by a flulike illness, headache, neck stiffness, fever, chills, myalgia, arthralgia, malaise, and fatigue. In severe cases, early systemic involvement may consist of lymphadenopathy, splenomegaly, hepatitis, nonproductive cough, testicular swelling, and conjunctivitis.

A later phase of Lyme disease occurs a few weeks after the tick bite. Musculoskeletal symptoms follow a pattern of migratory pain in joints, tendons, bursae, muscles, or bones. More than half of patients develop frank arthritis with involvement of the large joints. The knee is a particularly common site of Lyme arthritis and can be associated with large effusions. A persistent arthritis can develop late in the disease course in as many as 60% of patients in the United States.

In about 15% of untreated patients, neurologic symptoms can occur weeks to months after the initial infection. This may include meningitis, cranial neuritis, motor and sensory radiculoneuritis, and chorea. Also late in the clinical course, cardiac involvement may be noted (about 5% of untreated patients), including atrioventricular blocks, left ventricular dysfunction, or cardiomegaly. Although cardiac involvement lasts only a few weeks, it can be fatal.Treatment is with oral or parenteral antibiotics.

KEY POINTS

- OA is a local degenerative joint disorder associated with aging and wear and tear from repetitive stress.
- OA is characterized by loss of articular cartilage, deterioration of underlying bone, and the formation of bone spurs. The process is noninflammatory. Weight-bearing joints are often affected.
- Signs and symptoms of OA are localized (not systemic) and include joint pain and crepitus with movement.
- Joint infection may be due to a variety of infectious agents, but bacteria are the most problematic. The route of infection is usually by way of the bloodstream. Signs and symptoms are due to localized infection and the systemic manifestations of inflammation.

SYSTEMIC DISORDERS OF JOINT FUNCTION

Immune-Mediated Disorders

Rheumatoid Arthritis

Rheumatoid arthritis is a systemic autoimmune inflammatory disease. In the United States approximately 0.5% to 1.1% of the population is affected, with a lower prevalence being reported in Asian countries. Women are two to three times more likely to develop RA than men, with a peak incidence in the fourth and fifth decades. Gender difference disappears in older age, such as after menopause. RA affects all races, and its prevalence is not affected by climate. RA occurs two to three times more often in women with a familial history of the condition.

Etiology and pathogenesis. The specific cause of RA is unknown. An infectious agent has long been sought as a cause, but reproducible evidence of a particular bacterium or virus has not been found. A more likely theory is that RA is caused by an abnormal autoimmune response (possibly triggered by a bacterial or viral antigen) occurring in individuals who have a genetic predisposition to the disease. This is indicated based on twin studies, which show a monozygotic twin of a patient with RA has a risk of developing RA at approximately 15% (up from the general population risk of 1%).

Current research suggests that susceptibility to RA is influenced by the structure of the major histocompatibility complex molecules of antigen-presenting cells, particularly genes that code for class II human leukocyte antigens (HLA), although it is likely that several genes are involved. Past studies have defined genes for certain HLA-DR4 alleles to be associated with higher rates of RA in Caucasian populations. A particular subgroup of alleles in the HLA-DRB1 gene was found to have a key sequence of five amino acids in common, called the *shared epitope.* These genes are suspected to be associated with control of humoral and cell-mediated immune responses, such as the development of autoreactive T cells. Further studies have demonstrated that different HLA-DR4 alleles may be linked to higher risk of RA development in different ethnicities.

In patients at risk of getting RA, the cause and type of stimulus (or trigger) of immunologic abnormalities are not known, but they might be due to an infectious agent, environmental influences, or other lifestyle factors (such as tobacco use). The higher incidence of onset in premenopausal females or after a pregnancy also suggests an influence of hormonal factors. An increase in physical and/or psychological stress has also been associated with precipitating acute exacerbation of the disease.

Initially, pathologic changes in RA occur when the immune response localizes in synovial tissue. Here lymphocytes (T and B cells) and macrophages are activated by an unknown antigen trigger. Activated B cells help perpetuate the escalating inflammatory response by stimulating more lymphocytes and other immune cells. B cells also produce rheumatoid factor (RF) antibodies against immunoglobulin G (IgG) as well as anti–cyclic citrullinated peptide (anti-CCP) antibodies. Although immunoglobulins are natural human antibodies, the body produces an abnormal antibody (RF) against its own antibodies (IgG). Anti-CCP antibodies target peptides modified by converting the amino acid arginine to citrulline. Activated lymphocytes, macrophages, and antigen–antibody complexes activate the complement system, stimulate recruitment of other immune cells into the synovium, and produce an extensive array of inflammatory cytokines, metalloproteinases, and other mediators. These products of macrophages and lymphocytes are believed to be critical in RA pathogenesis because they stimulate and perpetuate the inflammation in the joint. Key proinflammatory cytokines demonstrated in the synovium include tumor necrosis factor-α (TNF-α); interleukin-1β (IL-1β); and interleukins 6, 8, 15, 17, 18, and 23, although many more are thought to be involved. Newer biological therapies are designed to target these cytokines as well as activated inflammatory cells and cell costimulatory markers.

The escalating inflammatory response in the rheumatoid joint leads to accumulation of dense aggregates of immune cells and infiltration of the synovial membrane that lines the joint cavity. The cells produce more cytokines and growth factors, which also stimulate edema, neovascularization, and proliferation of the synovium (which expands in a tumorlike manner). This hypertrophied synovium invades such surrounding tissue as cartilage, ligaments, joint capsules, and tendons. Granulation tissue forms, covering articular cartilage and leading to pannus formation. Pannus is vascularized tissue composed of lymphocytes, macrophages, histiocytes, fibroblasts, and mast cells. Pannus can erode and destroy articular cartilage, resulting in bone erosion, bone cysts, and fissures (Fig. 52.5). The expansion and destruction of joint structures can lead to inflammation, shortening, and even rupture of tendons as well as ligament laxity, joint subluxations, contractures, and deformities.

Clinical manifestations. RA has a wide range of clinical features, but the classic presentation is that of a bilateral symmetric polyarthritis

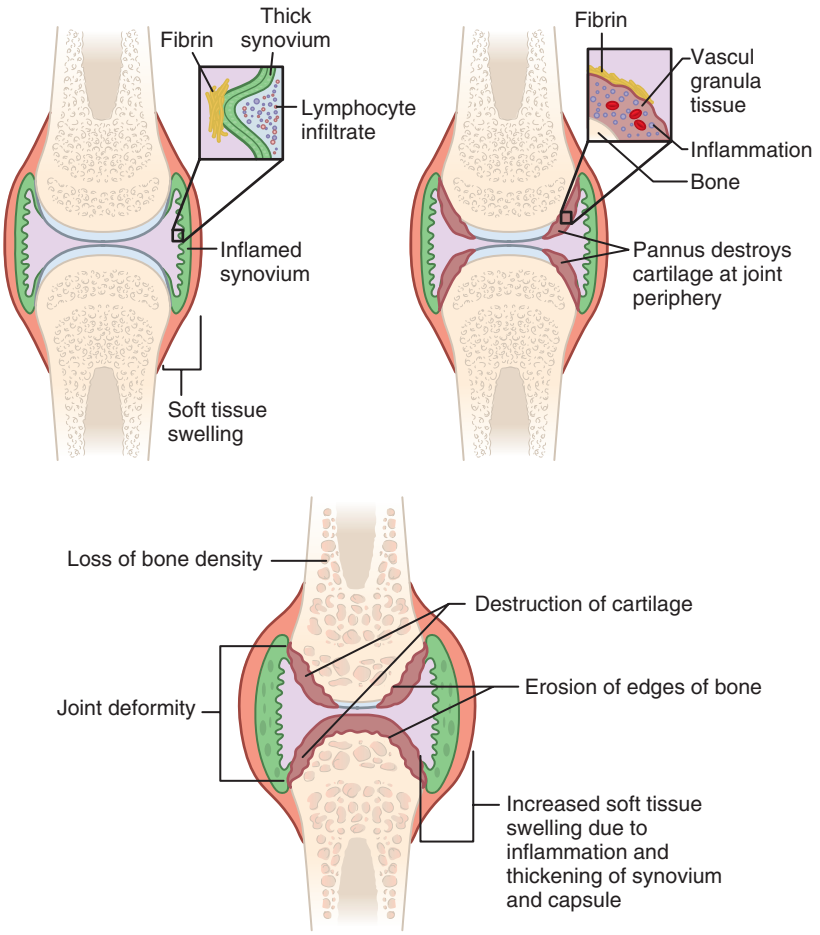


FIG 52.5 Schematic presentation of the pathologic changes in rheumatoid arthritis. The inflammation (synovitis) leads to pannus formation, obliteration of the articular space, and, finally, ankylosis. The periarticular bone shows disuse atrophy in the form of osteoporosis.

especially involving smaller joints. Malaise, fatigue, and diffuse musculoskeletal pain are common manifestations during acute exacerbations of the disease. It is interesting to note that although symmetric patterns involving the joints of the hands, wrists, elbows, and shoulders are evident, DIP joints are usually spared. This symmetry and the noninvolvement of the DIP joint assist in making the diagnosis of RA.

The hands, wrists, knees, and feet are most commonly involved. In the spine, the upper cervical area is most often affected. However, any diarthrodial joint is potentially at risk. The development of pannus followed by inflammatory destruction of the soft tissue leads to laxity of the ligaments and tendons and results in biomechanical dysfunction. This mechanical stress causes the typical deformities of RA.

Swelling in the hands is a typical sign of metacarpophalangeal (MCP) and proximal interphalangeal (PIP) joint involvement. Pain is elicited on palpation of the joints. Gradually, progressive synovial damage leads to characteristic ulnar deviation in the MCP joint (Fig. 52.6). In advanced situations, a swan-neck deformity develops in the fingers. Swan-neck deformity is a hyperextension of the PIP joint with flexion of the MCP and DIP joints attributable to contractures of intrinsic muscles and tendons. A boutonnière deformity, consisting of flexion of the PIP joints and hyperextension of the DIP joints, is also a common dysfunctional position caused by rupture or excess laxity of extensor tendons over the fingers (Fig. 52.7). Loss of strength and the ability to achieve a strong pinch is frequently noted in the hand affected by RA. A rupture of tendons and loss of the ability to extend the fingers are common findings in later stages of the disease.

The wrist is commonly involved. The synovium around the wrist becomes boggy and affects the tendon sheaths. Limitation of movement, especially dorsiflexion of the wrist, is often noted. Proliferation of the synovium on the volar or palmar aspect of the wrist may cause compression of the median nerve and development of carpal tunnel syndrome. Flexion contractures and swelling of the elbow are other common manifestations; in later stages of the disease, shoulder involvement may occur. Typical signs of shoulder involvement are limitations of movement and pain on palpation in the area of the coracoid process. Dislocation, subluxation, or rupture of the joint capsule may occur as the disease progresses.

Involvement of the upper cervical vertebrae is another common finding. Destruction of structures of the atlantoaxial vertebrae (such as the transverse ligament) creates the potential for subluxation of this joint and endangerment of spinal cord compression. Laxity in the cervical region may also allow compression of the vertebral artery, leading to vertebrobasilar insufficiency. Limitation of motion (especially rotation), pain on palpation, and headache in the occipital region are common.

Abnormalities in gait and limitations of movement are signs noted when RA affects the hip. Groin pain attributable to capsular involvement may be present. If synovitis of the hip becomes extensive, severe pain may be noted on evaluation. RA involvement in the knee is often extensive. Effusion, quadriceps atrophy, contractures, and synovitis of the semimembranous bursa (popliteal cyst) may be observed. Destruction of the articular surface, bone, and soft tissue may result from joint instability. One common clinical sign of involvement of the foot is

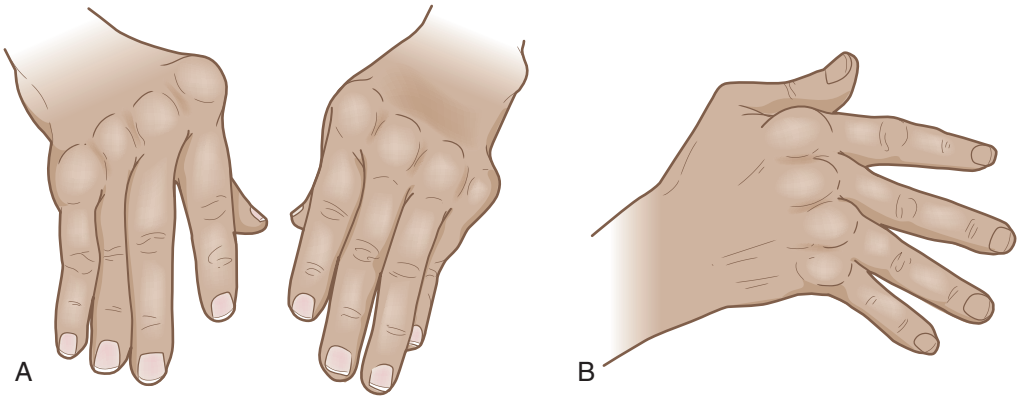


FIG 52.6 **A,** Typical deformity of the hand seen in patients with rheumatoid arthritis. Note ulnar deviation involving the right hand. **B,** Schematic of ulnar deviation.

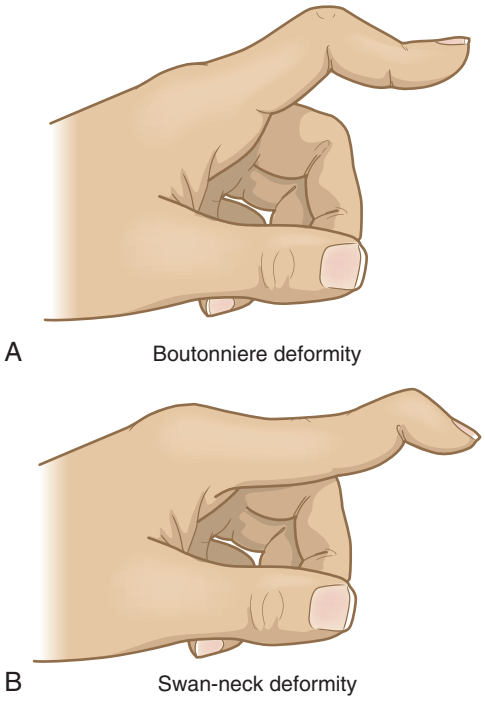


FIG 52.7 **A,** Palmer view of hand shows ulnar drift of the index through small fingers and boutonniere deformity of the thumb. **B,** Side view of hand shows the swan neck deformities of the fingers due in part to palmar subluxation of the digits at the MCP joints.

retrocalcaneal bursitis. Other features of foot involvement include swelling of joints, a cocking-up of the toes attributable to subluxation of the metatarsal heads (claw toes), and lateral deviation of the toes at the metatarsophalangeal (MTP) joints.

These clinical manifestations may develop rapidly or progress over many years. Usually symptoms develop over weeks and months. Initially, the patient may feel fatigued or chronically tired and may complain of systemic aching in the musculoskeletal system. Specific joint pain, tenderness, swelling, redness, and nodules are quite common.

Prolonged inactivity, such as sitting, initiates complaints of stiffness and swelling. As the disease progresses, walking, climbing stairs, opening jars or doors, and precise movement of the digits become quite difficult. Weight loss, depression, and a low-grade fever often are noted in these patients.

Unlike OA, RA is a systemic autoimmune disease. It is imperative that the diagnosis be made as early as possible so that intervention can be initiated quickly. RA may be confused with a number of disease entities such as Lyme disease, systemic lupus erythematosus (SLE), gout, and other inflammatory conditions. RA may also cause subcutaneous nodules in 15% to 20% of patients. These are made of dense perivascular collections of inflammatory cells surrounded by palisading fibroblasts. RA can also be associated with cardiac, pulmonary, skin, and ophthalmologic manifestations.

Cardiac manifestations may include pericarditis (often asymptomatic), myocarditis (rare), mitral valve disease, and conduction system disease or complete heart block. Patients with long-standing RA also have a higher risk of atherosclerotic coronary artery disease than age-matched controls. Pulmonary manifestations may occur as pleural effusions (about 20% patients), pulmonary fibrosis, or pulmonary nodules. Apart from the subcutaneous nodules, other skin manifestations can sometimes be seen, including palpable purpura or nail fold infarcts. The purpura is caused by inflammation of small arterioles in the skin (leucocytoclastic vasculitis). Ophthalmic manifestations might include episcleritis, scleritis, or secondary Sjögren syndrome (dry eyes and mouth).

A positive RF is found in the sera of approximately 60% to 80% of patients with RA. The titer of the RF does not fluctuate with disease activity and is not essential for the diagnosis of RA. Anti-CCP antibody can be found with a sensitivity of 60% to 70% and a high specificity of approximately 90% for RA. The presence of both RF and anti-CCP antibodies in a patient may signify a risk of more aggressive disease, although RA in the absence of these two markers can occur. Inflammatory markers (sedimentation rate, C-reactive protein) are often elevated. Other laboratory features may include hypergammaglobulinemia, thrombocytosis, and hypochromic microcytic anemia.

Radiography may demonstrate structural damage caused by RA. Typical findings include erosions on the bony margins of joints, joint space narrowing, periarticular osteopenia, and eventual malalignment and subluxation of the bones (Fig. 52.8).

Diagnosis. The American College of Rheumatology (ACR) in collaboration with the European League Against Rheumatism (EULAR) published guidelines for the classification and diagnosis of RA (Table 52.1). The criteria require the patient have at least one documented swollen joint, the absence of an alternative diagnosis that better explains the joint swelling, and a total score on the criteria scale of 6 or more (out of 10). The diagnosis takes into account the number and size of the swollen joints, the presence or absence of the RF antibody or CCP antibody, and the level of the patient's markers of inflammation

FIG 52.8 Rheumatoid arthritis. Radiographs of deformities of the feet, including bone erosions, osteopenia, lateral deviation, and subluxations, especially at the metatarsophalangeal joints. (Courtesy Dr. Douglas White.)

TABLE 52.1 2010 ACR/EULAR Classification Criteria for Rheumatoid Arthritis

Criterion	Points*
Joint Distribution (0–5 Points)	
1 large joint	0
2–10 large joints	1
1–3 small joints	2
4–10 small joints	3
>10 joints (at least 1 small joint)	5
Serology (0–3 Points)	
Negative RF *and* negative ACPA	0
Low-positive RF *or* low-positive ACPA	2
High-positive RF *or* high-positive ACPA	3
Symptom Duration (0–1 Point)	
<6 weeks	0
≥6 weeks	1
Acute-Phase Reactants (0–1 Point)	
Normal CRP *and* normal ESR	0
Elevated CRP *or* elevated ESR	1

Data from Aletaha D, Neogi T, Silman AJ, et al: 2010 Rheumatoid arthritis classification criteria, *Arthritis Rheum* 62(9):2574, 2010.
*A score ≥6 points indicates definite rheumatoid arthritis.
ACPA, Anti–citrullinated protein/peptide antibody; *ACR,* American College of Rheumatology; *CRP,* C-reactive protein; *ESR,* erythrocyte sedimentation rate; *EULAR,* European League Against Rheumatism; *RF,* rheumatoid factor.

(sedimentation rate and C-reactive protein). The goal of these criteria is to aid in the diagnosis of RA at an earlier stage of disease.

Treatment. Goals of therapy should include patient education, alleviation of pain and swelling, prevention of structural damage, and preservation of function. Initial therapy with antiinflammatory medications may include NSAIDs, COX-2 inhibitors, or corticosteroids (oral, parenteral, or intraarticular injections). Corticosteroids are potent antiinflammatory agents, effective at quickly controlling the pain, stiffness, and swelling of RA activity; however, every effort is made to avoid long-term steroid use because of adverse consequences such as steroid-induced osteoporosis, diabetes mellitus, cataracts, and more.

Disease-modifying antirheumatic drugs are used to achieve long-term control of RA activity and are recommended very early in the course of the disease to minimize later damage. The most common first-line agent used is methotrexate (oral or subcutaneously once weekly), which is a folate analog that blocks folate activity. Within cells, methotrexate may exhibit its antiinflammatory and immunoregulatory effects by inhibiting several different enzymatic pathways, leading to inhibition of adenosine monophosphate (AMP) deaminase and accumulation of intracellular and extracellular AMP and adenosine (which is a potent antiinflammatory agent). Leflunomide is another immune modulator that may work via decreased pyrimidine synthesis (and thereby may reduce the proliferation of lymphocytes), although the mechanism of action of this agent is not well understood.

Sulfasalazine and antimalarial drugs (especially hydroxychloroquine) have inhibitory effects on inflammatory cell function and may be useful in milder cases of RA or in combinations with methotrexate. Older immunosuppressive agents, such as azathioprine, gold (oral or intramuscular), or cyclosporine, are still available for management of RA but are much less commonly used.

Research into the use of biological agents in rheumatic disease, especially RA, has yielded a number of new and effective agents for controlling RA inflammation as well as preventing erosive and destructive mechanisms. TNF-α is a central and potent stimulator of inflammation in the rheumatoid joint, and there are currently five TNF-α inhibitors available in the United States. Etanercept is a soluble TNF-α receptor fusion protein, whereas adalimumab, golimumab, and certolizumab are monoclonal antibodies or modified fragments of antibodies that can bind and inhibit TNF-α. These agents are all self-administered injections given once weekly to once monthly depending on the agent. Infliximab is a chimeric monoclonal antibody against human TNF-α and is given as an IV infusion every 6 to 8 weeks.

Another biological agent available for treatment of aggressive RA is abatacept, which is a fusion protein that binds to CD80/CD86 on the surface of antigen-presenting cells. This binding prevents CD80/86 from binding with CD28 on T cells, a process that is necessary for T-cell activation and increased activity in the inflammatory response. Abatacept is administered once every 4 weeks via IV infusion or by subcutaneous injection every week. Rituximab, a monoclonal antibody against the B-cell marker CD20, has been approved since 1997 for the treatment of non-Hodgkin lymphoma. In RA, B cells are known to be active in several facets of the chronic inflammation, and therefore the B-cell depletion caused by rituximab therapy can have positive and sustained immunosuppressive effects. This agent is given as a series of two IV infusions administered 14 days apart every 6 months. Another agent approved for RA is tocilizumab, which is a humanized monoclonal antibody against the receptor of the proinflammatory cytokine IL-6. Blocking IL-6 can decrease effective activity of a variety of immune cells involved in the pathogenesis of RA inflammation. It is given as a monthly IV infusion or subcutaneous injection every 1 or 2 weeks.

In 2012 the first oral tyrosine kinase inhibitor, tofacitinib, was approved by the U.S. Food and Drug Administration (FDA) for treatment

of RA. Tyrosine kinases are enzymes on cell surfaces or within cells that transmit signals from the cell surface to the cell nucleus by a cascade of enzymatic reactions. By blocking this signaling, cell activation and protein transcription can be blocked. Tofacitinib inhibits primarily the kinase called *Janus kinase 3* (JAK3), and in doing so, blocks the action of certain proinflammatory cytokines on immune cells. Research on several other kinase inhibitors is ongoing.

Systemic Lupus Erythematosus

Systemic lupus erythematosus (SLE) is a chronic, multisystem, inflammatory, autoimmune disease. It is characterized by periods of exacerbations and remission, with multiple organ systems being affected at different times.

Etiology and pathogenesis. Genetic involvement has been demonstrated in familial occurrences of SLE. Although SLE occurs in all races, it occurs more often in the United States among African Americans than Caucasians and yet is uncommon in Africa. Environmental factors such as sunlight, thermal burns, and other types of physical stress may initiate the development of SLE. SLE is more common in women, with peak incidence between 15 and 40 years of age, suggesting that hormonal factors may influence onset. Other factors that can influence SLE development and disease activity include tobacco use and environmental exposures.

SLE is the result of an abnormal immune reaction of the body against its own tissues, cells, and serum proteins—the immune system has a decreased tolerance to itself. The main mechanisms include B-lymphocyte overactivity leading to excessive autoantibody production and increased reactivity of T lymphocytes. SLE patients can express a myriad of antibodies directed against many self molecules and antigens located in cell nuclei and cytoplasm. Antigen–antibody complexes form within the basement membranes of glomeruli in the kidneys, heart, skin, brain, joints, and other tissues. Immune complexes then activate, complement, and trigger the inflammatory responses, which are responsible for tissue destruction.

Clinical manifestations. SLE typically affects multiple organ systems such as the kidneys, heart, skin, nervous system, joints, lungs, and gastrointestinal tract. Not all systems are affected simultaneously. The characteristic clinical course is one of exacerbation and remission. A remission may last for many years.

Arthralgias and inflammatory arthritis are common features of SLE with most patients noting joint pain at some time during the course of the disease. In lupus arthritis, swelling, tenderness, pain on movement, and morning stiffness are noted. Involvement of the capsule, ligaments, and tendons can be extensive, causing reducible deformities in hands and feet (called *Jaccoud arthropathy*). Deformities range from contractures of the fingers, to hyperextension of the interphalangeal joint of the thumb, to subluxation of the MCP joint of the thumb, although bone erosions on x-ray are not seen. With steroid therapy, tendon rupture is not uncommon.

Skin manifestations may be quite extensive in SLE. Acute cutaneous lupus erythematosus often manifests with a classic butterfly (malar) rash, although some form of skin involvement can be present in 80% of patients. The skin lesion may be exacerbated during systemic flare-up. Swelling and redness are noted, and sunlight or artificial ultraviolet light may initiate a response. Skin involvement may occur on the shoulders, upper arms, upper back, chest, and neck. Scales or plaques develop on the scalp, ears, face, and neck. A latticelike venular skin change (livedo reticularis) is a common skin manifestation. Alopecia may also occur.

A number of systemic manifestations may also be present. Cardiac complications include pericarditis, valvular heart disease, and rarely myocarditis with congestive heart failure. Premature atherosclerotic heart disease is recognized as an important cause of morbidity and mortality. Lung and pleural involvement includes pleuritis or pleural effusion. More aggressive lung involvement may include acute interstitial pneumonitis, pulmonary embolus, or pulmonary hypertension. Renal involvement can be one of the more serious developments in SLE, although severity can vary. Glomerulonephritis (inflammation in the glomeruli of the kidneys) can be associated with proteinuria, hematuria, and progressive renal failure. Central nervous system involvement has also been recognized (ptosis, diplopia, ataxia, seizures, psychosis). Lymphadenopathy or Raynaud phenomenon (small digital blood vessel vasospasm in response to cold) may also be noted at some time in the course of the illness.

When SLE involves autoantibodies against blood elements, laboratory testing can reveal hemolytic anemia, leukopenia, lymphopenia, or thrombocytopenia. Renal disease often causes proteinuria, hematuria, or cellular casts on microscopic urinalysis. Although many autoantibodies cannot be measured by conventional laboratory methods, testing for the antinuclear antibodies (ANA) is important in screening for SLE. The ANA test is positive in over 95% of patients with SLE, although its specificity for lupus is low (only about 20%). The ANA can be positive in many other autoimmune diseases. Up to 30% of healthy individuals can have a low-titer ANA, although this rate decreased to about 3% with evaluation for higher titers. The frequency of a positive ANA in the absence of SLE increases with age as well. Also found in the sera of some SLE patients are autoantibodies against double-stranded DNA and other extractable nuclear antigens such as SSA (Ro), SSB (La), Smith (Sm), and RNP. Complement levels (C3 and C4) can be low in the setting of active SLE because the complement proteins are "consumed" in the antigen–antibody-mediated immune activation.

Treatment. In addition to a focus on patient education, which is of prime importance in managing SLE, the choice of therapeutic agents often depends on disease manifestations. Application of topical corticosteroids, avoidance of the sun, and use of sunblock can help control skin disease. NSAIDs, antimalarial medications, and methotrexate are useful for the management of arthritis and serositis. For more aggressive disease, including renal, hematologic, or neural involvement, oral or parenteral corticosteroids may be needed for initial control followed by immunosuppressive medications, such as azathioprine, mycophenolate mofetil, cyclophosphamide, and others. Newer biological therapies targeting activated B lymphocytes, cell costimulatory molecules, and cytokines continue to be developed. An agent for the treatment of SLE is belimumab, which is a humanized monoclonal antibody that inhibits BLys (B-lymphocyte stimulator). Inhibition of BLys leads to a decreased number of activated B cells and lowered levels of autoantibody production (such as antibodies against double-stranded DNA).

Scleroderma

Scleroderma, also called *systemic sclerosis,* is a multisystem inflammatory connective tissue disease characterized by immune dysregulation, vasculopathy, and deposition of large quantities of collagenous tissue, which results in severe tissue fibrosis. Skin, blood vessels, synovium, skeletal muscle, and microvasculature of internal organs are all affected. Scleroderma may occur in a localized form that only involves areas of skin without any internal organ involvement, or it can be a systemic disease. Two major types of systemic scleroderma are limited systemic sclerosis (LSS) and diffuse systemic sclerosis (DSS), a distinction which is made based on the extent of skin disease. It affects women three to seven times more frequently than men. Onset is most common between 30 and 50 years of age.

Etiology and pathogenesis. The cause of scleroderma is unknown, although environmental or infectious triggers in genetically susceptible individuals are suspected. Early in the disease, inflammation and immune cell infiltration can be found in skin, lungs, and other tissues. Widespread vasculopathy and vascular injury are seen with proliferation of smooth

muscle cells within vessels causing vascular wall thickening and eventual lumen obliteration, especially of small arteries, arterioles, and capillaries. Tissue ischemia results followed by diffuse tissue fibrosis. Increased amounts of collagen and other connective tissue components are produced by fibroblasts that are stimulated by local cytokines and mediators.

Clinical manifestations. Clinical manifestations often begin with Raynaud phenomenon (blanching of the digits in response to cold) and puffiness of the fingers, which can precede the development of other clinical features by many months. Raynaud phenomenon can often be idiopathic and not associated with systemic disease. However, those patients with severe Raynaud phenomenon along with typical nail fold capillary changes (capillary loops are dilated and few in number seen on capillaroscopy) are at this highest risk of developing scleroderma. Polyarthritis involving small joints of the hands is common early in the disease but can resolve later. Evaluation of the skin discloses initial bilateral swelling of the fingers, hands, and, periodically, feet. After a few weeks or months, edema is replaced by thick, tight skin that usually begins distally and progresses proximally, a feature called *sclerodactyly.* This tightening can eventually lead to severe contractures of the digits, skin ulceration at the tips, and severe loss of function. Skinfolds are lost, and a shiny appearance is noted. Hyperpigmentation or hypopigmentation may occur. In LSS the sclerodactyly involves just the distal extremities (below elbows and knees), with or without changes to the face, and other internal organ features can still occur. DSS involves the skin of proximal and distal extremities and can spread rapidly to the face and trunk. These patients are far more likely to have internal organ involvement early in the disease course. Tenosynovial involvement may be seen with tendon friction rubs, carpal tunnel syndrome, and very severe flexion contractures.

Patients with scleroderma may experience disuse atrophy of muscle because muscular motion is limited by involvement of skin and joints. A low-grade inflammatory myositis can be seen, but this is uncommon. Gastrointestinal involvement is present in a majority of patients. Musculature of the esophagus is involved, with dysmotility leading to difficulties in swallowing. Involvement of the esophageal sphincter musculature may result in reflux of gastric contents and development of peptic esophagitis. Malabsorption problems may also result from gastroparesis, intestinal dysmotility, and bacterial overgrowth. Constipation related to hypomotility of the large colon can be severe.

Pulmonary involvement, particularly pulmonary fibrosis, is the leading cause of death in systemic sclerosis. Pulmonary arterial hypertension can also develop as a result of vasculopathy of the pulmonary vasculature or from cardiac dysfunction. Manifestations of myocardial involvement include congestive heart failure, pericarditis, and atrial or ventricular dysrhythmias. Renal involvement used to be the predominant cause of death before the availability of angiotensin-converting enzyme inhibitors. In *scleroderma renal crisis,* sudden malignant arterial hypertension may rapidly progress to oliguric renal failure without immediate treatment.

A form of limited systemic sclerosis, abbreviated CREST, refers to *c*alcinosis (deposits of calcium in tissues); *R*aynaud phenomenon; *e*sophageal dysmotility; *s*clerodactyly; and *t*elangiectasias (capillary dilation that causes formation of vascular lesions on the face, lips, and fingers). Some patients with CREST syndrome develop pulmonary hypertension and intestinal malabsorption, which can indicate a poor prognosis.

Treatment. Treatment for scleroderma is largely organ specific because systemic disease-modifying agents have had disappointing results. Raynaud phenomenon is managed with avoidance of cold exposure and minimizing other vasospasm triggers (tobacco use, certain cold medicines, medication for migraine headaches, and more). Use of vasodilator medications such as calcium channel blockers (especially amlodipine or nifedipine), sympatholytic agents (such as prazosin), or, in severe cases, IV iloprost can be effective. Newer investigation suggests that phosphodiesterase-5 inhibitors (such as sildenafil or tadalafil) may be useful. Symptoms from gastrointestinal disease may be controlled with antacids, H_2 antagonists, proton pump inhibitors, and promotility agents. Previously associated with very high mortality, acute renal crisis (renal failure with malignant hypertension) can now be successfully managed with angiotensin-converting enzyme inhibitors. Pulmonary arterial hypertension can be managed with prostaglandin derivatives to include IV epoprostenol, treprostinil, and others. Endothelin receptor antagonists (such as bosentan, sitaxsentan, or ambrisentan) and phosphodiesterase-5 inhibitors (sildenafil, tadalafil, vardenafil) can improve patients' exercise tolerance and quality of life.

Ankylosing Spondylitis

Ankylosing spondylitis (AS) literally means fusion (ankylosis) of inflamed vertebra (spondylitis). It is arthritis of the sacroiliac joints that involves the axial skeleton and, in some cases, peripheral joints. The disease often begins in the spine of young males in their late teens or early 20s. The male-to-female incidence ratio is 2:1, with symptoms being somewhat more variable in women.

Etiology and pathogenesis. A strong genetic component likely plays a role in the development of AS, because approximately 90% to 95% of Caucasian patients are positive for the HLA-B27 genetic marker, and the frequency of this arthritis in different ethnic groups roughly parallels the frequency of HLA-B27 presence. The role of the HLA molecule in the pathogenesis of this arthritis is not clear, although antigen-presenting cells (expressing these HLA markers) may interact with certain bacterial or environmental factors and cross-react with self antigens found in joint tissues. Activation of immune-mediated inflammation occurs within the sacroiliac joints of the pelvis and the ligaments supporting the vertebral column. This leads to persistent back pain, stiffness, and gradual loss of mobility.

Clinical manifestations. Clinical features include the insidious onset of low back pain that improves with exercise but is not relieved by rest and the presence of severe morning stiffness for more than 3 months. Back pain in the night is common. Initial evaluation of the spine suggests an increase in muscle tone, and a loss of normal lumbar lordosis that progresses to a marked limitation of mobility noted in both anterior and lateral planes. With this limitation of movement and a position of spinal flexion, the hips and knees must compensate, creating lower extremity joint degeneration and deformities. Because of restricted postural position and decreased chest expansion, tidal volume may be diminished. A typical postural position for advanced AS is shown in Fig. 52.9.

Asymmetric peripheral arthritis with swelling of knees, ankles, or toes can also occur along with enthesitis (inflammation at the sites of tendon attachment to bone), with the plantar fascia and Achilles tendon being the most common sites. Persistent diffuse swelling of individual toes (called *dactylitis* or "*sausage toes*") can last weeks to months. Other organ systems that may be affected include the eyes (anterior uveitis, iritis), heart (aortitis, aortic valve insufficiency), and nervous system (nerve root or spinal cord impingement related to spinal deformities or fractures).

Treatment. The primary objectives of treatment are to relieve pain, decrease inflammation, and strengthen and maintain posture and function. Regular stretching and range-of-motion exercises are often recommended. Medication options include NSAIDs to reduce pain and swelling, or in severe cases, short-term corticosteroids. Disease-modifying agents, such as sulfasalazine or methotrexate, can be used mostly for peripheral joint involvement. Agents that inhibit TNF-α (etanercept,

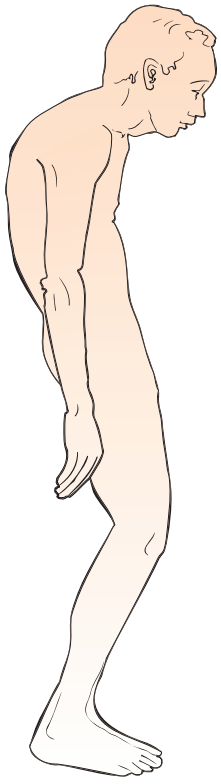


FIG 52.9 Typical posture of a patient with ankylosing spondylitis.

adalimumab, golimumab, certolizumab, or infliximab) have been shown to improve signs and symptoms of AS, although impact on bone structure changes remains a matter for ongoing research.

Postinfectious Systemic Disorders

Reactive Arthritis (Reiter Syndrome)

Historically, Reiter syndrome consisted of the triad of arthritis, urethritis, and conjunctivitis. Currently terminology renames this reactive arthritis (ReA), and it is defined as a seronegative arthritis preceded by urethritis, cervicitis, or gastroenteritis. Additional problems may include inflammatory skin lesions, oral ulcers, and keratoderma. The onset is most common between 20 and 40 years of age. Onset after a gastrointestinal tract infection is equally common in males and females, but young men are more frequently affected after a sexually transmitted infection (especially with *Chlamydia trachomatis*). ReA can be linked to the prevalence of HLA-B27. It develops in individuals who are genetically susceptible after an infection by bacteria such as *Chlamydia trachomatis* in the genitourinary tract, or *Salmonella, Shigella, Yersinia,* or *Campylobacter* in the gastrointestinal tract. It is thought that persistence of bacterial antigens and cross-reactivity of immune cells with these antigens trigger the inflammation seen in joints and tendons of persons affected by this syndrome.

Clinical manifestations. Clinically, an asymmetric oligoarthritis typically appears 2 to 6 weeks after the onset of the infectious episode. This acute arthritis onset predominantly affects knees and ankles, although upper extremity joints can sometimes be involved. Approximately 30% of patients can have acute inflammatory back pain or sacroiliitis, similar to AS. Additional features of musculoskeletal manifestations are typical: diffuse swelling of toes (dactylitis) or swelling at the Achilles tendon insertion or plantar fascia (enthesitis). The most common eye involvement in ReA is noninfectious conjunctivitis, although acute anterior uveitis (unilateral or bilateral) can occur less commonly and could lead to impairment of vision.

Cutaneous lesions related to ReA include development of small, shallow, painless ulcers on the glans penis and urethral meatus (circinate balanitis). A hyperkeratotic skin lesion (keratoderma blennorrhagicum) may form on the soles of the feet and palms of the hands. Hyperkeratosis (thickening beneath the nails) can occur.

ReA patients may have elevated inflammatory markers (sedimentation rate, C-reactive protein) as well as a mild normocytic anemia, transient leukocytosis, and thrombocytosis. RF test results are usually negative.

Treatment. Antiinflammatory medications, particularly NSAIDs, are usually effective at controlling pain and swelling, although intraarticular corticosteroid injections may also help. Second-line immune-regulating agents, such as sulfasalazine or methotrexate, can be used in persistent cases. For patients with severe refractory arthritis, biologic therapy (in particular the TNF-α inhibitors) can be effective.

Acute Rheumatic Fever

Acute rheumatic fever (ARF) is an inflammatory disease that follows a β-hemolytic group A streptococcal pharyngeal infection. The incubation period, or latent period from infection to onset of the disease, ranges from 2 to 6 weeks. The prevalence of ARF has been estimated at more than 15 million cases worldwide and may be increasing. This might be due to the reappearance of heavily encapsulated, highly virulent rheumatogenic streptococcal strains and/or to the decrease in awareness of the disease and less stringent adherence to disease control measures, especially prevention.

Etiology and pathogenesis. One theory of disease occurrence is the cross-reactivity of a patient's immune cells within the lymphoid tissue of the pharynx. Lymphocyte activity and antibody production are stimulated by streptococcal antigens, and then these cells cross-react with proteins in the target organs such as joints, heart, skin, and nervous system, leading to inflammatory reactions in these areas.

Clinical manifestations. The clinical aspects of ARF depend on the age of the affected individual. Peak incidence is between ages 5 and 15 years. Children and teenagers present with polyarthritis and carditis. Polyarthritis is usually the only manifestation in the adult. Fever is present in most patients.

Polyarthritis is the most common presenting symptom noted in patients with ARF. Pain may be quite severe over the initial few days to 1 or 2 weeks and then gradually subside. Synovial effusions and erythema may be noted. The knees, ankles, elbows, and shoulders are affected most often. Hips, wrists, and small joints of the hands and feet may also be compromised. Onset may be monoarticular with spread to numerous other joints. Joint symptoms usually respond rapidly to treatment with antiinflammatory medications, especially salicylates.

Children and teenagers with ARF are more likely than adults to develop carditis, with the youngest children (younger than 3 years) being the most likely affected. The signs of carditis include murmurs, cardiomegaly, congestive heart failure, and pericarditis. Mitral valve regurgitation is the most common murmur, followed by aortic regurgitation. Evidence of rheumatic heart disease may not be apparent for many years after the acute incident.

A rash is noted in less than 2% of individuals affected by ARF. The rash, called *erythema marginatum,* begins as a pale, erythematous, blanching macular rash over the trunk and proximal regions of the extremities. Painless nodules may cover the extensor surfaces.

Throat cultures may be negative by the time the symptoms of ARF are recognized, but certain antibody tests (antistreptolysin O, anti-DNase B) can be useful to aid in diagnosis.

Treatment. The most common therapy is NSAIDs, especially aspirin, and response can be rapid. Corticosteroids may be required in cases of severe cardiac involvement. Antibiotic therapy, including long-term prophylaxis in some cases, is generally recommended.

KEY POINTS

- A number of immune-mediated systemic connective tissue diseases result in joint dysfunction, including RA, SLE, and scleroderma. Most are classified as autoimmune disorders of unclear cause. The signs and symptoms of systemic joint disorders are generalized but involve multiple joints and usually other connective tissue structures.
- Differentiation among types of systemic joint disorders is based on patterns of joint dysfunction, immunologic factors (e.g., RF), and related lesions (e.g., the butterfly rash of SLE).
- Joint destruction in some systemic joint disorders is inflammatory in nature and involves the synovial membrane, cartilage, joint capsule, and surrounding ligaments and tendons.
- Rheumatic fever–related arthritis is a sequela of group A streptococcal infection.

JOINT DYSFUNCTION SECONDARY TO OTHER DISEASES

Psoriatic Arthritis

Psoriatic arthritis (PsA) is an inflammatory arthritis associated with psoriasis occurring in 0.056% to 0.28% of individuals in the United States, although some studies suggest a higher prevalence. Peak age of onset is 30 to 55 years of age, and the arthritis can occur in patients who have had psoriasis for many years.

Etiology and Pathogenesis

Studies have shown a strong familial tendency for PsA, suggesting that genetic factors may cause an increased predisposition to the disease. Environmental factors, such as infection or physical trauma, may trigger the onset of the arthritis. Immunologic features also have a dominant role, with activated T lymphocytes, macrophages, and neutrophils infiltrating skin and joint tissue and producing multiple inflammatory cytokines (TNF-α, IL-1, and IL-15, among others). These immune reactions cause proliferation of synoviocytes within joints, angiogenesis (new blood vessel formation), and expansion of inflammatory tissue. In the skin, keratinocytes are stimulated and will proliferate extensively.

Clinical Manifestations

There are several different patterns of joint involvement in PsA. The majority of patients have peripheral joint involvement in the form of asymmetric oligoarthritis. Other patients have a polyarthritis that is difficult to distinguish from RA. In some PsA patients, the DIP joints of the hands can be predominantly affected, which is different from RA. These patients often have psoriatic involvement of the nails with pitting or distortion of the nails (called *onycholysis*). Sacroiliac and spinal involvement can also occur, especially in patients who are positive for the HLA-B27 genetic marker.

Commonly, PsA is characterized by a combination of soft tissue and peripheral joint disease. Inflammation occurs in the joints as well as the periosteum, along the tendon, and at tendon insertions in bone (enthesitis). Fusiform swelling of the digits (dactylitis) is common.

Evidence of skin or nail changes, typical of psoriasis, is noted in PsA. These skin changes typical for psoriasis include erythematous papular lesions with characteristic scales. Nail involvement includes pitting and onycholysis (raised transverse thickening and longitudinal ridging). Subungual hyperkeratosis and oil droplet discoloration suggest psoriasis.

Inflammatory markers may be elevated, and RF and anti-CCP are usually negative. Radiographs may show minimal changes; however, some cases involve aggressive disease with bone erosion, fluffy new bone formation (periostitis), and marked loss of joint spaces. Erosive disease can develop in approximately 47% of patients within 2 years of the onset of the arthritis.

Treatment

Psoriasis may respond to topical corticosteroids, emollients, and keratolytic agents. Light therapy, utilizing ultraviolet A radiation, can also be effective. Management of the arthritis centers on NSAIDs or corticosteroids to control pain and swelling, but in many cases, more aggressive immunosuppressive therapy is needed, including methotrexate, cyclosporine, leflunomide, or sulfasalazine. In severe cases of skin and/or joint disease, TNF-α antagonists (etanercept, adalimumab, certolizumab, and infliximab) can be effective. Apremilast is an oral phosphodiesterase-4 inhibitor that can reduce signs and symptoms of PsA by regulating inflammatory mediators. Ustekinumab is a human monoclonal antibody that blocks receptor binding of IL-12 and IL-23. Its current use is primarily in the management of psoriasis, but studies have demonstrated some efficacy in PsA. There are several investigations of the use of other biologic agents that are currently used for RA management in the treatment of PsA, such as abatacept and rituximab, in addition to other cytokine targets.

Enteropathic Arthritis

Enteropathic arthritis refers to articular manifestations of two inflammatory bowel diseases (IBDs): ulcerative colitis and Crohn disease. A peripheral or axial arthritis can occur in up to 25% of patients with IBD, and in some cases the arthritis can precede the onset of gastrointestinal symptoms.

As in the case of the other spondyloarthropathies, the cause of enteropathic arthritis is unclear, but it is postulated to be associated with immune cross-reactivity with bacterial antigens. In the setting of IBD, inflammation of the gut lining may permit entrance of bacteria from the bowel lumen into the lymphoid tissue and bloodstream, stimulating an autoimmune response that can target joints.

Clinical Manifestations

Articular manifestations include peripheral arthritis, spondylitis, and involvement of muscle and bone. Peripheral arthritis is most commonly asymmetric and pauciarticular (affecting few joints), usually in the knees and ankles. Synovial inflammation may be mild to severe, and in mild cases may be transient but prone to relapse. In severe cases, a chronic and potentially erosive arthritis can develop. Spinal involvement occurs in 10% to 20% of patients, with inflammation in the spinal ligaments and sacroiliac joints, similar to features seen in patients with AS. It can be associated with HLA-B27 or HLA-B44 positivity. Like other spondyloarthropathies, enthesitis can also be noted.

Various cutaneous and ocular manifestations can occur in IBD. Skin lesions, such as erythema nodosum (painful inflamed nodules under the skin) or leg ulcers (pyoderma gangrenosum), may also be associated with the disease. Ocular manifestations, particularly acute episcleritis or bilateral anterior uveitis, occur in 5% to 8% of patients. Up to 5% of patients with IBD can develop primary sclerosing cholangitis, which can cause progressive hepatic fibrosis and failure. Amyloidosis with involvement of major organs can be observed in Crohn disease.

Anemia is common in IBD as well as leukocytosis. In the setting of arthritis, elevated inflammatory markers are often seen, and HLA-B27, when positive, may identify patients at higher risk of developing spondylitis.

Treatment

Treatment focuses on management of the gastrointestinal disease with immunosuppressive agents. Control of the arthritis can be pursued

with the use of NSAIDs (although this is limited due to gastrointestinal toxicity), COX-2 inhibitors, or corticosteroids. In some cases, second-line medications, such as sulfasalazine or azathioprine, may be required for treatment of the arthritis as well as the bowel disease. Other agents that are effective in the bowel inflammation, such as mesalamine, have limited effect on the joint disease. The use of TNF-α antagonists (adalimumab, infliximab, certolizumab) have been shown to reduce symptoms related to joint and intestinal inflammation. One of the newer agents for treatment of Crohn disease is vedolizumab, which is a monoclonal antibody against alpha-4 beta-7 integrin. It has been shown to have efficacy in bowel inflammation, although efficacy against related arthritis is limited. Other ongoing studies are looking at usefulness of agents that target other inflammatory cytokines, such as interferon-γ or IL-6.

Neuropathic Osteoarthropathy

Commonly called *Charcot joint,* neuropathic osteoarthropathy is a neurologic disease that leads to bone and joint abnormalities. The mechanism is not clear, but because of a loss in normal proprioception and pain responses, damage occurs to the joint. The mechanics of disease development are probably a combination of neurovascular and neurotraumatic processes. Peripheral nerve injuries, diabetes mellitus, pernicious anemia, alcoholism, and multiple sclerosis can lead to Charcot joint. Motor neuron involvement can affect both upper and lower motor neurons. Diabetes, tabes dorsalis, and syringomyelia are the three most prevalent disease processes that lead to neuropathic osteoarthropathy.

Clinical Manifestations and Treatment

Clinically, the patient presents with a swollen, deformed, and unstable joint. Radiographs reveal advanced joint destruction and pathologic fractures. Management requires protection of the involved joint through immobilization and less weight bearing. Surgical intervention has shown poor results attributable to nonunion, dislocation, or infection.

Hemophilic Arthropathy

Bleeding into joints, as noted in hemophilia, causes extension of the joint capsule and a limitation of movement. Hemorrhage stimulates a synovial proliferative response, chronic inflammation with a release of degradative proteinase, and changes in cartilage composition with less resistance to stress. Chronic synovitis alters the synovial lining and eventually leads to joint destruction.

Clinical Manifestations

Three stages of hemophilic arthropathy are described. The acute stage manifests with bleeding in the joint and occurs as the child begins to walk. Bleeding into the confined area of the capsule causes the joint to be positioned in flexion and increases stress to articular structures. Atrophy of muscles around the joint predisposes it to further hemarthrosis. The second stage is due to repetitive hemorrhages into the joint, resulting in chronic synovitis. The joint is edematous and warm but painless. The third stage is characterized by destruction of joint integrity.

Larger joints are affected more frequently than smaller joints. Seldom are structures of the wrist involved. Elbows, hips, and knees are subject to major destruction. Muscle hemorrhage (iliopsoas, forearms, gastrocnemius), muscle cysts, and pseudotumors (attributable to osseous hemorrhage) may develop. Medical treatment to enhance clotting is imperative. Education and prevention of joint deformity are essential in the management of hemophilia.

Gout

Gout is a heterogeneous disorder in which disturbance of uric acid metabolism leads to deposition of monosodium urate salts in articular, periarticular, and subcutaneous tissue (Fig. 52.10). It is also characterized by hyperuricemia and urate crystal–induced arthritis. Gout arises in humans because of a lack of the enzyme uricase and subsequent inability to oxidize uric acid to a soluble compound. Uric acid is a normal waste product of purine metabolism and therefore must be filtered primarily by the kidneys. When production of uric acid exceeds removal, hyperuricemia results and the deposition of crystalline sodium urate increases. The acute attack is often triggered by a traumatic event, a surgical procedure, an acute illness, or use of alcohol or medications.

Clinical Manifestations

Epidemiology surveys in 2007 to 2008 showed the prevalence of gout in the United States to be approximately 3.9%, with the rate being 9.3% in individuals over the age of 60. Risk of developing gout increases with age and with an increase in serum urate concentrations. Clinically, acute gouty arthritis is the form most frequently observed. Gouty arthritis is common in middle-aged men and postmenopausal women. A familial tendency is often noted.

Manifestations of gout include recurrent episodes of articular and periarticular inflammation (acute gouty arthritis), accumulation of tophi (crystalline deposits) in bony and connective tissue, impairment of the renal system, and the presence of uric acid calculi. There are four phases in gout: asymptomatic hyperuricemia, acute gouty arthritis, intercritical gout, and chronic tophaceous gout.

Asymptomatic Hyperuricemia

Hyperuricemia has been estimated at highly variable rates in Americans, from 2% to 40%, who are asymptomatic. A number of factors and clinical syndromes can be associated with hyperuricemia either by excessive urate production or by reducing renal excretion. Diet can play a significant role in increasing uric acid levels and subsequent gout development. Foods rich in purines are often associated with this, including organ meats, red meat, and seafood. Vegetables with high purine content do not increase risk. Carbonated beverages or colas containing high-fructose sweeteners can also lead to hyperuricemia. Ample data linking alcohol consumption with hyperuricemia and gout exist, with beer or ales being the highest risk. Interestingly, consumption of dairy products, such as milk, may reduce hyperuricemia somewhat.

Congenital enzyme deficiencies can lead to excessive urate production and hyperuricemia (Lesch–Nyhan syndrome, for example). Other clinical syndromes that are associated with extensive cell turnover can also lead to hyperuricemia. These can include hematologic disorders (hemolytic anemia, polycythemia vera, myelodysplastic syndromes, and others). Cancer treatment can lead to tumor lysis syndrome, which can also cause a marked increase in uric acid production.

Chronic kidney disease is also often associated with hyperuricemia as the decline in renal function impairs uric acid excretion.

In a patient with asymptomatic hyperuricemia, there are no clinical signs; however, the serum urate level is elevated. In the male, hyperuricemia can begin at puberty. In women, hyperuricemia usually does not appear before menopause. No treatment is required at this stage.

Acute Gouty Arthritis

Gouty arthritis is the most common early clinical sign. An affected joint is warm, red, and very tender to even light palpation. The MTP joint of the great toe is most often the earliest joint involved, although this is not always the case. Ankle, tarsal, and knee joints are often affected, and attacks can also occur in hands, wrists, and elbows. The first attack of acute gouty arthritis is often sudden onset with an intense pain that can awaken the patient from a sound sleep. Diffuse periarticular erythema often accompanies the attack. Diagnosis relies on classic clinical presentation, hyperuricemia, and the demonstration of urate crystals in synovial fluid of the involved joint.

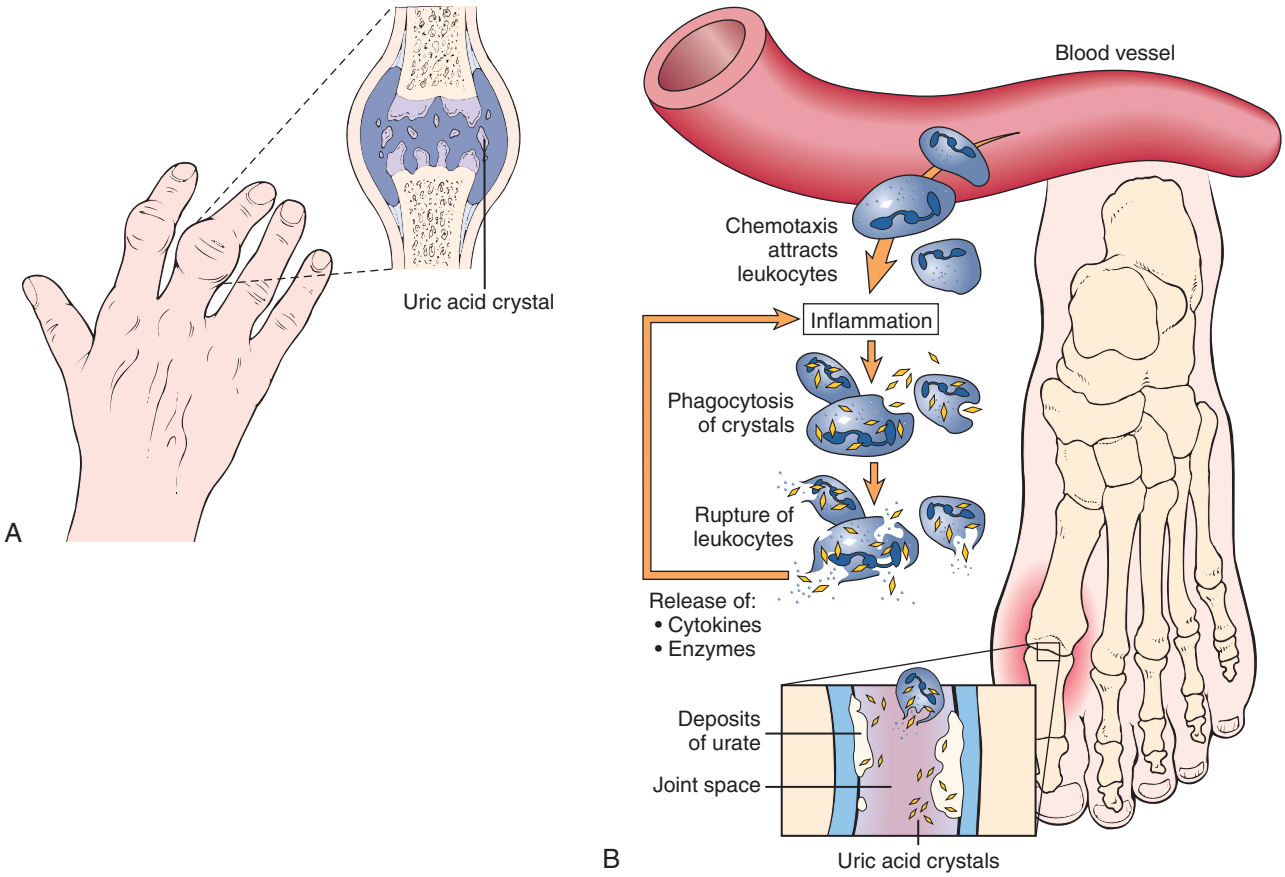


FIG 52.10 **A,** Gout. **B,** Gouty arthritis. Deposits of uric acid crystals in the connective tissue have a chemotactic effect and cause exudation of leukocytes into the joint. The inflammation most often affects the metatarsophalangeal joint. (A, From Frazier MS, Drzymkowski JW: *Essentials of human diseases and conditions,* ed 5, Philadelphia, 2013, Saunders. B, From Damjanov I: *Pathology for the health professions,* ed 4, Philadelphia, 2012, Saunders.)

Initial attacks subside within 1 or 2 days or may last 1 to 2 weeks, after which the patient is symptom-free until the next episode. Later attacks tend to become more frequent, and mild arthralgia may occur between episodes.

Intercritical Gout

Intercritical gout, the name for the disease in the intervals between acute attacks, presents no symptoms. Even during asymptomatic periods, urate crystals can be aspirated from involved joints, and deposition in soft tissues can be ongoing.

Chronic Tophaceous Gout

This is an advanced stage of gout. Tophi begin to appear approximately 10 years after the initial onset of gout. Tophi occur commonly in the synovium, subchondral bone, olecranon bursa, and infrapatellar and Achilles tendons. Tophi have been noted in walls of the aorta, valves of the heart, ear cartilage (pinna), corneas, sclerae, and kidneys.

As a result of deposition of crystals and chronic inflammation, deforming arthritis can develop. Development of tophi in tendon sheaths of the hand and wrist can cause a trigger finger or carpal tunnel syndrome.

Patients with gout may have involvement of the kidneys and develop renal malfunction. These individuals have a higher incidence of arterial hypertension, diabetes mellitus, and cardiac and cerebral atherosclerosis. Hypertriglyceridemia occurs more frequently in the patient diagnosed with gout.

Treatment

Management of an acute gouty attack usually requires aggressive antiinflammatory medication such as NSAIDs or corticosteroids (oral, parenteral, or intraarticular). Colchicine may also be used early in the course of an attack or in lower doses as a prophylactic agent. Medications to correct hyperuricemia and prevent gout exacerbation may target uric acid excretion by the kidneys (uricosuric agents) or uric acid production (allopurinol or febuxostat). Pegloticase is a porcine uricase enzyme recently approved for treatment of patients with refractory gout. It is administered as an IV infusion every 2 weeks and has a number of side effects, such as infusion reactions, worsening gout flare-ups, gastrointestinal side effects, and congestive heart failure. The practical usefulness of pegloticase is still uncertain and may be best reserved for those with chronic, severe active gout that is not responsive to other treatment options or in patients in whom other agents are contraindicated.

Adult-Onset Still Disease

Adult-onset Still disease is a form of seronegative (i.e., negative RF) polyarthritis with a number of symptoms similar to those of systemic-onset juvenile rheumatoid arthritis (JRA) in children (see the "Pediatric Joint Disorders" section in this chapter). Adult-onset Still disease may follow diagnosis of RA. Its cause is unknown.

Clinical Manifestations and Treatment

Clinical features include high-spiking fever, a rash on the trunk and extremities, and, possibly, a sore throat. Polyarthritis usually affects the

PIP and MCP joints of the hands, but can also involve the wrist, knees, hips, and shoulders. Visceral involvement includes hepatic insufficiency, chronic respiratory failure, cardiac tamponade, congestive heart failure, and splenomegaly. Laboratory features may include anemia, leukocytosis, elevated sedimentation rate, thrombocytosis, and elevated liver enzymes.

Some patients respond well to high-dose aspirin or NSAIDs during the acute illness, although in severe cases corticosteroids are used.

KEY POINTS

- PsA differs from RA in that PsA is more often asymmetric, is associated with psoriasis in the skin, and is not associated with autoantibody (RF or CCP) production. It can, however, lead to joint erosions and damage.
- Neurovascular, hematologic, and metabolic disorders may lead to associated disorders of joint function. Diabetes, for example, results in neurovascular changes that desensitize the joint and predispose it to traumatic joint dysfunction. Hemophilia predisposes to intraarticular bleeding.
- Altered uric acid metabolism leads to deposition of uric acid crystals in joints, causing inflammation and gouty arthritis.

PEDIATRIC JOINT DISORDERS

Pediatric rheumatic diseases include more than 100 illnesses associated with arthritis and musculoskeletal syndromes. Soft-tissue pain and restrictions constitute a major proportion of the complaints presented to pediatric rheumatologists.

Nonarticular Rheumatism

"Growing pain," or nonarticular rheumatism, is a common benign syndrome in about 10% of children, most often between 4 and 15 years of age. Nocturnal pain, usually occurring in the calves, shins, and thighs, is the most common symptom. Although this problem seems to be benign, medical consultation and education concerning the problem are essential.

Hypermobility of Joints

Hypermobility of joints is a common cause of complaints of pain in the joints. Mobility may be excessive in any joint, but it is most apparent in passive apposition of the thumb to the forearm, hyperextension of the fingers parallel to the forearm, and excessive extension (greater than 10 degrees) of the knees and elbows.

Children with one of several types of congenital disorders of connective tissue, the Ehlers–Danlos syndromes (EDS), can have marked joint hypermobility and increased skin elasticity and fragility. There are six major clinical subsets of EDS with the classification based on clinical features, including extraarticular involvement (eyes, vascular defects, valvular heart dysfunction, and much more).

Juvenile Idiopathic Arthritis

Juvenile idiopathic arthritis (JIA; previously called JRA) is a heterogeneous group of chronic inflammatory childhood syndromes. Some forms of JIA peak in early childhood but can abate in adulthood though sometimes with residual joint damage. JIA begins with synovial inflammation of unknown cause and affects approximately 100 per 100,000 children under the age of 16 years in the United States and Europe. JIA may be defined as arthritis in one or more joints (pain, redness, swelling, warmth, and limited range of motion) with a duration of at least 6 weeks. The classification and nomenclature of the types of JIA have been reorganized and rewritten a number of times over the years. One such classification by the International League of Associations for Rheumatology divides JIA into seven types based on joint involvement and extraarticular features present in the first 6 months of disease. These types include (1) oligoarticular, (2) polyarticular with RF negative, (3) polyarticular with RF positive, (4) PsA, (5) enthesitis-related (B27+ spondyloarthritis, (6) systemic, and (7) undifferentiated arthritis.

Clinical Manifestations

By definition, a child with oligoarticular (also called *pauciarticular*) onset of JIA has arthritis in four or fewer joints. This type is most common in young girls (2 to 3 years of age), although boys are occasionally affected. One subset of this form of JIA includes patients who have a positive ANA test result, which can signal an increased risk of developing inflammatory ocular disease (uveitis), a complication that may start with minimal or no symptoms yet can lead to severe irreversible vision impairment.

A polyarticular onset of JIA (i.e., involving five or more joints) is seen in approximately 10% to 28% of patients (somewhat less when including only those who are RF positive), with peak onset at age 8 to 16 years. RF positivity is also more common in girls with later onset of disease that can resemble adult RA. These patients are at a higher risk of developing progressive bone erosions, nodules, and poor functional outcome. Malaise, growth retardation or weight loss, low-grade fever, and anemia are other clinical manifestations. The patients with RF-negative polyarticular JIA can have a more variable presentation and may be younger age, and some may resemble those with oligoarticular disease, including a subset that are ANA positive.

The onset of joint swelling in the presence of psoriasis or psoriatic-like findings (nail pitting, dactylitis) constitutes another subgroup of JIA. These patients may have a family history of psoriasis and can present with an oligoarticular form (similar to oligoarticular JIA) or a later-onset form that resembles adult PsA.

Another type of childhood inflammatory arthritis is a juvenile spondyloarthritis or enthesitis-related JIA. These children are more commonly boys, may be HLA-B27 positive, and can sometimes be found to have sacroiliitis on MRI. They can occasionally present with a peripheral arthritis of large joints (hips, knees, shoulders) as well as eye inflammation (uveitis), but this is not common.

Systemic-onset is noted in approximately 2% to 17% of children with JIA, with peak onset at 1 to 6 years of age. Clinical manifestations include spiking fevers (103°F to 104°F), daily or twice daily and usually in the afternoon with return to baseline without antipyretics; transient, pale pink rash; lymphadenopathy; hepatosplenomegaly; and pericardial or pleural effusions. Fatigue, muscle atrophy, and weight loss can be severe. Anemia, leukocytosis, and thrombocytosis are common. RF is usually negative. Musculoskeletal findings in the early stages of the disease include recurrent arthralgia, myalgia, and transient arthritis, which are concurrent with fever spikes. Polyarthritis can develop weeks to months after the onset of the disease. Severe chronic arthritis may continue after the systemic symptoms subside.

A final subgroup is undifferentiated JIA, which is a miscellaneous group of children with inflammatory arthritis that does not fit into any of the other categories or whose disease has overlapping clinical features.

Treatment

All forms of JIA can cause general growth retardation, although it is more of a risk in children with systemic or polyarticular onset of disease. Inflammation close to the epiphyseal plates can result in altered growth of long bones. It is therefore imperative to achieve early diagnosis and implement appropriate treatment to minimize deformity and disability. Education and counseling are also important. Relief of symptoms and maintenance of joint position and muscle function are immediate goals of treatment.

Pharmacologic intervention is an important component of the treatment regimen. Drug therapy is instituted to decrease pain and arrest progression of the disease. The following categories of drugs are used: (1) antiinflammatory analgesics (aspirin or NSAIDs), (2) corticosteroids, (3) disease-modifying drugs (methotrexate, sulfasalazine, leflunomide), and (4) biological disease-modifying agents (such as the TNF-α inhibitors etanercept and adalimumab). Additional biological agents that can be used in severe JIA include tocilizumab, which inhibits activity of IL-6, and abatacept, which blocks T-cell activation. Biological agents targeting other inflammatory cytokines, such as IL-1 (anakinra), have also shown some efficacy in JIA. Physical and occupational therapy assessment and treatment plans are important, and daily activity should be an integral part of the child's lifestyle. Joint support and physical activity help prevent joint contracture.

KEY POINTS

- JIA has seven subtypes: oligoarticular onset (which involves four or fewer joints); polyarticular onset (in which symptoms are primarily localized to five or more joints, with or without RF positivity); psoriasis related; enthesitis related (juvenile spondyloarthritis), systemic onset (which has more systemic manifestations, including rash, high fever, lymphadenopathy, splenomegaly, fatigue, and polyarthritis); and undifferentiated arthritis (arthritis that does not fit well into any of the other categories.

SUMMARY

This chapter has provided an overview of major rheumatic disorders. Broadly speaking, interventions must assist in controlling disease activity, managing pain, minimizing deformity, and maintaining or restoring function. Depending on the stage of the disease, correct intervention must be implemented using knowledge of joint physiology biomechanics and pathologic changes resulting from the disease.

The challenge to the health professional is to ensure that an inflammatory response is not exacerbated while the body structures are being stimulated to increase strength, enhance nutrition, and improve tolerance to stress. Long periods of immobilization, bed rest, and sedentary behavior are counterproductive in the patient with arthritis. Lack of activity poses particular problems in people with arthritis. Deleterious effects on muscle strength, reflexes, connective tissue extensibility, and cardiovascular fitness are identifiable in the immobilized arthritic patient. It is imperative that the health professional look beyond disease-specific interventions and prescribe a well-developed exercise program for patients with arthritis. Achievement or retention of as much function as possible is the ultimate goal for an individual with arthritis. Every level of intervention must be directed to achievement of specific performance goals cooperatively developed by clinician and patient.

RESOURCES

Aletaha D, Neogi T, Silman AJ, et al: 2010 Rheumatoid arthritis classification criteria. *Arthritis Rheum* 62(9):2569–2594, 2010.

Bennett JE, Dolin R, Blaser MJ, editors: *Mandell, Douglas, and Bennett's principles and practice of infectious disease*, ed 8, Philadelphia, 2015, Saunders, Elsevier.

Firestein GS, Budd RC, Gabriel SE, et al: *Kelleyand Firestein's textbook of rheumatology*, ed 10, Philadelphia, 2017, WB Saunders.

Fleischmann R, Kremer J, Cush J, et al: Placebo-controlled trial of tofacitinib monotherapy in rheumatoid arthritis. *N Engl J Med* 367(6):495–507, 2012.

Furie R, Petri M, Zamani O, et al: A phase III, randomized, placebo controlled study of belimumab, a monoclonal antibody that inhibits B lymphocyte stimulator, in patients with systemic lupus erythematosus. *Arthritis Rheum* 63(12):3918–3930, 2011.

Hochberg MC, Silman AJ, Smolen JS, et al: *Rheumatology*, ed 6, Philadelphia, 2015, Elsevier.

Hunder GG, editor: *Atlas of rheumatology*, ed 2, Philadelphia, 2001, Current Medicine.

Kavanaugh A, Mease PJ, Gomez-Reino JJ, et al: Longterm (52-week) results of a phase III randomized, controlled trial of apremilast in patients with psoriatic arthritis. *J Rheumatol* 42:479–488, 2015.

Klippel JH, Stone JH, Crofford LJ, White PH, editors: *Primer on the rheumatic diseases*, ed 13, New York, 2008, Springer.

Stoll ML, Cron RQ: Treatment of juvenile idiopathic arthritis in the biologic age. *Rheum Dis Clin Nth Am* 39(4):751–766, 2013.

Sundy JS, Baraf HSB, Yood RA, et al: Efficacy and tolerability of pegloticase for the treatment of chronic gout in patients refractory to conventional treatment. *JAMA* 306(7):711–720, 2011.

53

Alterations in the Integumentary System

Lee-Ellen C. Copstead, Ruth E. Diestelmeier, and Michael R. Diestelmeier

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- How does the aging process affect the integumentary system?
- Why is it important to differentiate primary from secondary skin lesions?
- What lesion characteristics are assessed to aid in determination of the lesion's cause?
- How do systemic disorders affect nail and hair growth?
- Which skin disorders are likely to occur more commonly in certain age groups, including infants, children, adolescents, and the elderly?
- How does ultraviolet radiation affect the skin?
- How do superficial and deep pressure ulcers differ in clinical and etiologic features?
- How can malignant melanoma be differentiated from other skin lesions?

CHAPTER OUTLINE

Age-Related Changes, 1059
Epidermis, 1060
Dermis and Subcutaneous Tissue, 1060
Appendages, 1061
Hair, 1061
Nails, 1061
Glands, 1061
Evaluation of the Integumentary System, 1061
Primary and Secondary Lesions, 1061
Lesion Descriptors, 1062
Selected Skin Disorders, 1063
Infectious Processes, 1064
Viral Infections, 1064
Verrucae, 1064
Herpes Simplex Virus, 1064
Herpes Zoster Virus, 1065
Fungal Infections, 1065
Superficial Fungal Infections, 1065
Yeast Infections, 1065
Bacterial Infections, 1067
Impetigo, 1067
Syphilis, 1069
Leprosy, 1069
Inflammatory Conditions, 1069
Lupus Erythematosus, 1069
Seborrheic Dermatitis, 1069
Psoriasis, 1069
Lichen Planus, 1071
Pityriasis Rosea, 1071
Acne Vulgaris, 1071
Pemphigus, 1072
Allergic Skin Responses, 1072
Atopic Dermatitis, 1072
Contact Dermatitis, 1073
Drug Eruptions, 1074
Vasculitis, 1074
Parasitic Infestations, 1075
Scabies, 1075
Fleas, 1075

Lice, 1075
Chiggers, 1075
Bedbugs, 1075
Mosquitoes, 1076
Blood Flukes, 1076
Ticks, 1076
Rocky Mountain Spotted Fever, 1076
Lyme Disease, 1076
Other Disorders of the Dermis, 1076
Scleroderma, 1076
Localized Scleroderma, 1077
Diffuse Scleroderma, 1077
Sunburn and Photosensitivity, 1077
Effects of Sunlight, 1077
Ulcers, 1078
Altered Cell Growth: Epidermal Proliferation, 1078
Tumors, 1079
Cancer, 1079
Pigmentation Alterations, 1080
Vitiligo, 1080
Albinism, 1081
Special Characteristics of Dark Skin, 1081
Integumentary Manifestations of Systemic Disease, 1083
Skin, 1083
Color, 1083
Sensation, 1083
Texture, 1083
Temperature, 1083
Hair, 1084
Growth, 1084
Amount, 1084
Color, 1084
Texture, 1084
Lubrication, 1084
Nails, 1084
Shape, 1084
Color, 1084
Texture, 1085
Treatment Implications, 1086
Topical Treatment, 1086
Wet Dressings, 1086
Lotions, 1086
Gels, 1086
Creams, 1086
Ointments, 1086
Aerosols and Foams, 1086
Intralesional Injection, 1086
Selection of a Delivery System, 1087
Corticosteroids, 1087
Systemic Steroids, 1087
Topical Steroids, 1087
Developmental Considerations, 1087
Infancy, 1087
Childhood Skin Disorders, 1088
Rubella, 1089
Roseola Infantum, 1089
Measles, 1089
Chickenpox, 1090
Scarlet Fever, 1090
Adolescence and Young Adulthood, 1090
Geriatric Considerations, 1090

This chapter focuses on altered structure and function of the integumentary system. The etiologic factors, pathogenesis, and clinical manifestations of selected skin disorders, as well as general considerations regarding treatment modalities and their therapeutic application, are described.

AGE-RELATED CHANGES

The skin undergoes dramatic changes from birth through the mature years. Healthy infants and young children have relatively smooth and unwrinkled skin characterized by elasticity and flexibility. Because skin tissues are in an active phase of new growth, healing of skin injuries is often rapid and efficient. Young children and elderly individuals have fewer sweat glands than adults do, so their bodies rely more on increased blood flow to maintain a normal body temperature.

As adulthood begins at puberty, hormones stimulate the development and activation of sebaceous glands and sweat glands. After the sebaceous glands become active, especially during the initial years, they may overproduce sebum and thus give the skin an unusually oily appearance. Sebaceous ducts may become clogged or infected and form acne pimples or other blemishes on the skin. Activation of apocrine sweat glands during puberty causes increased sweat production, an ability needed to maintain an adult body properly, and also the possibility of increased "body odor." Body odor is caused by wastes produced by bacteria that feed on the organic compounds found in apocrine sweat and on the surface of the skin.

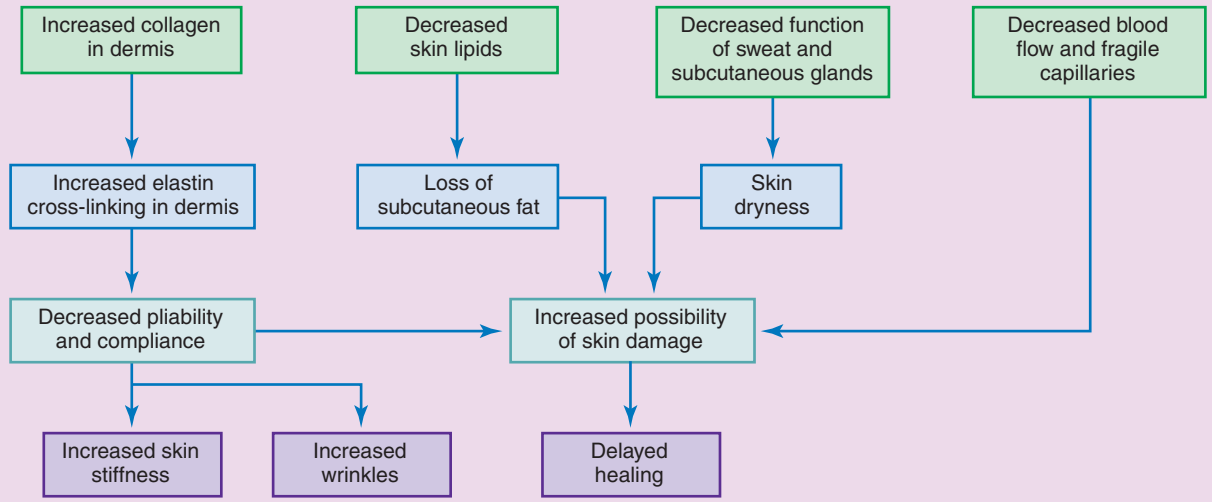
Past early adulthood and into middle age, the sebaceous and sweat glands become less active. Although this can provide relief to those who suffer from acne or other problems associated with overactivity of these glands, it can affect the normal function of the body. For example, the reduction in sebum production can cause the skin and hair to become less resilient.

Changes in the appearance and function of the skin, perhaps more than in any other organ, reflect the continual aging process (see Geriatric Considerations: Changes in the Integumentary System). One need only look at a person to determine an approximate age. Evidence of advancing age includes wrinkling and sagging skin, gray hair, and baldness. Aging changes are also linked to environmental influences, genetic makeup, and other bodily changes (Fig. 53.1).

Exposure to sunlight is one of the greatest factors in age-related skin changes. The result of such exposure can be seen in people who work outdoors in sunlight. Results are also evident when skin exposed to sunlight is compared with unexposed skin. Skin that is usually covered shows little change with age. Blue-eyed, fair-skinned individuals are more susceptible to solar skin damage than are people with darker, more heavily pigmented skin.

GERIATRIC CONSIDERATIONS

Changes in the Integumentary System



With aging, the skin's protective functions decline. Although the thickness of the stratum corneum remains the same, the properties of the water and chemical barriers in this layer of the integument are less effective. In the epidermis, mitosis decreases and cellular variation increases. The thickness of the epidermis is unchanged. The number of melanocytes decreases in Caucasians, with declining function. The melanocytes are less efficient and lack uniformity in pigment production with sun exposure.

There is a decrease in both the thickness and the amount of subcutaneous fat in the dermis, and there is an increase in the amounts of collagen and decreased elastin with cross-linking and calcification of elastin fibers. These changes cause a loss of skin pliability, compliance, and resiliency. There is an accompanying escalation in skin stiffness and an increase in wrinkling of the skin. Sebaceous and sweat gland function declines, resulting in drier, less oily skin. The number of sensory nerves and blood vessels in the skin declines, resulting in decreased sensation and loss of effective vasoactivity by dermal arterioles.

Nail and hair growth declines. The nails may become yellowed and thickened. Graying of the hair is due to the loss of melanocytes at the hair follicle base. The degree and pattern of hair loss are affected by genetic and endocrine factors. Body hair patterns change, with thinning of leg, axillary, and pubic hair.

The cumulative effect of these skin changes is loss of the regulatory, secretory, and excretory properties of the skin. The skin becomes injured more easily and, once injured, heals more slowly.

FIG 53.1 Physiologic signs of aging human skin. (Photo by Lee-Ellen C. Copstead.)

Epidermis

The epidermis shows a generalized thinning with advancing age, although there may be some thickening in sun-exposed areas. Although there is an increased variation in epidermal thickness, the average number of cell layers remains unchanged. The prickle cells of the inner layer of the epidermis show greater variation in nuclear and cytoplasmic size with a less orderly arrangement of cells. Cells reproduce more slowly and are larger and more irregular; however, exposed epidermal cells may divide more frequently than unexposed cells.

Dermis and Subcutaneous Tissue

The dermis contains blood vessels, nerves, hair follicles, and sebaceous glands, but the major portion is composed of collagen and elastin. The elasticity of the skin is largely due to dermal elastin. Decreased skin strength and elasticity with aging are attributed to a decreased amount of elastin and a proportionate increase in the collagen-to-elastin ratio. Collagen fibers change with age, becoming cross-linked and rearranged into thicker bundles. This condition is called elastosis and is closely associated with exposure to sunlight (solar elastosis). It produces a weather-beaten or tanned appearance.

Aging also produces a decrease in the vascularity of the dermal skin, as evidenced by decreasing numbers of epithelial cells and blood vessels. There is greater vascular fragility, leading to the frequent appearance of hemorrhages (senile purpura); cherry angiomas; venous stasis; and venous lakes on the ears, face, lips, and neck. The decreased vascularity and circulation in the dermis and the underlying subcutaneous tissue also have an effect on drug absorption. Drugs administered

subcutaneously are absorbed more slowly, thus prolonging their half-life. The amount of subcutaneous fat tissue also decreases, especially in the extremities, so that arms and legs appear to be thinner.

Appendages

Hair

The most obvious change in aging hair is its color. Half of the population over age 50 years has at least 50% gray body hair, regardless of gender or hair color. Gray hair is determined by an autosomal-dominant gene and results from a decreased rate of melanin production by the hair follicle. Hair color generally darkens with age, but this process is reversed with the onset of graying. Graying usually begins at the temples of the head and extends to the vertex of the scalp. It may not occur in the axilla, especially in women, and occurs to a lesser extent in the presternum or the pubis.

Changes in hair growth and distribution are also associated with aging. The amount and distribution of hair are determined by racial, genetic, and sex-linked factors; however, almost all older people have a diminution of body hair except on the face. Adults develop a full terminal hair pattern by age 40 years, and this is followed by a progressive loss of hair in reverse order of development. Postmenopausal Caucasian women lose trunk hair first, then pubic and axillary hair. Unopposed adrenal androgens produce coarse facial hair in 50% of Caucasian women older than 60 years, especially on the chin and around the lips.

Men also show a general thinning of hair distribution, with the hairs of the eyebrows, ears, and nose becoming longer and coarser. Baldness is often a concern, particularly in aging men, although women also tend to show some thinning of scalp hair. Frontal recession of the hairline occurs in 80% of older women and 100% of older men. Baldness in men is inherited from the mother and occurs only in the presence of testosterone. Onset is variable and is manifested by an M-shaped pattern of hair loss on either side of the midline or by a thinning patch over the vertex.

In general, the hair of both men and women changes from darker, thicker, and more numerous to lighter, thinner, and less numerous with aging. Hair changes begin in midlife and become highly noticeable in later life, especially after age 60 years. Women seem to manifest more hair loss on the trunk and extremities, whereas men have greater hair loss on the head.

Nails

With aging, nails become dull, brittle, hard, and thick. Most nail changes are due to a diminished vascular supply to the nail bed. There is approximately a 30% to 50% decrease in the growth rate of fingernails, from 0.1 mm/day in 30-year-olds to 0.07 mm/day in 90-year-olds. Aging nails show an increase in longitudinal striations, which can cause splitting of the nail surface.

Toenails are particularly prone to hyperkeratosis and resultant thickening. Pressure and trauma from poorly fitting footwear may be a significant factor, but onychomycosis, which affects approximately 20% of individuals over age 60, is the primary factor.

Glands

Sebaceous glands show little atrophy or histologic change with age; however, their function tends to diminish, as evidenced by a decrease in sebum secretion. In men the decrease is minimal, but in women there is a gradual diminution in sebum secretion after menopause, with no significant changes after the seventh decade. There are fewer sebaceous glands in older individuals, which appear related to the loss of hair follicles. The decrease in sebum secretion and in the number of sebaceous glands results in the drier, coarser skin associated with aging.

Sweat glands generally decrease in size, number, and function with age. In the eccrine glands, the secretory epithelial cells become uneven in size, ranging from normal to small, and there is a progressive accumulation of lipofuscin in the cytoplasm. In the very old, the secretory coils of many eccrine glands are replaced by fibrous tissue, which drastically diminishes their capacity to produce sweat. The thermal threshold for sweating is raised so that the amount of sweat output at a body temperature of 38°C (100.4°F) decreases. This may be due to the fact that there are fewer blood vessels and nerve cells around the glands that enable the body to respond to temperature changes. Apocrine glands do not decrease in number or size, but they do decrease in function. An accumulation of lipofuscin has also been noted in apocrine glands. The diminished functioning of sweat glands in the elderly greatly impairs the ability to maintain body temperature homeostasis.

Box 53.1 summarizes the morphologic features of aging human skin, and Fig. 53.2 summarizes the histologic changes associated with aging in normal human skin.

BOX 53.1 Morphologic Features of Aging Human Skin

Epidermis	Dermis	Appendages
Flat dermoepidermal junction	Atrophy	Graying of hair
Variable thickness	Fewer fibroblasts	Loss of hair
Variable cell size and shape	Fewer blood vessels	Conversion of terminal to vellus hair
Occasional nuclear atypia	Shortened capillary loops	Abnormal nail plates
Loss of melanocytes	Abnormal nerve endings	Fewer glands

From Gilchrest BA: Skin. In Rowe JW, Besdine RW, editors: *Health and disease in old age,* Boston, 1982, Little, Brown, p 383.

KEY POINTS

- The glandular function of skin varies considerably with age. Young children and elderly adults have fewer functional sweat glands and therefore less efficient evaporative heat loss capabilities. Sebaceous glands are particularly active during puberty, causing a predisposition to acne; they become less active with age, causing a predisposition to dry skin.
- The epidermis and dermis undergo degenerative changes with aging. The epidermis thins, and the dermis becomes less elastic and less vascular. The amount of subcutaneous fat decreases. Exposure to sunlight is an important factor in the development of aged skin.
- Graying of hair results from decreased melanin production by the hair follicle. After age 40 years, progressive hair loss occurs. Male pattern baldness is an inherited trait that is mediated by testosterone.

EVALUATION OF THE INTEGUMENTARY SYSTEM

A careful examination of the skin yields valuable information that may aid in identifying a systemic disease or a specific problem of the skin or appendages. Diagnostic evaluations include a careful history, and Table 53.1 provides a general guide. A proper skin examination also describes the objective signs of dermatologic disease, including all types of lesions and their distribution.

Primary and Secondary Lesions

Physical descriptions should include the lesions and their classification, generally primary (original appearance) or secondary (appearance modified by normal progress over time or by such external agents as scratching). Fig. 53.3 shows clinical examples of primary and secondary lesions.

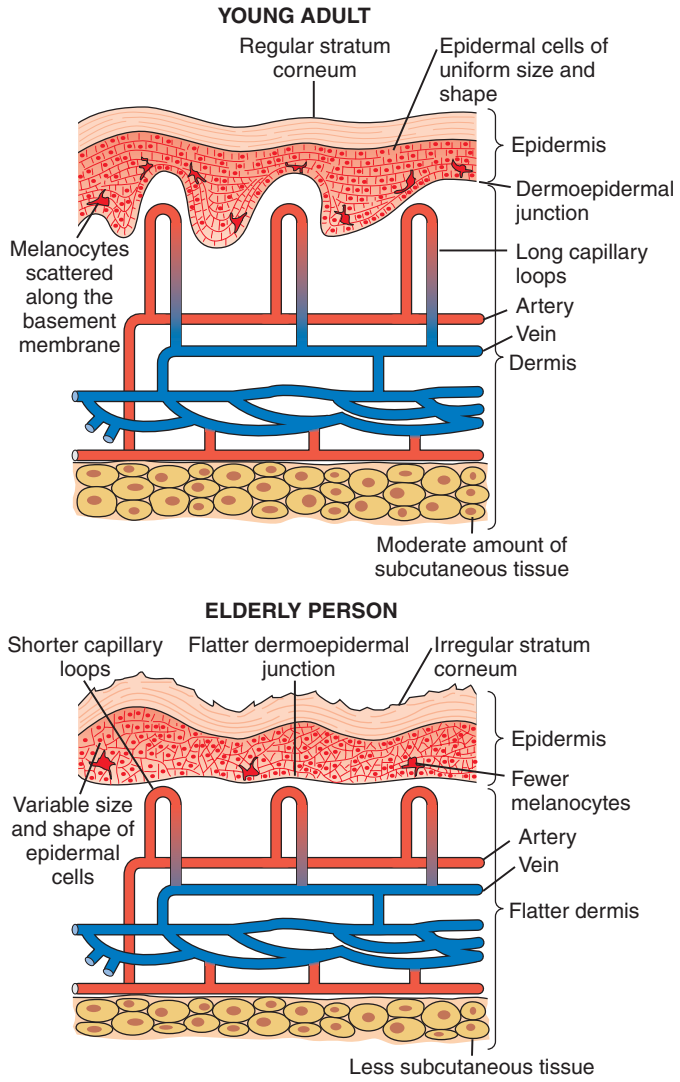


FIG 53.2 Histologic changes associated with aging in normal human skin. Note the flattening of the dermoepidermal junction and the shortening of capillary loops in older skin. Variability in size and shape of epidermal cells, irregularity of stratum corneum, and loss of melanocytes are also apparent. Age-associated loss of dermal thickness and subcutaneous fat is also illustrated.

Lesion Descriptors

After a skin lesion has been classified as primary or secondary, other features should be noted, particularly size, symmetry of color and shape, and distribution if more than one lesion is present. Skin lesions may assume a wide range of colors: red–salmon pink, brown-black, blue-purple, bone white–slate gray, and yellow, to name a few. Each color suggests certain diagnoses. Skin lesions may be solitary, few, or profuse. When more than one lesion is present, the distribution pattern may be important in suggesting the diagnosis. Look for the following common patterns: symmetric (affecting mirror-image portions of the body), sun exposed (affecting skin sites that routinely receive solar irradiation), intertriginous (affecting warm, moist, apposed skin sites), acral (affecting the distal extremities, ears, and nose), genital, and flexor or extensor predominance. Additional descriptors are often used to further characterize and describe a skin lesion or the relationship between various skin lesions such as confluent or clustered. Table 53.2 lists common morphologic and configurational terms.

TABLE 53.1 Summary of Key Assessment Items

Assessment Item	Purpose and Relevant Questions to Ask
Family history	Some skin diseases are familial or hereditary. When hereditary skin disease is ascertained, one may have the opportunity to both correct misconceptions and allay fears about the presence, absence, or prognosis of disease. What are the current familial dermatologic diseases?
Personal history	What was the age at onset of the problem? How has the patient adjusted to the problem? By social withdrawal? Cosmetic coverup? Withdrawal from school athletic activities that require showers (e.g., football, tennis)? Does the problem threaten the patient's self-image of masculinity or femininity? What is the patient's ethnic origin? (Some skin diseases are more common in certain ethic groups.)
Geographic origin and present abode	Length of time spent living in each area? Some skin diseases are indigenous, which may be important because of increased exposure. Occasionally, a contact of only 5 min is all that is necessary for acquisition of a disease.
Season	Seasonal occurrence of a problem? Pollen? Sunlight?
Occupation	Type of work? Skin contact material (e.g., chemicals, dust, gas), excessive heat and abnormal lighting, unhygienic surroundings, possible infective insects, other family members' occupational exposures?
Leisure activities	Does the problem occur only on weekends? After yard activities? Painting? Woodworking? Camping? Fishing? Hiking? In association with children's play?
Accompanying diseases	Collagen disease? Drug therapy for collagen disease? Other diseases and their drug therapy?
Previous treatment	Self-treatment? Other drugs prescribed?
Special history	Onset of skin lesions (abnormality)? Remissions, exacerbations, or recurrences? Site of onset? Character of lesions? Original character and subsequent changes? Course or extension? Symptoms? Itching? Ability to perform duties? Topical therapy? Self-treatment? Psychological factor? What does the patient associate with exacerbations of the problem (e.g., stress of a family argument, tax time, report time)?

Data from Rosen T et al, editors: *Nurse's atlas of dermatology,* Boston, 1983, Little, Brown.

KEY POINTS

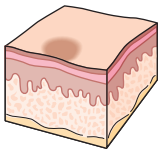
- Skin lesions may be categorized as primary or secondary.
- Primary lesions retain their original appearance, unmodified by time and external processes such as scratching.
- Secondary lesions are those whose appearance has been modified over time; they may look quite dissimilar to the original lesion. The differentiation of primary from secondary lesions aids in establishing a correct diagnosis.
- A description of lesion color, shape, number, and distribution is helpful in determining the cause of a lesion.

SELECTED SKIN DISORDERS

Diseases of the skin are divisible into two broad etiologic categories: inflammatory/infectious and proliferative/neoplastic. Inflammatory disorders of the skin often occur in individuals who have hypersensitivity reactions to substances in the environment. Infectious agents ranging from viruses to insects may infect the skin. Proliferative conditions include psoriasis, seborrheic keratosis, cysts, warts, and papillomas. Other benign tumors arise from other cells in the skin: nevi, lipomas, dermatofibromas, neuromas, and hemangiomas. Kaposi sarcoma is a malignant, opportunistic neoplasm that occurs in persons with preexisting immunodeficiency.

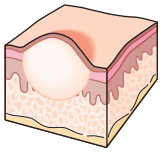
Skin cancer is the most common malignancy in the United States; however, with the exceptions of malignant melanoma and a few squamous carcinomas, skin cancers are not life threatening. Ultraviolet light damages sun-exposed skin and is a major factor in the development of skin

NONPALPABLE **PRIMARY LESIONS** (Original Appearance)

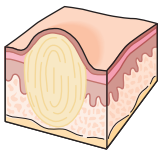


Macule: A spot, circumscribed, up to 1 cm; not palpable; not elevated above or depressed below surrounding skin surface; hypopigmented, hyperpigmented, or erythematous. **Example:** Freckles. Referred to as **patch** if greater than 1 cm. **Examples:** Café au lait spots, mongolian spots.

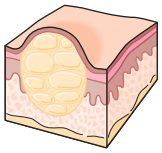
PALPABLE, SOLID



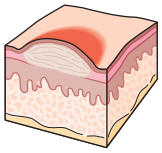
Papule: A bump, palpable and circumscribed, elevated and less than 5 mm in diameter; may be pigmented, erythematous, or flesh-toned. **Example:** Elevated nevus (mole).



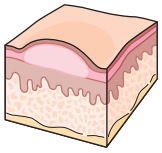
Nodule: A lesion similar to a papule, with a diameter of 5 mm to 2 cm; may have a significant palpable dermal component. **Examples:** Fibroma, xanthoma, intradermal nevi.



Tumor: Any mass lesion; generally larger than a nodule; may be either malignant or benign. **Example:** Lipoma.

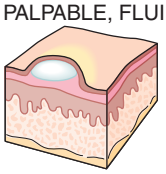


Plaque: Usually well-circumscribed lesion with large surface area and slight elevation. **Examples:** Psoriasis, lichen planus.

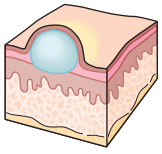


Wheal: An elevation in the skin, with a smooth surface, sloping borders, and (usually) light pink color; caused by acute areas of edema in the skin; may appear, disappear, or change form abruptly within minutes or hours; size ranges from 3 mm to 20 cm. **Example:** Mosquito bite.

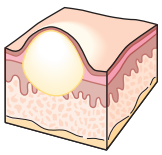
PALPABLE, FLUID-FILLED



Vesicle: A small blister (up to 5 mm in diameter); fluid collection may be subcorneal, intraepidermal, or subepidermal. **Example:** Herpes simplex (early stages).



Bulla: A blister larger than 5 mm; fluid may be located at various levels. **Examples:** Pemphigus, pemphigoid.

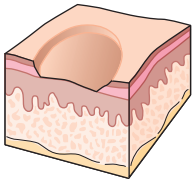


Pustule: An elevated, well-circumscribed lesion containing purulent exudate. **Example:** Acne vulgaris.

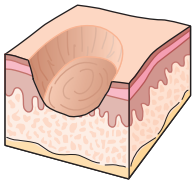
FIG 53.3 Characteristics of common skin lesions. *Continued*

SECONDARY LESIONS (Modification of Original Appearance)

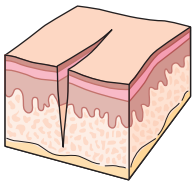
DAMAGED OR DIMINISHED SKIN SURFACE



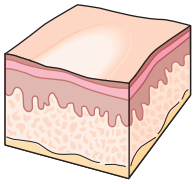
Erosion: Loss of epidermis that does not extend into dermis. **Example:** Ruptured chickenpox vesicle.



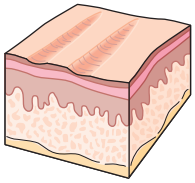
Ulcer: Loss of skin through the epidermis; healing results in scar formation. **Example:** Stasis ulcer.



Fissure: A split in all epidermal layers of skin. **Example:** Athlete's foot.

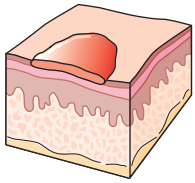


Atrophy: Diminution of epidermal surface; skin looks thinner and more translucent than normal; atrophy of the dermal layers may result in wasting or depression of the skin surface. **Example:** Arterial insufficiency.

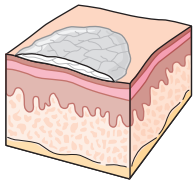


Excoriation: Loss of outer skin layers from scratching or rubbing. **Example:** Scratched insect bite.

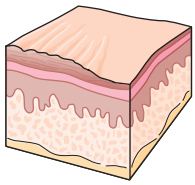
AUGMENTED OR INCREASED SKIN SURFACE



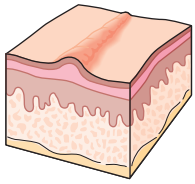
Crust: A collection of serous exudate and debris on the surface of damaged or absent outer skin layers. **Example:** Impetigo.



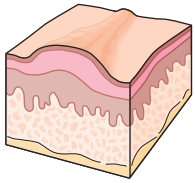
Scale: A compact portion of desquamating stratum corneum; may vary in size, thickness, and consistency. **Examples:** Psoriasis scale (compact and thick), pityriasis rosea scale (thin and small).



Lichenification: Epidermal thickening and roughening of the skin with increased visibility of skin surface furrows. **Example:** Chronic atopic dermatitis.



Scar: A collection of fibrous tissue that forms to replace lost epidermal and dermal tissue. **Examples:** Surgical scar, acne scar.



Keloid: Augmentation of scar tissue, creating a significant elevation on the skin surface after healing. **Examples:** Postsurgical scar, postacne scar.

FIG 53.3, cont'd

cancer. Although many of the disorders described in the following section are not life threatening, they can affect the quality of life.

INFECTIOUS PROCESSES

Viral Infections

Verrucae

Etiology and pathogenesis. Verrucae, or warts (Fig. 53.4), are common benign papillomas caused by DNA-containing papillomaviruses. Although warts vary in appearance depending on their location, the histologic characteristics of all lesions are similar. A wart is actually an exaggeration of normal skin composition, with the stratum corneum being irregularly thickened. The human papillomaviruses, the subgroup of papovaviruses that causes human warts, are not found in other animals and invade only the skin and mucous membranes of humans.

Warts may resolve spontaneously if immunity to the virus develops, but the immune response can be delayed for years and is not reliably activated in every case. In 95% of cases, untreated warts will resolve within 5 years, but they may multiply into hundreds of lesions and can involve any body site. Current surgical treatment may be directed at removal of the wart by laser. Liquid nitrogen or acid chemicals, cryotherapy, and salicylic acid paint or plasters have also been effective medical treatments. Topical blistering agents, immunomodulators, and intralesional injections of various agents may also be effective treatment modalities.

Herpes Simplex Virus

Etiology and pathogenesis. Herpes simplex virus (HSV) infections of the skin and mucous membranes are common (Fig. 53.5). Two types of herpesviruses infect humans: type 1 and type 2. Most HSV-1 infections occur above the waist. HSV-1 may result when external infection is spread to other parts of the body through the occupational hazards that exist in professions such as dentistry and medicine and some athletics. HSV-2 is responsible for most infections in the genital region.

Herpesvirus lesions usually begin with a burning or tingling sensation. Vesicles and erythema follow and progress to pustules, ulcers, and crusts before healing. The lesion is most common on the lips, face, and mouth. Pain is common, and healing takes place in 10 to 14 days. After the initial infection, the herpesvirus persists in latent form in the trigeminal nerve and other ganglia. Recurrent lesions are common and may be precipitated by stress, sunlight exposure, menses, or injury. The vast majority of patients have at least one episode of herpesvirus reactivation, and some individuals may have 10 or more outbreaks per year. Recently concern has arisen over the identification of infectious viral shedding in the absence of symptomatic lesions.

Treatment. No cure for herpes simplex is known, and most treatment measures are palliative. Lidocaine (Xylocaine) or diphenhydramine (Benadryl) application and aspirin administration help relieve pain. Cold compresses help in the acute stages. Acyclovir, famciclovir, or valacyclovir is recommended to shorten the duration of active disease

TABLE 53.2 Lesion Descriptors

Term	Definition
Confluent	Blending together
Diffuse	Generalized or widespread
Discrete	Remaining separate but close together
Eczematous	Vesicles with an oozing crust
Herpetiform	Closely grouped vesicles (herpeslike)
Linear	Set in a straight line
Localized	Found only in one area
Pedunculated	On a stalk
Reticulated	Netlike array
Round lesions	Annular (ring shaped, active edge, clear center)
	Arcuate (arc shaped, incomplete circle)
	Circinate (circular)
	Guttate (small droplet–like)
	Iris (concentric circles such as a bull's eye)
	Nummular (coin shaped)
	Ovoid (oval shaped)
Serpiginous	Wandering, snakelike
Telangiectatic	Characterized by dilated surface vessels
Verrucous	Rough, wartlike surface
Zosteriform	Similar to shingles, following along a nerve root dermatome

Data from Sauer GC: *Manual of skin diseases,* ed 6, Philadelphia, 1991, Lippincott.

FIG 53.4 Plantar warts. (From Callen JP et al: *Color atlas of dermatology,* ed 2, Philadelphia, 2000, Saunders.)

outbreaks; in certain situations, these drugs may be used for daily prophylaxis.

Herpes Zoster Virus

Etiology and pathogenesis. Herpes zoster (shingles) is an acute localized inflammatory disease of a dermatomal segment of the skin (Fig. 53.6). It is caused by the same herpesvirus that causes chickenpox (varicella zoster virus). It is believed to be the result of reactivation of a latent varicella zoster virus that has been present in the sensory dorsal ganglia since childhood infection. During an attack of shingles, the reactivated virus travels from the ganglia to the skin of the corresponding dermatome.

Clinical manifestations. The clinical manifestations of shingles include the eruption of vesicles with erythematous bases that are restricted to skin areas supplied by sensory neurons of a single or associated group of dorsal root ganglia. Eruptions generally follow a unilateral dermatomal distribution and most often occur on the thorax, trunk, and face. In immunosuppressed persons, the lesions may extend beyond the dermatome. New crops of vesicles erupt for 3 to 5 days along the nerve pathway. Lesions are deeper and more confluent than those of chickenpox. The vesicles dry, form crusts, and eventually fall off. Lesions usually clear in 2 to 3 weeks. Severe pain and paresthesias are common. In the elderly, herpes zoster virus is a particularly serious condition that may be long lasting. Pain reports from elderly individuals indicate an increased severity and lengthy episodes of up to 1 year. Systemic treatment with acyclovir, famciclovir, or valacyclovir should be initiated as soon as possible, preferably within the first 48 to 72 hours.

Postherpetic neuralgia is the most important complication occurring in people older than 50 years. Eye involvement can result in permanent blindness. Postherpetic neuralgia is characterized by recurring pain that lasts long after the rash and blisters of shingles disappear. Pain can persist for months or years and is more common with people over the age of 60 years. A live attenuated vaccine introduced in 2006 (Zostavax) has demonstrated a reduction in the incidence rate of herpes zoster. It is approved for patients 50 years and older.

Treatment. Management of shingles includes oral antiviral drugs; acyclovir (Zovirax) is one example. Topical agents such as Burow compresses or aqueous alcohol shake lotions may also be used. Pain medication may be indicated in severe cases. Systemic corticosteroids have also been effective in healthy persons older than 50 years with severe pain, but their use remains controversial. High doses of interferon, an antiviral glycoprotein, have been used in persons with cancer when the herpetic lesions are limited to the dermatome. Additionally, vaccination is becoming an important tool in preventing herpes zoster (e.g., Zostavax).

Fungal Infections

Superficial Fungal Infections

Three genera of fungi (dermatophytes) commonly infect human skin: *Microsporum, Trichophyton,* and *Epidermophyton.* These organisms can cause an infection termed tinea in any cutaneous area, including the hair and nails. Infections in different locations are named after the location: tinea capitis (scalp) (Fig. 53.7A), tinea barbae (beard), tinea faciei (face) (Fig. 53.7B), tinea corporis (trunk) (Fig. 53.7C), tinea manus (hand), tinea cruris (groin), and tinea pedis (foot).

Clinical manifestations. The clinical signs of superficial fungal infection vary depending on the physical location and the host's response to the invading organism. Often fungal infections are manifested as erythematous macules or plaques with peripheral scaling and some central clearing. Vesicular lesions often accompany the dry scaling on the feet. Because of the variability of signs and symptoms, superficial dermatophytosis must be considered when evaluating even a weeping, crusted area more suggestive of eczema or impetigo. Dermatophyte infection of the nails, or onychomycosis, is usually seen as a white or yellow opaque discoloration that often progresses to a thickened, crumbed, or deformed nail (Fig. 53.8).

Treatment. Topical management of localized superficial dermatophyte infections is very effective. Among the topical antifungal preparations available in cream and solution form are miconazole nitrate, clotrimazole, econazole nitrate, ciclopirox olamine, and terbinafine. A 4-week course of twice-daily applications will usually clear the symptoms. For more extensive infections involving the hair, nails, or resistant organisms, systemic therapy (e.g., griseofulvin or itraconazole and terbinafine) is required. Treatment duration ranges from 3 or 4 weeks (tinea corporis) to 12 months (onychomycosis).

Yeast Infections

The yeast *Candida albicans* is another common source of superficial infection (Fig. 53.9). It is manifested in newborns as the white lesions

FIG 53.5 Herpes simplex virus (HSV), types 1 and 2. **A,** Herpes simplex labialis—recurrent lesions. **B,** Herpetic whitlow: herpes simplex infection of the finger. **C,** Vaginal primary herpes simplex virus. **D,** Penile lesions of recurrent herpes simplex infection. **E,** Recurrent blistering on the elbow of this patient was treated as recurring impetigo before being correctly diagnosed as HSV infection. **F,** The histopathologic appearance of HSV reveals an intraepidermal bulla with bizarre multinucleated balloon cells. **G,** A positive Tzanck smear from a herpes simplex blister. (A–E, From Callen JP et al: *Color atlas of dermatology,* ed 2, Philadelphia, 2000, Saunders. F and G, From Callen JP et al: *Color atlas of dermatology,* Philadelphia, 1993, Saunders.)

FIG 53.6 Herpes zoster. **A,** Recurrent infection with the varicella zoster virus. The eruption is usually dermatomal but can become generalized. **B,** Disseminated herpes zoster. **C,** Herpes zoster ophthalmicus. (From Callen JP et al: *Color atlas of dermatology,* ed 2, Philadelphia, 2000, Saunders.)

of thrush, in infants and bedridden patients as intertrigo, and in immunoimpaired individuals as the systemic disorder mucocutaneous candidiasis. Mucocutaneous candidiasis may actually be the presenting sign in an individual with a previously undiagnosed immunodeficiency disorder.

Localized yeast infections such as oral candidiasis (thrush) may be managed with nystatin mouth rinse or clotrimazole troches (throat lozenges). The topical antifungal medications mentioned earlier may also be used in the management of localized yeast infections. Widespread or systemic infections respond well to oral ketoconazole or fluconazole (Diflucan).

Bacterial Infections

Impetigo

Etiology and clinical manifestations. Impetigo is an acute, contagious skin disease characterized by the formation of vesicles, pustules, and yellowish crusts (Fig. 53.10). The most common cause of infection of the skin, impetigo is caused by staphylococci or streptococci. Approximately 5% of the population each year sustains staphylococcus infections of a severity sufficient to require medical attention. Approximately 20% of adults are chronic carriers of the bacterium *Staphylococcus aureus,* and another 60% are intermittent carriers. The bacterium is carried in

FIG 53.7 Tinea infections. **A,** Tinea capitis, localized patch. **B,** Tinea faciei. **C,** Tinea corporis. Annular scaly plaques in superficial basal cell epithelioma. (From Callen JP et al: *Color atlas of dermatology,* Philadelphia, 1993, Saunders.)

FIG 53.8 Dermatophyte infection of the nails resulting in onycholysis. (From Callen JP et al: *Color atlas of dermatology,* Philadelphia, 1993, Saunders.)

FIG 53.9 *Candida albicans* infection of the tongue in chronic mucocutaneous candidiasis. (From Berger TG et al: *Andrews' diseases of the skin: clinical dermatology,* ed 12, Philadelphia, 2016, Saunders.)

FIG 53.10 Impetigo. (From Swartz MH: *Textbook of physical diagnosis,* ed 7, Philadelphia, 2014, Saunders.)

the nasal area and may pass onto the skin and produce disease. Staphylococcal infections are a special problem for hospitalized patients, who may become infected from the infected hospital staff.

Treatment. Treatment for impetigo includes topical application of 2% mupirocin ointment (Bactroban) or 1% retapamulin (Altabax) ointment. If a large area of skin is involved or if the person is febrile, impetigo may be managed systemically with oral dicloxacillin, cephalexin, or erythromycin.

Syphilis

Etiology and clinical manifestations. A variety of sexually transmitted diseases caused by bacteria can infect the genitalia. The most serious is syphilis, which is caused by *Treponema pallidum.* If the person remains untreated, three stages can occur. In primary syphilis, a chancre (ulcer) generally occurs as a single lesion on the genitalia; the spirochetal microorganism that causes syphilis can be seen in a scraping of the chancre. Secondary syphilis is characterized by a disseminated rash that cannot be clearly distinguished from other rashes. Both the primary and the secondary stages of syphilis are contagious.

Treatment. Studies to detect serum antibodies against syphilis (such as the Venereal Disease Research Laboratories) and examination of the pustules for the spirochete are required to achieve a diagnosis. Penicillin is highly effective in eradicating syphilis in the primary and secondary stages, but unfortunately damage caused by tertiary syphilis to the cardiovascular and central nervous systems is permanent.

Leprosy

Leprosy is a chronic infectious disease of the skin caused by the intracellular bacillus *Mycobacterium leprae.* Approximately 11 million people worldwide have leprosy. The diagnosis is made with a skin biopsy. Leprosy has a low rate of infectivity and is usually responsive to sulfone drugs such as dapsone. For chronic deformities, corrective orthopedic surgery may be required.

INFLAMMATORY CONDITIONS

Lupus Erythematosus

Lupus erythematosus (LE) is an inflammatory disease that has cutaneous manifestations. Systemic LE and chronic discoid LE are clinically dissimilar but basically related diseases. The two diseases differ with regard to characteristic skin lesions, subjective complaints, other organ involvement, LE cell test findings, response to treatment, and eventual prognosis. Discoid lupus presents with scaly red plaques with scarring that involve sun-exposed skin. Classically, systemic lupus presents with a butterfly-shaped erythema involving the cheeks and nose; discoid lesions may be seen as well. A comparison of the two conditions is found in Table 53.3. Fig. 53.11 illustrates characteristic skin lesions of both conditions.

Seborrheic Dermatitis

Clinical manifestations and treatment. Seborrheic dermatitis (Fig. 53.12) is a papulosquamous skin disease manifested by various degrees of scaling and erythema in areas of high oil gland concentration such as the scalp, eyebrows, glabellae, eyelids, nasolabial folds, pinna and posterior sulcus of the ears, sternum, axillae, umbilicus, and anogenital area. Common manifestations of this disease are cradle cap in newborns and dandruff in adolescents and adults.

Although seborrheic dermatitis is not curable, it may be controlled with topical medication. The regular use of tar, zinc, selenium sulfide, or salicylic acid shampoos often clears the symptoms and signs of seborrheic dermatitis in the scalp; mild topical corticosteroids (e.g., 1% hydrocortisone) clear lesions on the face and ears.

Psoriasis

Etiology and clinical manifestations. Psoriasis is a common chronic skin disease characterized by papules and plaques with an overlying silvery scale. The specific cause of psoriasis is unknown, but it appears

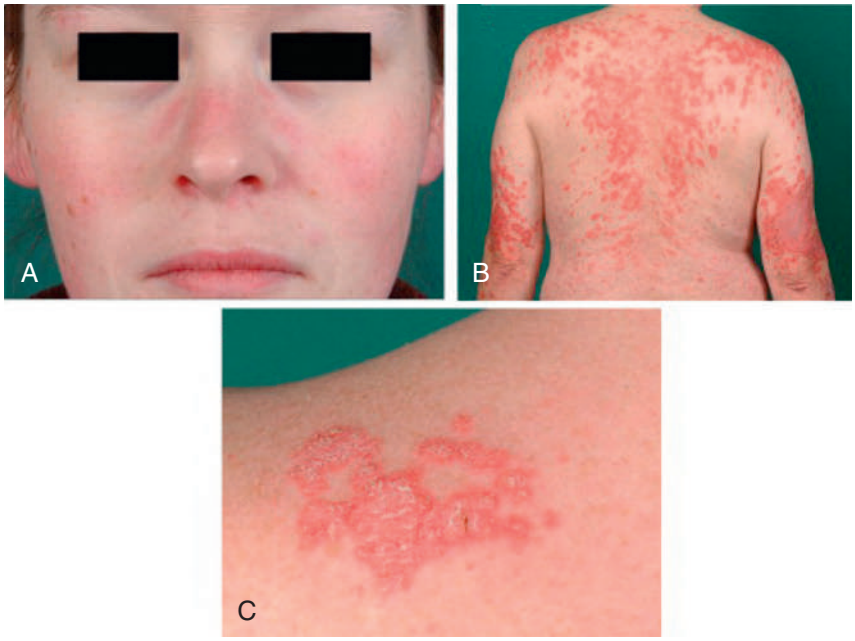


FIG 53.11 Localized acute cutaneous lupus erythematosus (ACLE) with discrete butterfly rash. **B,** Annular type of subacute cutaneous lupus erythematosus (SCLE) with multiple confluent lesions in typical sun-exposed distribution. **C,** Disseminated discoid lupus erythematosus (DLE) lesions with active hyperkeratotic infiltrated plaques on the left shoulder in a patient with systemic lupus erythematosus (SLE). (From Kuhn A, Landmann A: The classification and diagnosis of cutaneous lupus erythematosus. *J Autoimmun* 2014;48(49):14–19.)

TABLE 53.3 Comparison of Chronic Discoid With Systemic Lupus Erythematosus

Parameter	Chronic Discoid LE	Systemic LE
Primary lesions	Red, scaly, thickened, well-circumscribed patches with enlarged follicles and elevated border	Red, mildly scaly, diffuse, puffy lesions; purpura also seen
Secondary lesions	Atrophy, scarring, and pigmentary changes	No scarring; mild hyperpigmentation
Distribution	Face, mainly in the "butterfly" area, but also on the scalp, ears, arms, and chest; may not be symmetric	Face in "butterfly" area, arms, fingers, and legs; usually symmetric
Course	Very chronic with gradual progression; slow healing under therapy; no effect on life	Acute onset with fever, rash, malaise, and joint pains; most cases respond rather rapidly to steroid and supportive therapy, but prognosis for life is poor
Season	Aggravated by intense sun exposure or radiation therapy	Same
Gender incidence	Almost twice as common in females	Same
Systemic pathology	None obvious	Nephritis, arthritis, epilepsy, pancarditis, hepatitis, etc.
Laboratory findings	Biopsy characteristic in classic case LE-cell test negative, as are other laboratory tests	Biopsy less useful

From Sauer GC: *Manual of skin diseases,* ed 6, Philadelphia, 1991, Lippincott, p 253.
LE, Lupus erythematosus.

FIG 53.12 Seborrheic dermatitis: **A,** Scalp shows greasy scales not spilling onto forehead. **B,** Retroaricular erythema and scaling with crusted fissures. **C,** Nasolabial and eyebrow scaling. **D,** Petaloid variant: erythematous annular plaques in interscapular region. Eczematous dermatitis. (From Khanna N: *Illustrated synopsis of dermatology and sexually transmitted diseases,* ed 4, 2012, Elsevier India.)

to be a multifactorial inherited condition in which minor aberrations of the immune system promote inflammation and hyperproliferation within the skin. The disease may affect people of all ages, with varying degrees of severity. Lesions can appear on any area of the body; however, they seem to have a predilection for the knees, elbows, lower part of the back, scalp, and nails (Fig. 53.13). Disease progression is unpredictable, and the patient may periodically experience spontaneous exacerbations or remission.

Treatment. No cure for psoriasis is known. Treatments, both topical and systemic, are directed at clearing and controlling the lesions. Therapies include topical corticosteroids (most commonly used), a vitamin D derivative (calcipotriene ointment [Dovonex]), ultraviolet light exposure, topical tar preparations, and combinations of ultraviolet light with topical tar or systemic psoralen. Systemic therapies with methotrexate and hydroxyurea are also effective in clearing psoriasis but carry considerable risk of toxicity. Biological agents have become standard for treating moderate to severe plaque psoriasis. All patients should be screened for tuberculosis and hepatitis B virus before initiating any biological therapy. Patient response to this treatment needs to be weighed with the potential risk of significant side effects.

Lichen Planus

Etiology and pathogenesis. Lichen planus is a relatively common, chronic, pruritic disease involving inflammation and papular eruption of the skin and mucous membranes. Idiopathic lichen planus is of unknown cause but can be stimulated by a variety of drugs and chemicals in susceptible persons. The characteristic lesion is a shiny, white-topped, purplish, polygonal papule (Fig. 53.14). Lesions appear on the wrists, ankles, and trunk. Mucous membrane lesions are white and lacy and may become bullous. Pruritus is severe, and new lesions develop as a result of scratching (Koebner phenomenon). Nails are affected in approximately 10% of people with lichen planus.

Treatment. In the majority of people, lichen planus is a self-limiting disease. Treatment measures include discontinuation of all medications, followed by the administration of topical corticosteroids and occlusive dressings. Systemic corticosteroids may be indicated in severe cases, and antipruritic agents are helpful in reducing the itching.

Pityriasis Rosea

Etiology, pathogenesis, and treatment. Pityriasis rosea is a rash of unknown origin that primarily affects young adults. The incidence is highest in the spring and fall seasons. It has been speculated to be viral in origin, but to date no virus has been isolated. The characteristic lesion is a macule or papule with surrounding erythema. The lesion spreads with central clearing, much like tinea corporis. This initial lesion is a solitary lesion, called the *herald patch,* and is usually located on the trunk or neck. As the lesion enlarges and begins to fade away (2 to 10 days), successive crops of lesions appear on the trunk and neck. The extremities, face, and scalp may be involved, and mild to severe pruritus may occur. The disease is self-limiting and usually disappears within 2 to 10 weeks. Treatment is palliative and includes topical steroids, antihistamines, and colloid baths. Systemic corticosteroids may be indicated in severe cases. Systemic antibiotics, especially erythromycin, may also shorten the course.

Acne Vulgaris

Etiology and pathogenesis. Acne, an extremely common disease of the pilosebaceous unit, affects up to 90% of all individuals and produces unsightly lesions and sometimes permanent scarring and disfigurement (Fig. 53.15). Etiologically, acne involves multiple factors such as sex hormones, heredity, bacterial flora of the skin, stress, mechanical occlusion, and cosmetic use. Acne arises when sludging of sebaceous oils and deposition of loose epithelial cells cause an obstruction of the follicular canal. Continued oil production and bacterial growth in this

FIG 53.13 Psoriasis vulgaris (chronic plaque psoriasis). (From Berth-Jones J: Psoriasis. *Medicine* 2013;41(6):334–340.)

FIG 53.14 Linear lichen planus. Discrete lesions along the lines of Blaschko on the lower extremity of an adult. Lichen planus and lichenoid dermatoses. (From Shiohara T, Kano, Y: *Dermatology,* ed 3, Philadelphia, 2012, Elsevier Saunders.)

FIG 53.15 Acne vulgaris typically starts on the central forehead, nose, and chin (T-distribution). (From Habif TP et al: *Skin disease,* ed 3, Philadelphia, 2011, Saunders.)

obstructed follicle may cause rupture of the wall or sebaceous gland and result in an inflamed lesion.

Treatment. No cure for acne is known. Treatment modalities are directed at clearing the lesions and maintaining a clear complexion. Topical therapy is effective for most patients. Such medications are designed to cause increased peeling of the stratum corneum and loosening of the follicular plugs.

Many products are available to achieve this goal. Soaps, lotions, and gels containing sulfur, resorcinol, salicylic acid, or benzoyl peroxide all enhance drying and peeling. Astringents, which are liquids primarily composed of alcohol with acetone, are used as solvents to remove the surface lipid and loose skin cells, as well as to enhance drying. Topical retinoids, of which there are several, are a mainstay of treatment. Retinoic acid, a derivative of vitamin A, is one example of this class of drugs. Retinoic acid is an exfoliative agent and is very useful in dealing with open comedones and papules. Topical antibiotics are also available, the most effective being liquid preparations of erythromycin and clindamycin (Cleocin T) with an alcohol base.

For cases characterized by inflammatory lesions, pustules, or nodules, systemic therapy can be useful. Antibiotics, especially tetracycline and erythromycin, have long been used in such treatments. Concerns that continued use of systemic tetracycline-group antibiotics may result in colonization with tetracycline-resistant *S. aureus* have not been supported. This is significant because tetracycline-group antibiotics are currently one of the primary options for outpatient treatment of methicillin-resistant *S. aureus*. In cases that are resistant, minocycline, sulfamethoxazole-trimethoprim, and sulfones are occasionally used. Isotretinoin, a vitamin A derivative, is effective in the management of nodular and cystic acne. Birth control pills, especially the estrogen-dominant type, can be of value in managing severe recalcitrant acne in females. However, androgen-dominant contraceptives can aggravate or precipitate acne.

As with any medication regimen, both systemic and topical acne treatments can produce unwanted side effects in sensitive patients. Systemic tetracycline may cause gastrointestinal upset, nausea, diarrhea, and vaginal *Monilia* overgrowth. Tetracycline should not be used in children because their unerupted teeth may be severely and permanently discolored. Topical antibiotics can cause irritant or allergic contact dermatitis.

Other useful acne treatments include corticosteroid injection into cysts and nodules and surgery, which involves extraction of the comedones and drainage of fluctuant cystic lesions.

FIG 53.16 Pemphigus vulgaris. (From McCall MC, Cassoobhoy LM: *Comprehensive hospital medicine,* Philadelphia, 2007, Elsevier.)

Pemphigus

A group of related disorders (pemphigus group of vulgaris, vegetans, foliaceus, and erythematosus) is characterized by bullous eruptions (blisters). These disorders are thought to be caused by autoimmune reactions. Patients show antibodies against keratinocytes and basement membranes. The autoantibodies perhaps cause the keratinocytes to separate from one another to form blisters. Of the group of related diseases, pemphigus vulgaris has the worst prognosis (Fig. 53.16). Bullae can erupt on the skin and mucous membranes (e.g., esophagus), and toxemia and infection can cause death if proper treatment (cortisone) is not administered.

ALLERGIC SKIN RESPONSES

Atopic Dermatitis

Etiology and clinical manifestations. Atopic dermatitis is a complex genetic disease that results from gene–gene and gene–environment interactions. Genetic defects in the epidermal barrier protein filaggrin have been cited as a major cause of atopic dermatitis. It is suggested that the trait is inherited via a maternal gene located on chromosome 11. The stratum corneum layers of those individuals with loss-of-function mutations in the filaggrin gene have lower levels of natural moisturizing factor and also are deficient in extracellular lipids including ceramides. Repair of barrier function has become one of the important developments in treatment of atopic dermatitis; these treatments include emollients that contain ceramides. Atopy, or allergy, is indicated by a personal and sometimes family history of asthma, allergic rhinitis, or the most commonly seen manifestation, eczematous dermatitis (Figs. 53.17 and 53.18). The highest incidence of atopic dermatitis occurs in children, with most cases developing in those younger than 5 years. The characteristic features

FIG 53.17 Severe generalized atopic dermatitis. The dermatitis has generalized to involve the entire body. Secondary skin infection with *Staphylococcus aureus* is almost always present with this degree of inflammation. (From Habif TP et al: *Clinical dermatology*, ed 6, Philadelphia, 2016, Elsevier.)

FIG 53.18 Ill-defined patchy erythema, papules, and erosions on the bilateral forearms. This eczematous dermatitis may be atopic, irritant, or contact in etiology or a hybrid of these. (From Habif TP et al: *Skin disease,* ed 3, Philadelphia, 2011, Saunders.)

depend on the age at onset, but pruritus is always present. In infants, the disease characteristically appears on the face, scalp, or extensor surfaces of the extremities; the predominant lesion is an oozing, crusting, coalescent papule. The disease in children is most often manifested as erythema, papules, and lichenification of the flexor surfaces of the extremities, especially the antecubital and popliteal areas, the wrists, and the nape of the neck. Older children and young adults have thickening of the skin, or lichenification, along with fine, dry scaling and some papules. These changes are again seen on the flexor surfaces of the extremities and the scalp, face, and upper chest. Retrospective studies show that in nearly half of all patients with childhood atopic dermatitis, the disease improves or clears with age.

Treatment. Treatment of atopic dermatitis is usually conducted on an outpatient basis. The most important considerations are moisturization of the skin and prevention of continued drying and water loss. The drying and scaling that are characteristic features of atopic dermatitis impair the skin's ability not only to retain moisture but also to repel such external invaders as chemical irritants and surface bacteria. Milder cases of atopic eczema can be managed conservatively by decreasing the frequency of bathing, using tepid water in baths, eliminating alkaline soaps, and using moisturizing creams (especially after baths and washing). In more severe cases that involve an inflammatory response to skin breakdown, topical steroids are an important part of therapy. Short courses of systemic antibiotics such as erythromycin have also been helpful in controlling the severity of atopic eczema by reducing the concentration of cutaneous bacterial flora. Even after all these measures have been executed, some patients with severe atopic dermatitis are hospitalized for application of continuous wet dressings and topical steroids. Systemic immunomodulatory agents are indicated for adult and pediatric patients in whom optimized topical regimens using emollients, topical antiinflammatory therapies, adjunctive methods, and phototherapy do not provide adequate control of the disease.

An important feature of all atopic dermatitis cases that must be addressed is pruritus. The topical treatments mentioned previously are helpful in reducing pruritus. If additional measures are needed, systemic antihistamines (e.g., hydroxyzine and diphenhydramine) are effective.

Contact Dermatitis

Etiologies and clinical manifestations. Contact dermatitis is a cutaneous reaction to topical irritation or allergy. Irritant contact dermatitis develops in any person exposed to a sufficiently high concentration of the irritating agent. Some of the more active irritants are acids, alkalis, and hydrocarbons.

Allergic contact dermatitis indicates delayed acquired hypersensitivity to a specific allergen. Dermatologic problems may appear after years of asymptomatic exposure to the precipitating agent. Chromates, nickel, ethylenediamine, paraphenylenediamine, neomycin, formaldehyde, and lanolin components may cause allergic contact dermatitis.

Aside from reactions to various industrial chemicals, the most common type of allergic contact dermatitis reaction is to plants. Rhus dermatitis encompasses allergy to poison ivy, poison oak, and poison sumac. Clinically, rhus dermatitis begins within 48 hours of contact. The first symptom is pruritus, followed by erythema and vesicle formation, sometimes in linear fashion (Fig. 53.19). As long as the allergen remains on the surface of the skin, it can be spread to

FIG 53.19 Rhus dermatitis with the characteristic linear groups of vesicles. (From Berger TG et al: *Andrews' diseases of the skin: clinical dermatology,* ed 12, Philadelphia, 2016, Saunders.)

nonexposed areas. Therefore thorough washing can help prevent spread by hand contact. Exposure to blister fluid does not spread poison ivy lesions.

Treatment. Contact dermatitis from exposure to poison ivy can range from mild to severe. For the mildest cases, application of topical steroids or cooling shake lotions of camphor and menthol may effectively decrease discomfort. Severe cases may require hospitalization for cooling baths and wet dressings, which dry the lesions and decrease the tense, pruritic blisters. Discomfort and generalized edema often respond to systemic steroids administered over a 10- to 14-day period.

Drug Eruptions

Etiology and clinical manifestations. Adverse or undesirable reactions to medically administered drugs are common, yet cutaneous reactions are uncommon (0.1%) within the overall prescription-taking population. Cutaneous reactions to medication usually begin within a week of drug exposure, although reactions to penicillins may occur later. Women

FIG 53.20 Fixed-drug eruption. (From Marks JG, Miller JM: *Lookingbill and Marks' principals of dermatology,* ed 4, Philadelphia, 2006, Saunders.)

experience more cutaneous drug eruptions than men. The drugs that most frequently result in adverse cutaneous eruptions are ampicillin, penicillin, cephalosporins, and barbiturates. Blood transfusions also occasionally produce cutaneous reactions identical to those of a drug eruption.

The most common type of adverse cutaneous drug eruption is an erythematous maculopapular exanthem (rash). These often pruritic lesions are usually widely dispersed, and clearing is gradual and continues for several weeks after the drug has been discontinued. Other common drug reactions include urticaria (i.e., hives), erythema multiforme (including Stevens–Johnson syndrome), exfoliative dermatitis, photosensitivity, vasculitis, and fixed-drug eruption.

Exanthem-type eruptions can be caused by such medications as barbiturates, griseofulvin, penicillin, thiazides, and sulfonamides. Urticarial eruptions may result from the use of barbiturates, penicillin, chloramphenicol, phenolphthalein, salicylates, sulfonamides, or tetracycline. Erythema multiforme is seen with erythromycin, penicillin, phenolphthalein, salicylate, diphenylhydantoin, and thiazides. Exfoliative dermatitis can be caused by barbiturates, gold, penicillin, phenothiazides, and sulfonamides, and photosensitivity is seen with chlordiazepoxide, fluoroquinolones, griseofulvin, phenothiazines, sulfonamides, tetracycline, and thiazides. Cutaneous vasculitis may be triggered by iodines, erythromycin, penicillin, quinidine, sulfonamides, and thiazides. A fixed-drug eruption (Fig. 53.20) is a round-to-oval, violaceous macule or slightly palpable plaque that is often recurrent, especially in previously affected sites, on reexposure to the irritating medication. This effect can be caused by barbiturates, gold, phenolphthalein, sulfonamides, and tetracycline. These drug lists are not inclusive, and several substances are known to cause multiple adverse cutaneous reactions.

Treatment. Management of drug eruptions includes discontinuation of the offending drug and administration of oral antihistamines and antipruritic lotions of hydrocortisone, menthol, camphor, or other proven substances for relief of pruritus. For more severe eruptions, a 2- to 3-week course of systemic corticosteroids should be considered. In addition, the patient should be counseled regarding use of the offending medication and an appropriate notation placed in the patient's medical record.

Vasculitis

Etiology. When antigen and antibody react in blood vessels in the skin, severe necrotizing inflammation (vasculitis) can appear. This condition can be caused by drug allergies; disorders such as systemic lupus erythematosus (SLE), rheumatoid arthritis, and glomerulonephritis; and certain infectious diseases such as hepatitis B. Polyarteritis nodosa is a form of systemic vasculitis that can cause inflamed arteries in visceral organs, brain, and skin.

Treatment. Immunofluorescent studies reveal antigens and serum immunoglobulins trapped in the wall of the blood vessel that is inflamed by neutrophils. Acute vasculitis can cause damage not only to skin but also to the brain and visceral organs. When the vasculitis is severe, systemic corticosteroids may be administered in high doses.

PARASITIC INFESTATIONS

Scabies

Sarcoptes scabiei is a mite, and infestation with this mite in humans is called scabies. Scabies begins with eggs laid in the stratum corneum. These eggs hatch into larvae within 3 to 4 days and grow to adulthood within 2 months. Scabies is usually contracted after close personal contact with an infested individual.

Clinically, scabies lesions are small (1 to 4 mm), erythematous papules, some with an overlying dry scale or crust (Fig. 53.21). In some cases, linear burrows are seen. Scabies mites have a predilection for the finger webs, wrists, umbilicus, and groin area. The history related by most patients is an intensely pruritic eruption that spreads over a period of weeks from a single area of the body to other areas.

Scabies treatment consists of topical permethrin cream (Elimite), γ-benzene hexachloride (Lindane), or crotamiton (Eurax). For infants, 5% to 6% precipitated sulfur in petrolatum applied twice daily for 1 week is usually adequate.

Fleas

Three types of flea commonly bite and cause cutaneous reactions in humans: the human flea *(Pulex irritans),* the cat flea *(Ctenocephalides felis),* and the dog flea *(C. canis).* Flea bites may appear as small erythematous macules, erythematous papules, wheals, or a vesicle (Fig. 53.22).

Diethyltoluamide or pyrethrin insect repellents are effective in preventing flea infestation. Indoor carpeting, an ideal environment for fleas, should be treated with an appropriate insecticide.

The milder papular form of flea bites can be managed with soothing shake lotions of menthol and camphor or with topical steroids. More severe reactions (e.g., vesicles or bullae) may require a course of systemic steroids.

Lice

Phthirus pubis (crab lice), *Pediculus humanus* var. *capitis* (head lice), and *P. humanus* var. *corporis* (body lice) are the types of lice most often found on humans. They are surface dwelling, unlike the burrowing scabies mite, and they usually can be seen without magnification. Control and eradication are possible with one of the following: permethrin cream rinse or pyrethrin and piperonyl butoxide liquid, gel, or shampoo.

Chiggers

Chiggers are mites that reside in grass and bushes. They are common in the southern United States but can be found as far north as Canada. Puncture of the skin by the mite to obtain nourishment produces pruritic papules commonly seen wherever it encounters resistance, such as at the top of socks, at the belt line, or around the neckband area (Fig. 53.23). Secondary lesions are excoriations from scratching that have become infected by bacteria. Treatment is palliative, and the use of insect repellent is encouraged for prevention.

Bedbugs

The common bedbug, *Cimex lectularius,* is a reddish-brown insect 3 to 6 mm long that turns purple after feeding. Like most parasites, bedbugs feed on human blood. Importantly, they can also alternate between human and animal hosts, and they live up to and sometimes beyond 1 year. When not feeding, bedbugs stay hidden in the cracks and crevices of furniture, mattresses, wallpaper, picture frames, baseboards, flooring, door locks, or any darkened area. Unless their source is eliminated, recurrence is inevitable. Professional extermination is advised because of their many hiding places. Bedbugs have been known to feed on animal populations when forced from their living quarters. On rehabitation in the same quarters, the bedbug can easily return to human hosts.

They are nocturnal feeders, and, when crushed, they emit a foul odor. The bedbug bite is painless and produces a pruritic oval or oblong wheal with a small hemorrhagic punctum at the center. Bullous lesions are not uncommon. Usually, lesions are multiple and arranged in rows or clusters on the face, neck, hands, and arms. No area is exempt. The wheal is probably a type 1 sensitivity reaction to the anticoagulant saliva of the bedbug. Secondary excoriation and bacterial infections may occur.

The diagnosis depends on the time of the day when the lesions appear. Because of the painless bite, it is not uncommon for the victim to awake with one or several pruritic papules. Topical antipruritics are used as treatment.

FIG 53.21 Scabies in an infant. Diffuse pruritic lesions on an infant are often confused with eczema. (From White GM, Cox NH [eds.]: *Diseases of the skin: a color atlas and text,* ed 2, St Louis, 2006, Mosby.)

FIG 53.22 Insect bites (fleas) led this patient to scratch. (From Swartz MH: *Textbook of physical diagnosis,* ed 7, Philadelphia, 2014, Saunders.)

FIG 53.23 Chigger bites on ankle. (From Berger HL et al: *Andrews' diseases of the skin: clinical dermatology,* ed 12, Philadelphia, 2016, Saunders.)

Mosquitoes

Most people have experienced mosquitoes and are familiar with their bites. The typical lesion is a raised wheal on an erythematous base, accompanied by pruritus within 45 minutes of the bite. A second type of reaction is the delayed response: 8 to 12 hours after the bite, the lesion becomes raised, erythematous, and indurated, with extensive pruritus or pain. This reaction peaks 24 to 72 hours after the bite. The saliva of the mosquito is believed to be the source of the skin reaction. Although severe skin reactions are possible, they are rare. Insect repellents are recommended for prevention; local antipruritics are used for treatment.

Blood Flukes

Bathers in the freshwater lakes of Wisconsin, Michigan, and Minnesota are prone to periodic attacks of inflammatory, papular, urticarial, and vesicular eruptions on the uncovered areas of the body, mainly the legs. This pruritic eruption, commonly called *swimmer's itch,* usually subsides within a week and is caused by invasion of the skin by cercariae (larvae) of the schistosomes (worms) of ducks and mammals. The life cycle of these various species of schistosomes includes the snail as an intermediate host. On invasion of the abnormal definitive host—the human skin—the cercariae die, and the resulting skin eruption is the skin's reaction in ridding itself of the foreign bodies. Repeated attacks are met with stronger resistance, and the dermatitis becomes increasingly severe. Secondary infection, edema, and lymphangitis can occur.

Swimmer's itch is best prevented by destruction of the snails through careful addition of a combination of copper sulfate and hydrated lime to the lake water. Rapid drying of the swimmer with a towel apparently prevents penetration of the cercariae. Active therapy is directed at relieving the itching and preventing secondary infection.

Ticks

Ticks are insects that live in woods and underbrush. They attach to human and animal hosts and burrow in the epidermis, where they feed on blood. The tick bite itself is not problematic, but the infectious bacteria or viruses that ticks carry to human hosts create problems. Many tickborne illnesses are known, including Central European encephalitis, Q fever, babesiasis, relapsing fever, Rocky Mountain spotted fever (RMSF), and Lyme disease. Both RMSF and Lyme disease are relatively common in the United States.

Rocky Mountain Spotted Fever

Etiology, pathogenesis, and clinical manifestations. Rocky Mountain spotted fever (RMSF) is caused by a tick that carries *Rickettsia rickettsii.* In the past RMSF was localized to the Rocky Mountain area, but by 1982 most states had reported at least one case.

The initial tick bite appears as a papule or macule, with or without a central punctate area. The tick burrows into the host and enlarges as it feeds. The tick must be attached to the human host for 4 to 6 hours before the rickettsiae are activated by the blood. Rickettsiae are found in the tick feces and body parts. The rickettsiae then enter the bloodstream and multiply in body tissues. Within 4 to 8 days the patient experiences fever, headache, muscle aches, nausea, and vomiting. A rash then appears on the wrist or ankle. The characteristic rash is a macular or maculopapular one that spreads to the rest of the body. Other symptoms include generalized edema, conjunctivitis, petechial lesions, photophobia, lethargy, confusion, and cranial nerve deficits.

Treatment. Treatment for RMSF requires hospitalization and antibiotic therapy. The most important measure is to prevent tick bites by using insect repellents while engaged in activities in wooded areas. Once a tick has attached itself, it is important to remove all the tick's body parts to limit the possibility of infection. One can remove ticks by dousing them with mineral oil or isopropyl alcohol before slowly pulling them out with tweezers. The practice of applying a hot match to the end of the tick is not an effective method for removal because the tick may regurgitate into the open wound.

Lyme Disease

Etiology. Lyme disease is caused by the bite of a tick that carries the spirochete *Borrelia burgdorferi.* White-tailed deer and white-footed mice are the main reservoirs of this disease-causing spirochete. Lyme disease causes multiple symptoms affecting the skin, nervous system, heart, and musculoskeletal system.

Pathogenesis, clinical manifestations, and treatment. The disease has three clinical stages. Stage I usually occurs in the summer and early fall with single or multiple erythematous papules that may itch, sting, or burn. The thighs, groin, and axillae are particularly common sites of involvement. This disease is often accompanied by flulike symptoms (fatigue, headache, chills, fever, sore throat, stiff neck, nausea, myalgias, and arthralgias). If the patient remains untreated, stage II Lyme disease appears weeks to months later. This stage is characterized by meningitis, cranial nerve palsies, and peripheral neuropathy; occasionally, cardiac involvement is noted. In stage III, oligoarticular arthritis occurs. In early Lyme disease, treatment includes administration of antibiotic therapy such as doxycycline, amoxicillin, or erythromycin for 10 to 21 days. Neurologic disease, arthritis, or cardiac disease is managed with doxycycline or amoxicillin for 1 month or with intravenous penicillin for 10 to 14 days.

OTHER DISORDERS OF THE DERMIS

Scleroderma

Scleroderma is characterized by massive collagen deposition with fibrosis accompanied by inflammatory reactions and vascular changes in the capillary network. The process by which these changes occur is not

known but may represent an autoimmune mechanism or primary vasculopathy.

The two forms of scleroderma, localized and diffuse, are clinically dissimilar except for some common skin histopathologic features. Localized scleroderma (morphea) is a benign disease; diffuse scleroderma (progressive systemic sclerosis) is serious, progressive, and fatal.

Localized Scleroderma

Localized scleroderma has an unknown etiology, no systemic involvement, and no known treatment. Disability is confined to the area involved. Lesions tend to involute (shrivel) slowly and spontaneously. Relapses are rare. Primary skin lesions are single or multiple, violet-colored, firm, inelastic macules and plaques that enlarge slowly. The progressing border retains a violet hue, and the center becomes whitish and slightly depressed beneath the skin surface. Bizarre lesions occur, such as long linear bands on extremities, "saber cut" lesions in the scalp, or lesions involving one side of the face or the body. Secondary lesions include mild or severe scarring after healing, permanent hair loss from the scalp lesions, and, rarely, ulceration. The trunk, extremities, and head are most frequently involved (Fig. 53.24).

Diffuse Scleroderma

Diffuse scleroderma is a rare systemic collagen disease of unknown cause characterized by a long course of progressive disability resulting from lack of mobility of the areas and the organs affected. The skin becomes hardened like hide, the esophagus and the gastrointestinal tract semirigid, the lungs and heart fibrosed, the bones resorbed, and the overlying tissue calcified. Fig. 53.25 illustrates the "hidelike" skin on the face of a woman with diffuse scleroderma.

Another rare collagen disorder, dermatomyositis is characterized by the acute or insidious onset of muscle pain, weakness, fever, arthralgia, and, in some cases, a puffy erythematous eruption that is usually confined to the face and the eyelids. Progression of the disease results in muscle atrophy and contractures, skin telangiectasias (vascular lesions formed by blood vessel dilation) and atrophy, and generalized organ involvement. Death occurs in 50% of cases.

Sunburn and Photosensitivity

Effects of Sunlight

Sunlight is an extremely harmful environmental agent because it produces the short ultraviolet wavelength that is responsible for sunburn, thickening of the stratum corneum, suntan, and increased melanin production. Sunlight produces direct local effects on the skin in the form of elastotic syndromes, keratoacanthomas, premalignant diseases, basal cell epitheliomas, and squamous cell epitheliomas. Both indirect and direct effects can produce malignant melanomas.

Sunburn is initially manifested as erythema, pain, heat, and occasionally blistering, edema, and tenderness. In severe sunburn, these symptoms may also be accompanied by the constitutional symptoms of chills, fever, nausea, and generalized discomfort.

The most effective treatment is to avoid or limit exposure to sunlight. Wearing protective clothing is effective; sunscreens are also quite useful in preventing sunburn and the chronic solar changes of the skin. *Para*-aminobenzoic acid is the most widely used sunscreen. People sensitive to *para*-aminobenzoic acid may use cinnamates and benzophenones as substitutes. Opaque screens such as zinc oxide and titanium dioxide also work well. However, these white preparations are not cosmetically

FIG 53.24 Extensive morphea (localized scleroderma). (From Berger TG et al: *Andrews' diseases of the skin: clinical dermatology,* ed 12, Philadelphia, 2016, Saunders.)

FIG 53.25 Hidelike skin on the face of a woman with diffuse scleroderma (progressive systemic sclerosis). (From Berger TG et al: *Andrews' diseases of the skin: clinical dermatology,* ed 12, Philadelphia, 2016, Saunders.)

elegant. Recently titanium dioxide has been incorporated into foundation makeup for women.

Sunburn can be managed symptomatically with cold water baths or compresses; topical steroids are often effective in relieving the discomfort of localized severe burns. For widespread sunburn, a 10- to 14-day course of systemic steroids may suppress the symptoms.

Ulcers

An unfortunate problem for a bedridden person may be the development of pressure sores, or decubitus ulcers. Because thinning epithelial cells and blood vessels have a slower rate of repair in older adults, the incidence of decubitus ulcers is higher and the ulcer more severe in elderly individuals, and healing of damaged skin is slower.

Pressure sores are localized areas of cellular necrosis resulting from prolonged pressure between any bony prominence and an external object such as a bed or wheelchair. The tissues are deprived of blood supply and eventually die. Areas frequently affected in older persons include the heels, greater trochanter, sacrum, dorsal (especially in thin kyphotic persons) and scapular regions of the spine, and elbows. Long-term pressure increases vulnerability to decubitus ulcer development. High pressure maintained for a short time is less dangerous than low pressure continued for a long time. Predisposing factors include poor nutrition, aging, immobility, superficial sensory loss, and disturbed autonomic function (loss of bowel and bladder control). Older people with dementia are particularly prone to the development of pressure sores because of arteriosclerotic changes in the vessels, loss of subcutaneous tissue and tissue elasticity, and clouding of the sensorium.

Pressure sores can be evaluated clinically using the staging system described in Table 53.4. Pressure sores are superficial (benign) or deep (malignant). Superficial sores are reddened areas involving only the outer skin layers. They are less dangerous than deep sores and are caused by friction, shearing stresses, trauma, infection, and saturation with urine or other wet agents. The lesions are frequently painful but are easily treated and prevented. Treatment consists of keeping the area clean, dry, and free from infection or further pressure; covering the lesion with a nonstick dressing also promotes healing. Measures such as performing frequent body position changes (every 2 hours), transferring the person out of bed and into a chair, ensuring the vulnerable areas are clean and dry, and keeping the weight of the bed coverings off the feet are most effective in the prevention of superficial pressure sores.

Deep sores develop quickly as a result of thrombosis of the vessels in deep tissue overlying bony prominences. Muscle and fat layers are more vulnerable than the dermis, and involvement of these layers causes deep, large ulcers. The sore begins as a reddening of the skin with unobservable necrosis in the deep underlying tissue. In 1 to 2 days, the lesion bursts through the skin like an abscess to reveal a deep cavity full of black or infected slough, which may penetrate the bone. Skin loss from such a large area results in extensive scarring. The development of deep pressure sores with an illness can delay recovery and may even be fatal.

Prevention is more difficult with deep pressure sores, especially in the elderly. The risk of these lesions developing is greatest during the 10 days after the onset of illness or admission to the hospital, which coincides with the period of greatest immobility. A sore that develops early and penetrates deeply is most dangerous to an older person. Early signs of deterioration include apathy, loss of appetite, and incontinence. Some measures that can help prevent deep pressure sores are described in Box 53.2.

Treatment consists primarily of reinforcing preventive measures, including maintenance of fluid and protein stores that are lost through serous and purulent discharge, repair of tissues by administration of vitamin supplements, avoidance of general infections such as pneumonia or cystitis, and remediation of anemia. The lesions should be cleaned and dressed, with care taken to manage local infection. To promote granulation and healing, the wound should be irrigated with warm saline daily. Irrigation washes out the debris, reduces the growth of anaerobes, promotes separation of the slough, and decreases the pocketing of infection in deeper tissues. Infection must be eradicated and the slough must separate before healing can take place.

Altered Cell Growth: Epidermal Proliferation

Keratinocytes produce keratin. Rare, inherited defects in keratinocytes can occur, and the inherited disease congenital ichthyosis is characterized by an excessive growth of keratinocytes and keratin, which gives the skin a fish-scale appearance (Fig. 53.26).

Corns and calluses result from hyperkeratosis. Stimulation of the epidermis by intermittent pressure elicits hyperkeratosis (corn and callus formation). By contrast, *atrophy* of the epidermis can arise from a decreased blood supply.

BOX 53.2 Prevention of Deep Pressure Sores

Change position every 2 hours.
Do not oversedate or undersedate.
Avoid or correct malnutrition.
Avoid dehydration; maintain blood pressure and cardiac output.
Use an alternating-pressure airbed or waterbed.

TABLE 53.4 Clinical Description of Pressure Sores

Grade/Stage	Description
1	Acute inflammatory response primarily in epidermis with minimal soft tissue swelling and warmth; erythema of intact skin; is erythematous, and, unless abraded, erythema will blanch; blancheable erythema (reactive hyperemia) can be expected to be present for 30 to 45 min after exposure to pressure; is usually very discretely bordered; reversible with intervention
2	Pressure sore representing an inflammatory and fibroblastic response extending through epidermis into dermis; is partial-thickness or superficial skin loss involving epidermis and/or dermis; may present as blistering with erythema and/or induration; ulcer may also present as abrasion or shallow crater; wound base is moist and pink; wound is painful but free of necrotic tissue
3	Pressure sore clearly penetrating subcutaneous layers; often there is exposed muscle, fat, and tendons; full-thickness tissue loss extends through dermis to involve subcutaneous tissue; damage or necrosis of dermis may extend down to, but not through, underlying fascia; ulcer presents clinically as deep crater with or without undermining of adjacent tissue; stage may also include sinus tract formation, exudates, and/or infection; wound base is usually painful
4	Pressure sore extending beyond deep fascia, almost always to bone; deep-tissue destruction occurs, extending through subcutaneous tissue and fascia; is full-thickness skin loss with extensive tissue necrosis and damage to muscle, bone, and supporting structures (tendons and joint capsules); undermining sinus infection may be present; wound base is usually not painful

FIG 53.26 Ichthyosis. (From Newsham J, Farquharson NR, Clayton TH: *Treatment of skin disease: comprehensive therapeutic strategies,* ed 4, Philadelphia, 2014, Saunders.)

Benign or malignant neoplasms commonly arise from keratinocytes. Warts (verrucae), for instance, are caused by a virus that provokes a benign proliferation of keratinocytes. Squamous cell carcinomas (arising from keratinocytes) often occur in areas of skin excessively exposed to sunlight.

Tumors

Each cell type of the skin can give rise to either benign or malignant tumors. Benign tumors, including squamous papillomas, arise from keratinocytes, common moles (nevi) from melanocytes, lipomas from adipose cells, vascular tumors (hemangiomas) from blood vessels, dermatofibromas from fibroblasts, and neuromas from nerves.

Kaposi sarcoma arises from reticulocytes and is multifocal, metastasizing, and malignant. Kaposi sarcoma is classified as an opportunistic neoplasm because it occurs in persons with preexisting immunodeficiency, for example, in individuals with primary immunodeficiency, persons who undergo therapeutic immunosuppression, and persons with human immunodeficiency virus (HIV) infection. Figs. 53.27 and 53.28 show some of the cutaneous diseases associated with HIV infection.

Cancer

Cancer of the skin is common. Most skin cancers are slowly progressive, but certain types can be rapidly lethal. Excessive exposure to sunlight by a person with fair skin often leads to skin cancer. In addition to sunlight, exposure to irritating chemicals, recurrent trauma, and irradiation are associated with a high risk of skin cancer.

Basal cell carcinomas are the most common skin tumors and the most benign (Fig. 53.29). Squamous cell carcinomas are the second

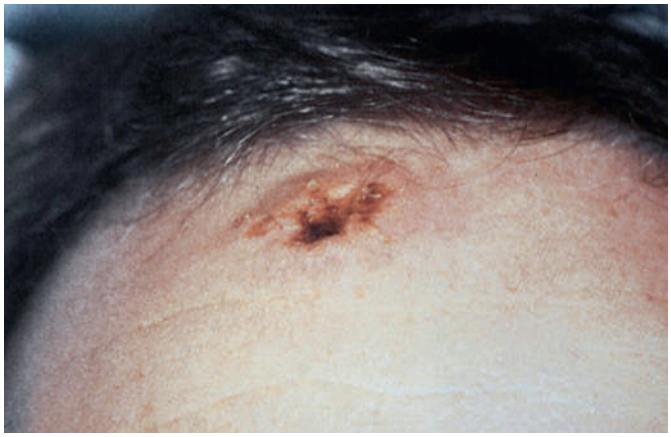


FIG 53.27 Forehead ulcer in a human immunodeficiency virus–infected host with *Cryptococcus neoformans* seen in histopathology. (From Perfect JR: *Mandell, Douglas, and Bennett's principles and practice of infectious diseases, updated edition*, Philadelphia, 2015, Saunders.)

FIG 53.28 Seborrheic dermatitis associated with HIV infection. (From Kliegman RM et al: *Nelson textbook of pediatrics,* ed 19, Philadelphia, 2011, Saunders.)

FIG 53.29 Basal cell carcinoma. Notice the rolled, well-defined margin. (From Swartz MH: *Textbook of physical diagnosis,* ed 7, Philadelphia, 2014, Saunders.)

most common skin malignancy (Fig. 53.30). They can occasionally metastasize. By contrast, melanoma is rare but can be highly malignant (Fig. 53.31). Melanoma is notoriously unpredictable; however, the prognosis is based on the size, the depth of invasion of the tumor, and the presence of metastasis. Lumps that increase rapidly in size, change color, ulcerate, or bleed should undergo biopsy and be examined microscopically to rule out malignancy. Complete surgical excision is the treatment of choice for skin cancers.

Pigmentation Alterations

Vitiligo

Vitiligo (leukoderma) is a condition in which pigment disappears from a patch of skin. The onset is sudden and may be associated with pernicious anemia, hyperthyroidism, and diabetes mellitus.

Vitiligo is a concern to darkly pigmented individuals of all races. It also affects light-skinned individuals, but not as often. The lesion is a depigmented patch with definite borders on the face, axillae, neck, or extremities (Fig. 53.32). The borders are smooth. Size varies from small to large macules involving large areas of the skin surface. The large macular type is much more common. Depigmented areas, which burn in sunlight, appear bone colored or grayish blue.

Vitiligo appears at any age, in men and women alike, and usually occurs before the age of 21. Its incidence has been increasing in India, Pakistan, and Far Eastern countries. Although the cause is unknown, inheritance and autoimmune factors have been implicated. Affected areas spread over time.

Treatment may consist of various topical immunomodulating agents, as well as the use of various types of ultraviolet light. Cosmetics such as Dermablend may be used to camouflage the areas of depigmentation.

FIG 53.30 **A** and **B,** Squamous cell carcinomas. (From White GM, Cox NH [eds.]: *Diseases of the skin, a color atlas and text,* ed 2, St Louis, 2006, Mosby.)

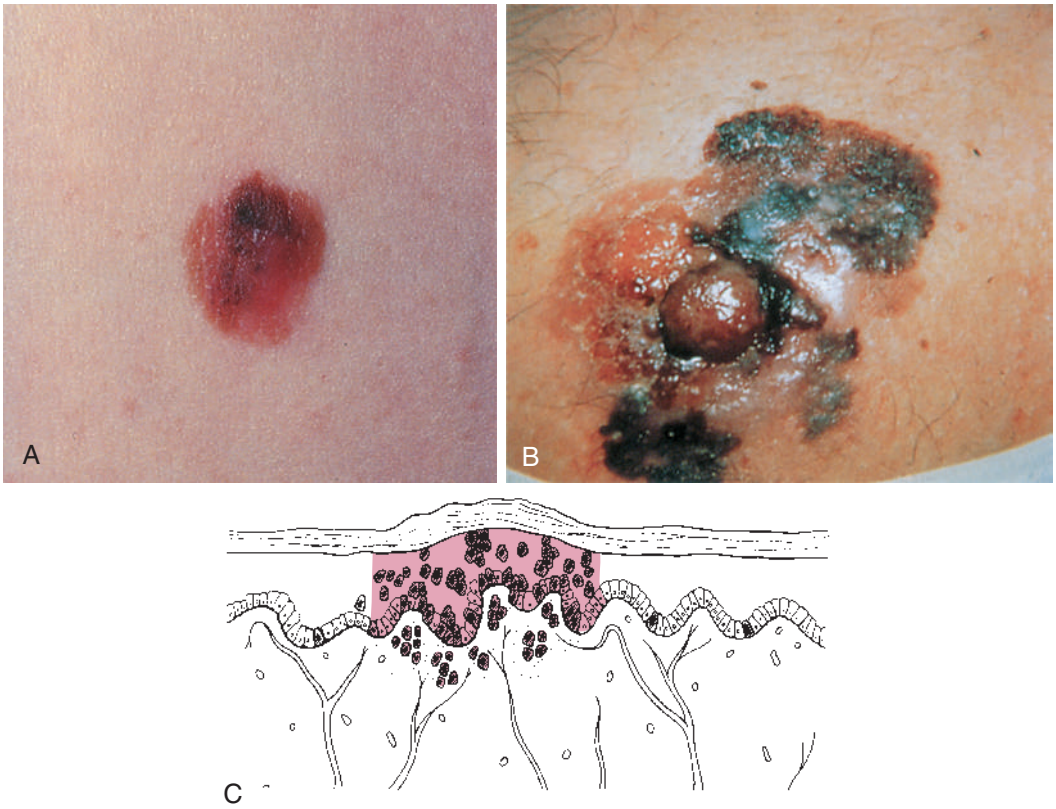


FIG 53.31 **A** and **B,** Superficial spreading malignant melanoma. **C,** Cross-section through a melanoma. Note the nests of melanoma cells in the dermis. (From Swartz MH: *Textbook of physical diagnosis,* ed 7, Philadelphia, 2014, Saunders.)

Albinism

Etiology and pathogenesis. Melanocytes produce melanin. A partial or total absence of melanin arises as an inborn error in metabolism in individuals with albinism. Albinism, also termed *oculocutaneous albinism,* is characterized by a generalized lack of pigmentation of the skin and the hair. In addition, the eyes may show nystagmus and a lack of pigmentation of the fundi and translucent irises. The condition is recessively inherited. Biochemically, albinism occurs because of impaired or absent melanin synthesis. The long-term consequences of albinism may include solar keratoses and basal and squamous cell cancers.

KEY POINTS

- Skin infections may be caused by viral, fungal, or bacterial organisms.
- Viruses are associated with warts (human papillomavirus), cold sores (herpes simplex), and shingles (herpes zoster). Warts are painless. They may be surgically removed but often resolve spontaneously. Herpes simplex lesions are painful, may be managed symptomatically, and often recur in times of stress. Herpes zoster inhabits sensory dorsal ganglia neurons and causes pain along a dermatome.
- Superficial fungal infections (tinea, ringworm) are often characterized by central clearing and peripheral scaling. They may be effectively managed with topical antifungals. Yeast infections tend to occur in moist areas such as mucous membranes and are managed with systemic or topical drugs.
- Impetigo is caused by staphylococcal or streptococcal infection and is characterized by yellowish pustules and crusts. It responds to antibiotic therapy.
- The cause of noninfectious inflammatory diseases is usually unknown. Lupus erythematosus, seborrheic dermatitis, psoriasis, lichen planus, pityriasis rosea, and acne are in this category. Treatment is aimed at reducing inflammation rather than cure. Antibiotics may be used to prevent or manage lesion superinfections.
- Skin allergies are associated with substances that cause erythema and pruritus. Atopic dermatitis (eczema), commonly seen in young children, may be aggravated by substances to which the individual is allergic. Contact dermatitis can occur in anyone exposed to a sufficiently high concentration of an irritating substance. Drug reactions are allergic responses manifested as widely dispersed, often pruritic rashes. Antigen–antibody reactions within cutaneous blood vessels can result in severe necrotizing vasculitis.
- The skin is subject to invasion by a number of different bugs, ticks, and parasites. Lesions tend to be singular or grouped and in areas exposed to the particular pest. Scabies commonly occurs on the hands and wrists and may appear as linear burrows. Bites from fleas, mites, bedbugs, and mosquitoes often induce pruritic macules or papules. Tick bites are usually painless but may be problematic because ticks can carry diseases such as Rocky Mountain spotted fever and Lyme disease.
- Scleroderma is a collagen disease of unknown cause. It may be localized to the skin or produce systemic involvement. The skin is discolored, thick, and hardened.
- Ultraviolet rays in sunlight are associated with acute damage to the skin (sunburn) and increase the long-term risk of skin cancer.
- Pressure ulcer is a significant problem of immobility caused by prolonged pressure on bony prominences. Superficial sores are reddened areas involving the outer skin layers. Deep sores are due to thrombosis of vessels deep in tissue. Deep sores may be unnoticed initially and then burst through the skin like an abscess.
- Abnormalities of skin cell growth may result in such benign processes as corns and calluses or the more serious consequence of cancer. Basal cell and squamous cell carcinomas are slowly progressive and generally amenable to surgical excision. Malignant melanoma is more prone to metastasis and carries a poorer prognosis.
- Abnormal pigmentation may occur in response to skin injury, infection, or inflammation or may be genetically determined. Albinism is due to lack of melanin production. Vitiligo is a depigmented patch of skin that is most noticeable in dark-skinned individuals. The cause of vitiligo is unknown.
- Education regarding the use of sunscreens and clothing for protection against ultraviolet light–induced damage is indicated. Sunglasses and magnifiers are beneficial for the ocular symptoms.

SPECIAL CHARACTERISTICS OF DARK SKIN

A number of disorders of the skin exclusively affect people with dark skin. Pigmentary disturbances from many causes, both hypopigmentation and hyperpigmentation, are common. Postinflammatory hyperpigmentation, for example, may occur in African American individuals when melanocytes are stimulated by inflammation. Hyperpigmentation in any person with dark skin can occur after traumatic injury, skin infection, or inflammatory skin disease. Patchy areas of depigmentation (vitiligo), described earlier, are more noticeable in persons with dark skin because of the color contrast. Some lesions, such as those causing erythema, may show no visible color change in darkly pigmented individuals. For example, petechiae, which cause pinpoint purplish-red lesions, are usually observable only on the oral mucosa or conjunctiva.

Disorders such as seborrheic dermatitis and keloids are seen with greater frequency in African Americans. The custom of tightly plaiting the hair or using hot oil and tension on the scalp leads to gradual damage to hair follicles, hair thinning, and, eventually, hair loss. Known as traumatic alopecia, this condition is also seen with greater frequency in African Americans (Fig. 53.33).

Conversely, many skin disorders that affect light-skinned people, such as squamous cell or basal cell carcinoma, senile keratoses, and psoriasis, only rarely affect darker-skinned persons.

Psoriasis is rare among the African American population. If present, it may be difficult to detect. The typical bright-red color is not present. The plaques assume a blue or violet hue because of stimulation of melanocytes. The characteristic silvery scale is often absent.

Literature related specifically to abnormalities of dark skin is also rare. Normal variants such as the Mongolian spot in infants, Futcher or Voigt lines, and linear nail pigmentation are frequently mistaken for disorders. Box 53.3 presents tips for assessing dark skin.

KEY POINTS

- Certain skin disorders are manifested differently or appear in greater or lesser frequency in individuals with dark skin. Normal variants such as the Mongolian spot in infants, Futcher or Voigt lines, and linear nail pigmentation are frequently mistaken for disorders.
- Disorders such as seborrheic dermatitis and keloids are seen more often in African Americans.
- Traumatic alopecia occurs more often in African Americans because of tight plaiting of the hair or the use of hot oil and tension on the scalp.
- Squamous cell or basal cell carcinoma, senile keratoses, and psoriasis only rarely affect darker-skinned persons.

FIG 53.32 Vitiligo. (From Ezzedine K et al: Vitiligo. *Lancet* 2015; 386(9988):74–84.)

FIG 53.33 Thinning on the temporal scalp in a patient with traction alopecia. Note how tightly the patient's hair is pulled back. Over time, if traction continues, hair loss becomes permanent. (From Bolognia JL, Jorizzo JL, Schaffer JV [eds.]: *Dermatology,* ed 3, Philadelphia, 2012, Saunders.)

BOX 53.3 Tips for Assessing Dark Skin

Skin color should be observed in the sclerae, conjunctivae, buccal mucosa, tongue, lips, nail beds, palms, and soles.

Inspection should be accompanied by palpation, especially if inflammation or edema is suspected.

Findings should always be correlated with the patient's history to arrive at a diagnosis.

Pallor in brown-skinned patients may appear as a yellowish-brown tinge to the skin. In a black-skinned patient the skin will appear ash gray. Pallor can be difficult to determine. In dark-skinned individuals it is characterized by absence of the underlying red tones in the skin.

Jaundice may be observed in the sclera but should not be confused with the normal yellow-pigmented sclera of a dark-skinned black patient. The best place to inspect is in the portion of the sclera that is observable when the eye is open. If jaundice is suspected, the posterior portion of the hard palate should also be observed for a yellowish cast, which is most effective when done in bright daylight.

The ***oral mucosa*** of dark-skinned individuals may have a normal freckling of pigmentation that may also be evident on the gums, the borders of the tongue, and the lining of the cheeks.

The gingiva may normally have a dark blue color that may appear blotchy or be evenly distributed.

Petechiae are best observed over areas of lighter pigmentation—the abdomen, gluteal areas, and volar aspect of the forearm. They may also be seen in the palpebral conjunctiva and buccal mucosa.

To differentiate petechiae and ecchymosis from erythema, remember that pressure over the area will cause erythema to blanch but will not affect either petechiae or ecchymosis.

Erythema is usually associated with increased skin temperature, so palpation should also be used if an inflammatory condition is suspected.

Edema may reduce the intensity of the color of an area of skin because of the increased distance between the external epithelium and the pigmented layers. Therefore darker skin would appear lighter. On palpation the skin may feel "tight."

Cyanosis can be difficult to determine in dark-skinned individuals. Familiarity with the precyanotic color is often helpful. However, if it is not possible to determine cyanosis from the skin, close inspection of the nail beds, lips, palpebral conjunctiva, palms, and soles should show evidence of cyanosis.

Rashes may be assessed by palpating for changes in skin texture.

INTEGUMENTARY MANIFESTATIONS OF SYSTEMIC DISEASE

The skin reflects the status of many organ systems. For example, the endocrine, cardiovascular, renal, respiratory, and hepatic systems all have possible dermal manifestations. Metabolic disorders and internal malignancies also cause cutaneous alterations. Certainly, skin manifestations of internal malignancy can be obvious. The late-appearing features of cachexia (wasting), pallor, and cutaneous metastases are obvious signs of malignancy. Abnormalities in endocrine function also produce a myriad of cutaneous changes. In general, systemic disease states are expressed through altered color, sensation, texture, and temperature of the skin; altered growth, texture, color, and lubrication of the hair; and changes in nail shape, color, and texture.

Skin

Color

Color changes in the skin can signal the presence of systemic disease. The entire color spectrum (red, orange, yellow, green, blue, indigo, and violet) is represented through possible coloration changes in the skin.

Redness (erythema) may be generalized, as with carbon monoxide poisoning, or localized, as with rashes or on the palms. Although erythema is often visible in lighter-skinned individuals, it may be less apparent in those whose skin is dark; however, the affected part may become an even deeper shade of brown. Redness may accompany inflammation.

When inflammation is suspected in a dark-skinned person, other parameters can be assessed by palpation, among them increased skin temperature, tight skin suggestive of edema, induration of deep tissue or blood vessels, and tenderness. Because the dorsal skin surface of the fingers is more sensitive to subtle skin temperature differences than the palmar surface is, the examiner should use the dorsal portion of the fingers to move from one skin area to another for comparison. The patient's family and friends are also helpful in validating color change, particularly when it has occurred gradually.

Orange discoloration can occur from the deposition of carotene. Protein-calorie malnutrition can cause hypopigmentation in African American children, with the hair and skin appearing orange.

Yellow discoloration can occur locally when lipids are deposited in skin secondary to a metabolic defect in blood lipids. More commonly, a generalized yellow (jaundiced) appearance arises because of liver disease. Bilirubin accumulates in blood and saturates the tissues. Jaundice is observed in the usual sites (e.g., mucous membranes, nail beds). Because many factors can alter these findings, one single positive finding should not be held as conclusive. Other parameters, such as environmental temperature, drug use, smoking, amount of hemoglobin, and the color of urine or stool, can support a description of cyanosis or jaundice. In both dark-skinned and light-skinned individuals, yellow sclerae may indicate jaundice, but other factors can cause yellow scleral pigmentation; fatty deposits that contain carotene are a common finding in dark-skinned individuals. To determine whether the yellow sclerae signify jaundice, observe the hard palate in bright daylight. Jaundice can be detected in this location quite early (i.e., when serum bilirubin level is 2 to 4 mg/100 mL) if the palate does not have heavy melanin pigmentation. If the hard palate does not show jaundice when the sclerae are yellow, the pigmentation may be due to some other factor, such as carotene accumulation. All these factors support the importance of repeated observation and accurate description of what is seen. As often as possible, the same individual should perform the entire examination and confirm specific findings in one area with additional data from other areas.

When jaundice is severe, biliverdin also accumulates. A person with obstructed bile ducts can become green-yellow because of biliverdin.

Blueness of the skin (cyanosis) often occurs on the tips of the fingers, toes, nose, and lips in individuals with cardiac or respiratory problems that prevent oxygenation of blood. Localized blueness with pain of the fingers on exposure to cold is termed *Raynaud disease.* It frequently arises from the presence of cryoglobulins, which solidify in the cold, and is also associated with disorders of the immune system, such as lymphoma and acquired immunodeficiency syndrome.

Indigo discoloration occurs locally, as in gangrene of the toes from severe generalized arteriosclerosis. The skin can darken from increased melanin synthesis, as in chronic adrenal insufficiency. Also, silver poisoning can make the skin dusky. Violet-colored palms (palmar erythema) can be seen in some individuals with liver disease and occasionally in pregnant women as a response to hyperestrogenism.

Shades of violet occur on the legs as a result of vascular insufficiency or when cardiopulmonary function is compromised.

The primary sites for assessing skin pallor are the nail beds, lips, and conjunctivae. When observing the lower eyelid (inferior palpebral conjunctiva) for pallor, the examiner should lower the lid sufficiently to see the conjunctiva near not only the outer canthus but also the inner canthus because the former is often darker. Greater perception is necessary when assessing a darkly pigmented individual for pallor, because the changes are subtle. Red tones may be absent; a brown-skinned person may appear more yellowish brown, and a black-skinned person may appear ash gray. This variability supports the need for accurate baseline data for comparison.

Sensation

Sensory innervation is generally responsible for the itching (pruritus) and pain that accompany most skin diseases. *Itching* is often the initial symptom in such conditions as atopic eczema, allergic contact dermatitis, scabies, dermatophytosis, psoriasis, and varicella. It can also be associated with systemic disorders, including carcinoma, diabetes, thyroid disease, uremia, and obstructive biliary disease. Other dermatologic conditions, such as herpes simplex, aphthous stomatitis, herpes zoster, furuncles, and cellulitis, produce considerable pain.

Texture

Normal aging produces an alteration in the texture of skin. Loose and wrinkled skin that lacks tone may also indicate dehydration (an abnormal finding). Dehydration may also be apparent through inspection of the oral cavity. On inspection, a dry, leathery appearance of the tongue is *not* a reliable indicator of dehydration inasmuch as mouth breathing frequently makes the tongue look dry even when the individual is well hydrated. A more reliable method of assessing hydration of the oral cavity is to palpate the mucous membranes along the area of the gum and cheek where the membranes approximate. If the membranes are dry and the finger does not slide easily, dehydration is evident.

To evaluate fluid excess, palpate the skin over the hands, feet, ankles, and sacrum. If the skin is firm and indents easily (pitting edema) on moderate pressure from the fingertips, fluid excess is present.

Feeling the deeper portions of the skin may reveal areas of induration (hardness), such as those resulting from multiple intramuscular or subcutaneous injections of medication. Lipodystrophies consist of smooth, large depressions in the skin that indicate atrophy of the subcutaneous fat layer, which has a spongy consistency. Both induration and lipodystrophy are often seen at sites of repeated insulin injections.

Temperature

If the skin feels warm and dry in a person who is febrile (feverish), the blood temperature is probably rising, an indication that the thermoregulatory mechanism of sweating may not be functioning. Likewise, if the

skin is warm and wet, the temperature can be expected to fall due to the cooling mechanism of sweating.

Sweating can also occur when the blood glucose concentration falls rapidly with a resultant rise in the blood epinephrine level. Hypoglycemic sweating can usually be distinguished from other causes of sweating because of the additional symptoms of weakness, tachycardia, hunger, headache, and "inward nervousness" manifested as mental irritability and confusion.

Because skin temperature depends on the amount of blood circulating through the dermis, decreased localized blood flow (resulting in coolness), often to the feet, may indicate a peripheral vascular dysfunction. Generalized skin coolness may indicate decreased metabolism such as that occurring after administration of a general anesthetic. If the temperature is very low, signs of shock may be evident.

On the other hand, an increase in skin temperature may indicate a hypermetabolic state, such as that occurring in hyperthyroidism and after sun exposure or sunburn.

Hair

Disturbances in body function are often reflected in changes in growth pattern, amount, texture, color, and lubrication of the hair.

Growth

The high speed of growth of the scalp hair makes it more susceptible to damage from systemic disease, toxic drugs, radiation, and stress. The rate of growth varies with general health and age, and hair growth is dependent on circulating hormonal factors (primarily testicular or adrenal androgens). Thus hormonal imbalances or shifts (e.g., those accompanying childbirth) may also result in disturbances in the hair growth cycle. Nutritional factors, although often promoted in the nonmedical literature, have little effect on hair growth except in cases of severe malnutrition.

Amount

Alterations in the amount of body hair can be extremely anxiety provoking for both males and females. In females with hypertrichosis, or hirsutism, hair growth is intensified on the upper lip, chin, cheeks, and chest; around the nipples; and from the pubic crest to the umbilicus (along the linea alba); the downy hair on the arms, legs, and back becomes coarse. The pubic hair often assumes the upright triangular distribution typical of the male as opposed to the female's usual inverted triangle. An endocrine malfunction such as excess androgen production may sometimes be associated with hirsutism, but ethnic background (Mediterranean groups predominantly) may also be responsible for the excessive hair growth. This propensity is especially true of the hair on the arms, legs, back, and face. Other ethnic group members such as full-blooded African American females and male Native Americans rarely have facial hair. Distribution of the hair in family members and ethnic background are thus important considerations in ascertaining hair growth.

Hypertrichosis lanuginosa is typically a congenital, autosomal-dominant disorder in which excessive hair is distributed over the entire body throughout life. The condition is usually associated with other congenital anomalies such as spina bifida. In some cases, such as with certain internal carcinomas, hypertrichosis lanuginosa is an acquired disorder; the degree of hairiness is variable and usually involves the face.

Color

Perhaps the most common color change in the hair is the generalized graying that accompanies the aging process.

Texture

Normal aging also produces a decrease in hair thickness. Disturbances of the thickness of scalp hair are common. Baldness (alopecia) or thinning of the hair that is generalized or creates a receding hairline is often genetically determined (Fig. 53.34). Some rare genetic defects in the hair shaft itself may produce breaking of the hairs and be erroneously diagnosed as alopecia. Generalized and localized baldness may result from treatment modalities such as radiation therapy or chemotherapy. In addition, various types of scalp diseases (e.g., fungal, lupus) and telogen effluvium (transient hair loss occurring 2 to 3 months after general anesthesia, febrile illness, or giving birth) can cause hair loss. Other traumatic types of hair loss may result from pulling of the hair because of a nervous habit, wearing hair styles such as tight braids or ponytails, or donning constrictive apparel such as a hat.

Lubrication

Hyperfunction of the sebaceous glands is associated with androgen stimulation such as occurs with the excessive scalp oiliness and facial acne in adolescence. Dry, brittle hair is commonly the result of excessive washing or the application of chemical agents (coloring, bleach, or detergent shampoos) to the hair.

In addition to direct observation of the scalp and face, correlation of the findings with data from the patient history helps determine dysfunctional states of health.

Nails

Because nails are derived from a highly active tissue, they may be affected by any serious systemic illness. Moreover, any local skin disease that affects the epidermis may also affect the nail matrix (epidermal cells that give rise to the nail plate) and lead to an abnormal (dystrophic) nail. By measuring the distance between abnormalities (pits, grooves, and lines) and the proximal nail border, one may estimate the time of initial illness.

Shape

Transverse furrows (Beau lines) in the nail indicate that nail growth has been disturbed (Fig. 53.35). These furrows can result from infection, systemic disease, or injury. Nails with a concave curve are known as spoon nails, or koilonychia. This may signal a form of iron-deficiency anemia and is also associated with other disorders such as coronary disease, syphilis, or the use of strong soaps. Destruction of the nails (onycholysis) may accompany a great variety of unrelated conditions ranging from the application of false nails to hyperthyroidism, fungal nail infection, or psoriasis (Fig. 53.36). Certain medications may also cause onycholysis. Splinter hemorrhages may be linked to bacterial endocarditis and trichinosis. These red or brown splinters or streaks run parallel to the finger in the nail bed (Fig. 53.37). Clubbing of the fingers is characterized by a flattening of the angle of the base of the nail. It may occur in association with cardiovascular disease, subacute bacterial endocarditis, and pulmonary disease.

Color

Nail color indicates the amount of blood oxygenation. Bluish or purplish discoloration of the nail beds occurs with cyanosis, whereas pallor often indicates anemia. To compare color of the nail beds, apply slight pressure on the free edge of the second or third fingernail. The blanching that results is then compared with the normal color of the nail. The rate of color return also indicates the quality of peripheral vasomotor function.

FIG 53.34 **A,** Androgenic alopecia in a male. **B,** Female pattern baldness in a woman. (From Calonje E et al: *McKee's pathology of the skin: with clinical correlations,* Philadelphia, 2012, Elsevier, pp 967–1050.)

FIG 53.35 Beau lines. This patient had major surgery 5 months previously. (From Bender AM, Cohen BA: *Pediatric dermatology,* ed 4, Philadelphia, 2013, Saunders.)

FIG 53.36 Onycholysis with distal separation of the nail plate from the nail bed. (From The Cleveland Clinic Foundation: *Current clinical medicine,* ed 2, Philadelphia, 2009, Saunders.)

Texture

Thickening of the nail may result from nutritional disturbances, repeated trauma, inflammation, and local infection. Along with thickening, toenails may become discolored and grooved, and debris may accumulate under the nail. This condition may be exacerbated as the distal portion of the nail works free from the underlying nail bed and more debris is accumulated; fungal infections may also follow. Treatment usually consists of periodic debridement of the nail plate; however, a return to normal nail structure rarely occurs after thickening.

FIG 53.37 Splinter hemorrhage in the ring finger of a man with infective endocarditis. There is an older, fading "splinter" under the nail of the index finger. Splinter hemorrhages are often smaller and darker than this. (From Epstein O et al: *Clinical examination,* ed 4, 2008, Saunders.)

KEY POINTS

- Many systemic diseases are associated with alterations in skin, hair, and nails. Skin reflects systemic inflammation and fever as erythema. A rising fever is manifested as warm, dry skin, whereas warm, moist skin indicates a fever beginning to decline. Poor oxygenation and circulation may be manifested by cyanosis, pallor, or coolness. Jaundice indicates altered bilirubin metabolism, usually caused by liver or biliary disease. Fluid balance may be manifested in the skin as decreased turgor or edema. Sympathetic activation may be indicated by cool, pale, diaphoretic skin.
- Hair growth, strength, texture, and color are affected by systemic diseases such as endocrine abnormalities, extreme malnutrition, and drugs. Excessive androgen may result in hirsutism. Alopecia may result from chemotherapeutic drugs or radiation therapy.
- Abnormalities of nail growth (pits, grooves, lines) occur as a result of nearly any serious systemic illness. Certain nail defects are characteristic of particular diseases: spoon nails may indicate iron-deficiency anemia; clubbing is associated with cardiopulmonary disease. Nail color is commonly assessed to determine the adequacy of oxygenation and perfusion.

TREATMENT IMPLICATIONS

A distinct advantage in treating the skin is the ease of direct observation of the pathologic process and the effects of treatment. Culture, macroscopic examination of skin scrapings, and biopsy also facilitate diagnosis. A correct diagnosis can help prevent complications from improper therapy but does not lessen the importance of choosing an appropriate delivery system.

Topical Treatment

Wet Dressings

Wet dressings, the application of a liquid in compress form, are a very important part of the dermatologic therapy delivery system. The applied liquid can be plain water or water with additives (e.g., sodium, magnesium, or aluminum salts).

Wet dressings are a versatile, even paradoxical, therapeutic approach in that they can dry or hydrate as necessary. Intermittently applied, they serve as an effective astringent for the weeping, oozing lesions that accompany stasis and decubitus ulcers and impetigo. Vesicular lesions, including those seen in dyshidrotic eczema, herpes zoster, and pemphigus, also respond nicely to treatment with intermittent wet dressings. By drying disease-related lesions, intermittent dressings help speed recovery.

Continuous wet dressings, on the other hand, are effective in rapidly hydrating the skin. This technique, used most often in severe cases of atopic eczema, normally requires hospitalization. Wet dressings of gauze soaked in tap water are applied directly to the skin and covered with an insulating agent such as towels, large thick gauze pads, or even long underwear to prevent evaporation. It is very important that the dressings remain moist. Therefore they must be resoaked and changed every 3 hours around the clock throughout the course of treatment. Once the desired state of hydration has been achieved, the dressings can be discontinued and emollient creams used to prevent redrying of the treated area.

Lotions

Shake lotions are mixtures of small suspended particles in a liquid vehicle such as water or alcohol. These are especially useful for application directly to moist or exudative processes such as rhus dermatitis. As the liquid phase dissipates, the evaporative effect cools and dries the skin.

Emollient lotions are a mixture of oil in water and have a slightly greasy consistency. These preparations are useful when skin moisturization is needed such as in xerotic conditions. Lotions are often used as a vehicle for other medications such as topical steroids that must be applied over large areas of skin.

Gels

Most gels are clear, colorless, volatile substances. They generally penetrate better than creams. Gels are very convenient to use on wet lesions because of their astringent tendencies. Because they do not leave the white or oily residue of creams and ointments, they are appropriate for use on scalp lesions.

Creams

Creams are the most widely used dermatologic delivery system. Many different bases are used in creams, but the "vanishing" type is most common and allows application with no surface residue. Creams penetrate well and have some moisturizing capability. They are used most frequently in the management of dry to slightly moist dermatoses.

Ointments

The medication in most ointments is carried in a petrolatum-type base, which facilitates penetration into the upper skin layers. Ointments are frequently used on skin lesions that have overlying dry scaling and crusting, but they are also very effective on severe dermatoses requiring an increased medication dosage. Ointments are semiocclusive and often not appropriate for use on lesions that are oozing and discharging a transudate or exudate.

Aerosols and Foams

Aerosols, fine particle sprays of medication usually delivered by gas under pressure, are a cosmetically elegant way of treating dermatoses, especially on hairy areas of the body. Many topical medications are now available as foams that can be used in a similar fashion.

Intralesional Injection

Intralesional injection, or the deposition of medication directly into the lesion, can be done with a conventional needle and syringe or with an instrument (Dermojet) that injects fine particles of medication

through the skin with air pressure. This delivery form is especially useful in delivering higher concentrations of corticosteroids to lesions (usually with deep dermal components) that do not respond to topical medication.

Selection of a Delivery System

Delivery system selection depends on the disease being treated, the type of lesions clinically present, and the practitioner's preferred medication routine. For instance, weeping exudative lesions require drying (wet dressings) and perhaps corticosteroids. Initial delivery as a gel would increase the drying tendency; as the lesion dries, a cream may be used to prevent overdrying and fissure formation.

In a disease state such as chronic atopic eczema with lichenoid or thickened skin, the prescriber may choose an ointment to enhance penetration of the medication into the lesion. The ointment's occlusive nature reduces moisture loss from the skin.

Seborrhea and psoriasis in the scalp may be treated with aerosols, which are quick and easy to use and are associated with a high degree of patient compliance. Patients often find them cosmetically superior to the identical medication in cream form. Keloids, which require highly concentrated medication to be delivered to a small area, are ideal candidates for intralesional injection of corticosteroids.

Corticosteroids

Corticosteroids are a very important tool in the practice of dermatology. Dermatologists administer steroids systemically and topically. Steroids may be characterized as short acting (cortisone or hydrocortisone), intermediate acting (prednisone, prednisolone, methylprednisolone, or triamcinolone), or long acting (dexamethasone or betamethasone).

Systemic Steroids

Administration of systemic steroids in dermatologic disease is usually oral. Intermediate-acting steroids (prednisone, prednisolone, methylprednisolone) are used most often. The greatest benefit of oral administration is the ability to adjust dosage schedules quickly if required. Once-daily doses, divided daily dosage, or alternate-day regimens are all effective.

Intramuscular administration of corticosteroids is also common. Preparations such as triamcinolone acetonide are used most often. These drugs, which may reduce inflammation for more than 4 weeks, ensure that an unreliable patient will receive appropriate doses of medication.

Systemic corticosteroids are generally used for relatively short periods. Therefore the complications commonly associated with corticosteroid use are not usually seen in dermatologic treatment. However, the long-term use of corticosteroids in diseases such as pemphigus and SLE often results in cushingoid features such as a round, puffy face and a "buffalo hump." Additional adverse effects include fatigue, weakness, and acne.

Topical Steroids

Corticosteroids can also be applied topically to suppress inflammation. Although this approach does not cure the disease, the reduction in erythema, edema, and pruritus promotes healing. Topical steroids are available in a variety of forms. Based on their capacity to cause cutaneous vasoconstriction, topical steroids are divided into seven groups, with group 1 being the most potent (augmented betamethasone dipropionate [Diprolene AF] and clobetasol dipropionate [Temovate]) and group 7 being the least potent (1% hydrocortisone).

KEY POINTS

- Selection of topical treatment depends largely on whether the goal is to moisturize or dry the affected area. Continuous wet dressings, lotions, creams, and ointments tend to be moisturizing. Intermittent wet dressings and gels tend to be astringents for weeping, oozing lesions.
- Corticosteroids are commonly administered to reduce inflammation. They may be given topically, intralesionally, or systemically.

DEVELOPMENTAL CONSIDERATIONS

The skin and the skin problems of special groups warrant consideration. Certain skin problems are seen only in infants and children (e.g., cradle cap and diaper rash) (Fig. 53.38). Other dermatoses are seen in both children and adults, but in children these dermatoses may appear different from the adult counterpart. Still other dermatoses affect primarily older individuals.

Infancy

Infancy connotes soft, flawless skin. In general, this is a true image. Several congenital skin lesions, such as Mongolian spots, hemangiomas, and nevi (moles), are nevertheless associated with the early neonatal period (see Pediatric Considerations: Integument in the Newborn).

Mongolian spots are caused by selective pigmentation. They usually occur on the buttocks or sacral area and are commonly seen in Asian Americans or African Americans.

Hemangiomas are vascular disorders of the skin. Two types of hemangiomas are commonly seen in infants and small children: bright-red, raised *strawberry hemangiomas* and flat, reddish-purple *port-wine stain hemangiomas.* Strawberry hemangiomas begin as small red lesions shortly after birth. They may remain as small superficial lesions or extend to involve subcutaneous tissue. Strawberry hemangiomas usually disappear before the child reaches 5 to 7 years of age without leaving an appreciable scar. Port-wine stain hemangiomas are rare, usually occur on the face and neck, and can be quite disfiguring. They do not disappear with age, and no satisfactory medical treatment is available, although laser surgery may be effective in some cases. Coverage using cosmetic makeup such as Dermablend may sufficiently conceal their disfiguring effects.

Infant skin is also exquisitely sensitive to irritation, injury, and extremes of temperature. Prolonged exposure to a warm humid environment can lead to prickly heat, and too-frequent bathing can cause excessive dryness. Soiled diapers, left unchanged, can lead to contact dermatitis and bacterial infections. Cradle cap is a harmless and usually self-limited scaly condition of the scalp. Fig. 53.38 illustrates common skin problems of infants and small children.

The primary factor in preventing infant skin disorders is careful and meticulous skin care. Baby lotions are helpful in maintaining skin moisture, whereas baby powder acts as a drying agent. Both are helpful aids when used selectively and according to the nature of the skin problem (excessive moisture or dryness).

Baby powders containing talc can cause serious respiratory problems if inhaled; therefore containers should be kept out of the reach of small children. Corn starch is preferable to talc, and baby powders containing corn starch are readily available. Unnecessary bathing should be avoided, and clothing should be comfortable and appropriate for environmental conditions.

Diaper rash results from the ammonia and alkaline by-products of urine breakdown. Disposable diapers or diapers washed in gentle

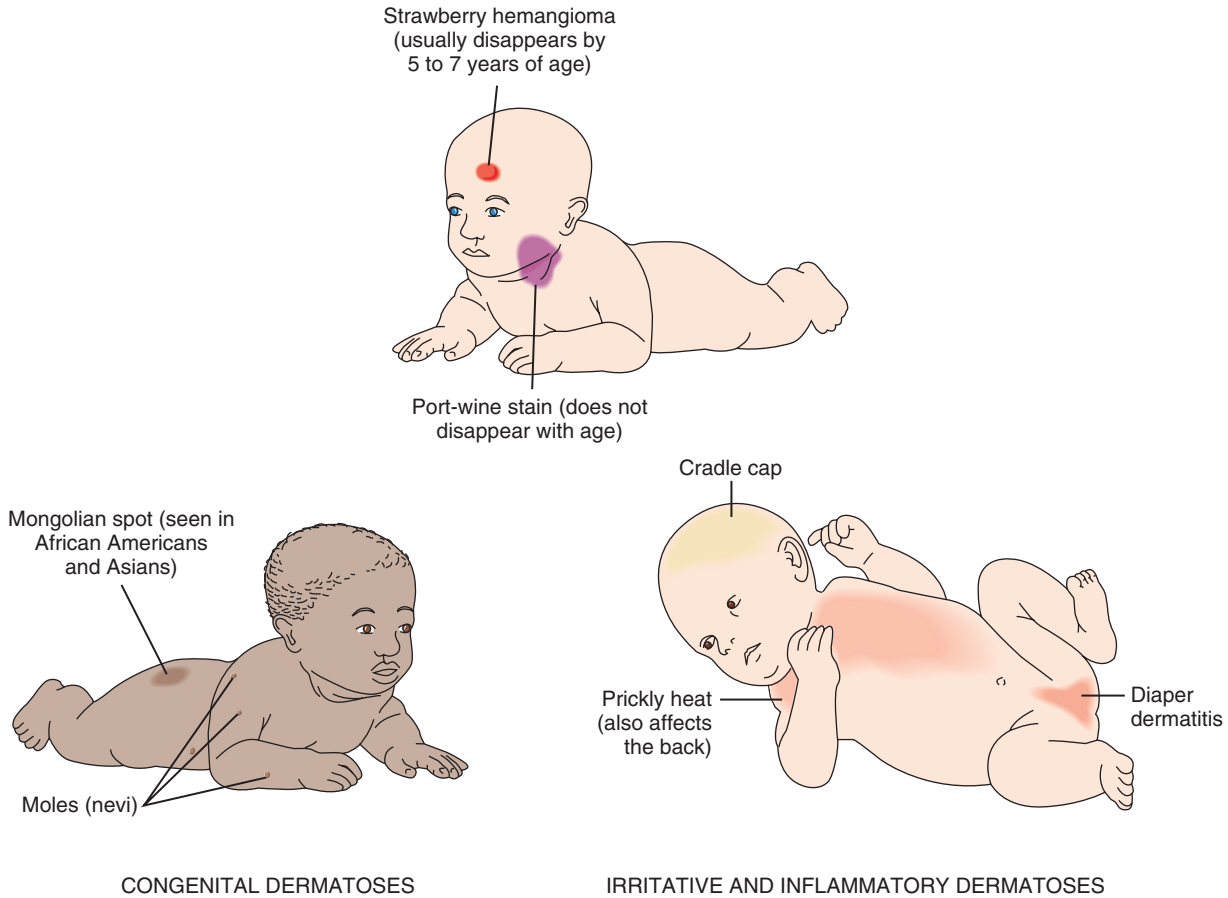


FIG 53.38 Sites of common dermatoses in infants and small children.

detergent and thoroughly rinsed to remove all traces of ammonia and alkali help prevent diaper rash. Treatment includes frequent diaper changes with careful cleansing of any irritated areas, especially in hot weather. Exposing irritated areas to air is also helpful. The use of plastic pants should be discouraged.

Prickly heat is caused by midepidermal obstruction and rupture of the sweat glands from prolonged exposure to a warm and humid environment. Treatment includes removing excessive clothing, cooling with warm water baths, drying with powders, and avoiding hot and humid environments.

Cradle cap is usually managed with mild shampooing and gentle combing to remove the scales.

Childhood Skin Disorders

As infants grow and develop into active young children, they become susceptible to the many skin disorders affecting people of all age groups who encounter environmental agents. Children, because of their physiologic development and playful nature, may also be more prone to accidents that result in major skin trauma such as lacerations or burns. (See Chapter 54 for further discussion of burn injury.) Careful activity supervision helps prevent such accidental trauma.

Besides interacting with the environment, children are frequently in close contact with other children. As a result, communicable diseases such as head lice, tinea capitis, and impetigo are more frequently seen in children (Fig. 53.39). Epidemiologically, the incidence of rubella, roseola, rubeola (measles), chickenpox, and scarlet fever is also highest in this age group.

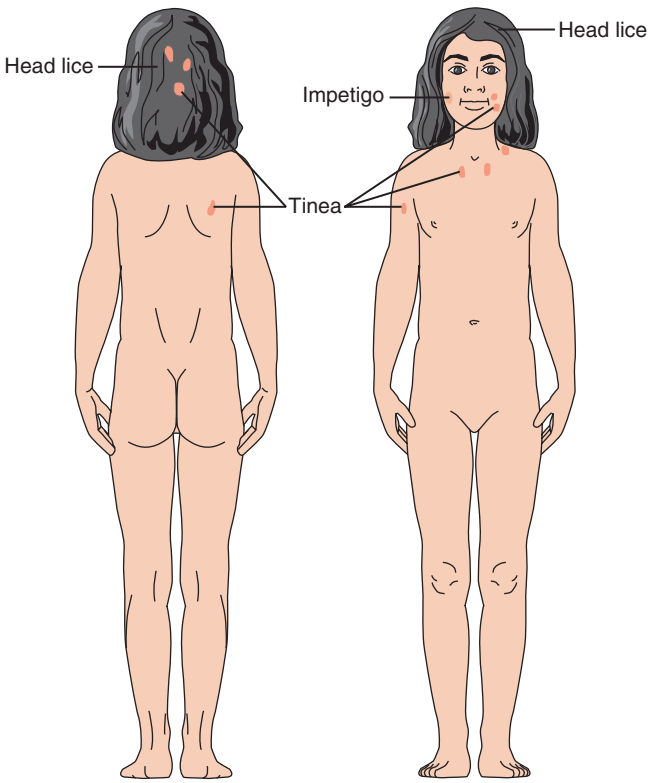
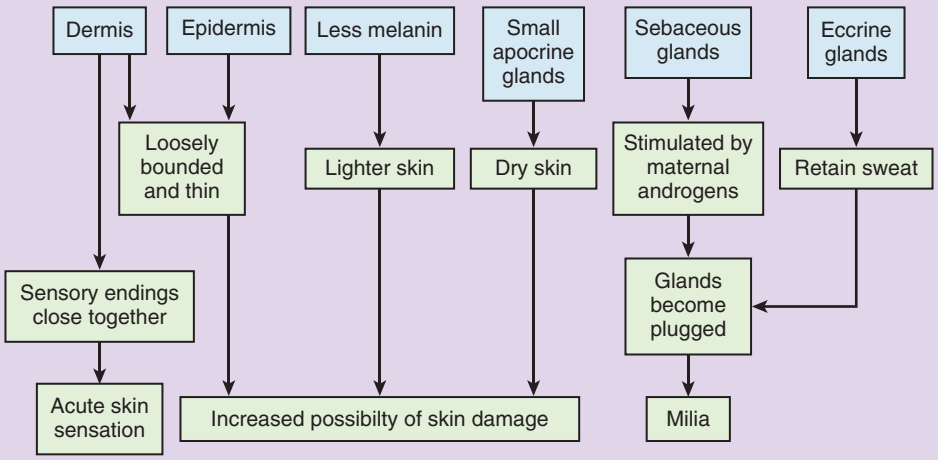


FIG 53.39 Sites of selected common communicable dermatoses affecting children.

PEDIATRIC CONSIDERATIONS

Integument in the Newborn



The integument in the newborn has sensory endings in the dermis that are closer together than seen in mature skin, creating more sensitive skin sensation. As the child develops, the skin receptors begin to become more dispersed, decreasing skin sensitivity. Several factors lead to increased possibility of skin damage. The dermis and epidermis are present and functional at birth, but the layers are thinner. The layers are loosely connected because rete pegs, which anchor the epidermis and dermis together, are not developed at birth. Therefore the dermis and epidermis are susceptible to skin damage with minimal friction. As a toddler, the dermis and epidermis become more closely linked. Another factor is that the newborn has less melanin at birth. The newborn will have lighter skin than he or she will have as a child. This lack of melanin makes the infant more susceptible to harmful rays from the sun, increasing skin damage. The newborn also has small, nonfunctional apocrine glands. These glands, located on hair follicles, do not develop until puberty. The lack of functional glands can lead to dry skin.

Several other glands in the newborn help create milia—small, white acne found on newborns. The sebaceous glands can be found on the scalp, face, and genitalia of the newborn. The sebaceous glands are activated by maternal androgens in the fetus, producing vernix at birth, and becoming plugged and producing milia after birth. The eccrine glands respond to heat and emotional stimuli to create sweat that can obstruct the sebaceous glands and create milia. The eccrine glands will be activated at a higher temperature than in adults and are most active in the palms of newborns.

Rubella

Etiology, pathogenesis, and clinical manifestations. Rubella (3-day measles, German measles) is a childhood disease caused by the rubella virus. It is characterized by a diffuse punctate, macular rash that begins on the trunk and spreads to the arms and legs. Mild febrile states occur; generally the child's temperature is less than 100°F. Postauricular, suboccipital, and cervical lymph node adenopathy is common. Coldlike symptoms usually accompany the disease in the form of cough, congestion, and coryza (profuse nasal mucous membrane discharge). Treatment is based on symptoms.

Rubella generally has no long-lasting sequelae; however, transmission of the disease to pregnant women early in the gestation period may result in severe teratogenic effects in the unborn fetus. Among the teratogenic effects are cataracts, microcephaly, mental retardation, deafness, patent ductus arteriosus, glaucoma, purpura, and bone defects.

Prevention. Most states require immunization to prevent the transmission of rubella to pregnant women. Immunization is with a live virus vaccine called *measles-mumps-rubella (MMR)*. One injection during infancy is followed by one booster dose when the child enters kindergarten or first grade or when the child enters middle school or junior high school. Administration of these two injections is considered adequate to prevent rubella. Cases of rubella in immunized children are rare.

Roseola Infantum

Pathogenesis and clinical manifestations. Roseola infantum is a contagious viral disease that generally affects children younger than 4 years and usually children about 1 year of age. It produces a characteristic maculopapular rash covering the trunk and spreading to the appendages. A rapid rise in temperature to 105°F and the appearance of coldlike symptoms accompany the disease. Unlike rubella, no cervical or postauricular lymph node adenopathy occurs. The symptoms usually subside within 3 to 5 days. Roseola infantum is frequently mistaken for rubella, which can usually be ruled out by the age of the child, as well as by the absence of lymph node adenopathy. Generally, rubella does not develop in children younger than 6 to 9 months because of the presence of maternal antibodies. Blood antibody titers may be assayed to determine the actual diagnosis. In most cases, no long-term effects from this disease are noted.

Treatment. Management of roseola infantum is palliative. As with rubella, antipyretic drugs such as acetaminophen (Tylenol) and cooling baths are used to reduce the fever. Ensuring sufficient rest and administering fluids are recommended for recuperation and body rehydration. Pruritus may rarely accompany the other symptoms. If severe, pruritus can be managed with topical lotions such as Caladryl.

Measles

Etiology, pathogenesis, and clinical manifestations. Hard measles, or 7-day measles (rubeola), is a communicable viral disease caused by the *Morbillivirus*. The characteristic rash is macular and blotchy; sometimes the macules become confluent. The rubeola rash usually begins on the face and spreads to the appendages. Accompanying symptoms are a temperature of 100°F or greater, Koplik spots (small, irregular red spots with a bluish-white speck in the center) on the buccal mucosa, and mild to severe photosensitivity. Coldlike symptoms

and general malaise and myalgia are often present. In severe cases, the macule may hemorrhage into the skin tissue or to another body surface, a condition called *hemorrhagic measles*. Measles is more severe in malnourished children. Complications include otitis media, pneumonia, and encephalitis. For a positive diagnosis, most states require antibody titer determination. Blood titers are usually determined during the disease process and 6 weeks after disappearance of the symptoms.

Prevention and treatment. Measles is preventable by vaccine (MMR), and immunization is required by law in most states. Immunization is accomplished by the MMR schedule (see discussion under Rubella).

Management of measles is based on symptoms. Children are kept in darkened rooms. Antipyretic medications are given to reduce the fever, and rest and fluids are recommended.

Chickenpox

Etiology, pathogenesis, and clinical manifestations. Chickenpox (varicella) is a common communicable childhood disease. It is caused by the varicella zoster virus, which is also the causative agent in shingles. The characteristic skin lesion occurs in three stages: macule, vesicle, and granular scab (Fig. 53.40). The macular stage is characterized by the rapid development (within hours) of macules over the trunk of the body that spread to the limbs, buccal mucosa, scalp, axillae, upper respiratory tract, and conjunctivae. During the second stage, the macules vesiculate (blister) and may become depressed or umbilicated (raised blisters with depressed centers). The vesicles burst, and a scab forms during the third stage. Crops of lesions occur successively, so all three forms of the lesion are usually visible by the third day of illness. Mild to extreme pruritus accompanies these lesions and can be a complicating factor by leading to scratching and the subsequent development of secondary bacterial infection. Other symptoms that accompany chickenpox are coldlike symptoms, including cough, coryza, and sometimes photosensitivity. Mild febrile states usually occur. Complications such as pneumonia, sepsis, and encephalitis may occur but are rare among healthy children. Disease severity is age dependent, and the risk of visceral involvement is considerably higher in adults.

Prevention and treatment. Treatment is based on symptoms. Antipyretic drugs such as acetaminophen are given for fever reduction; they may also relieve local discomfort. Pruritus is relieved with lukewarm baths. Oral administration of diphenhydramine (Benadryl) or other antihistamines may be prescribed to alleviate itching. Application of topical antipruritics such as Caladryl lotion is also helpful. However, in young children, care must be taken to avoid topical preparations of Caladryl containing diphenhydramine to circumvent possible overdose of this agent through systemic absorption. (This consideration is especially important if the young child is also taking oral Benadryl.) Home remedies such as baking soda baths also relieve pruritus, and rest and fluids are important in recuperation and rehydration. Some authorities recommend acyclovir, an antiviral agent, for the management of chickenpox.

Varicella zoster immune globulin provides passive immunity against chickenpox and is recommended after exposure, especially for high-risk groups. A vaccine against chickenpox that will provide active immunity is also available. Vaccination is currently recommended for all children and is sometimes required for school entrance.

FIG 53.40 Varicella (chickenpox). (From Lebwohl MG et al: *Treatment of skin disease,* ed 4, Philadelphia, 2014, Saunders.)

Scarlet Fever

Etiology, clinical manifestations, and treatment. Scarlet fever is a systemic reaction to the toxins produced by group A β-hemolytic streptococci. It occurs when the individual is sensitized to the toxin-producing variety of streptococci. Scarlet fever frequently occurs in association with streptococcal sore throat (strep throat), but it may also be associated with a wound, skin infection, or puerperal infection. Scarlet fever is characterized by a pink punctate skin rash on the neck, chest, axillae, groin, and thighs. When palpated, the rash feels like fine sandpaper. Flushing of the face with circumoral pallor is evident. Other symptoms include high fever, nausea and vomiting, strawberry- or raspberry-colored tongue, and skin desquamation. Complications of scarlet fever include otitis media, peritonsillar abscess, rheumatic fever, and acute glomerulonephritis. Penicillin is the treatment of choice.

Adolescence and Young Adulthood

The most common skin disorder of adolescence and young adulthood is acne vulgaris. The increased production of sex hormones and oils contributes to the development of acne. Childhood diseases are less common in adolescence; however, chronic skin diseases may be exacerbated.

Geriatric Considerations

Skin disorders are so common in elderly individuals that it is difficult to distinguish normal from abnormal. More than 90% of elderly individuals have some kind of skin disorder (Fig. 53.41). The most common skin disorders in the elderly are keratoses and skin cancers, followed by fungal infections, dermatitis, pigmentary disturbances, psoriasis, and urticaria (hives). Other skin disorders frequently seen in the elderly are comedones (blackheads), asteatoses (scaling), cherry angiomas (small, red, benign tumors), nevi (moles), skin tags (pedunculated fleshy growths), and lentigines ("liver spots"). In addition, the incidence of senile purpura and senile warts (papillomas) significantly increases, especially among the very old. Senile purpura is related to loss of the subcutaneous tissue that supports the skin capillaries. Minor trauma can cause small bruises or ecchymotic lesions, which largely occur on the extensor surface of the forearms. Approximately 40% of older men and 77% of older women show evidence of senile purpura. Senile papillomas are small yellow, brown, or black warts located on the trunk, limbs, and face; 63% of all older individuals have some senile papillomas.

Fig. 53.42 illustrates several of the common skin lesions associated with aging. Most of these lesions are considered normal concomitants of aging and cause little discomfort. The greatest concern regarding body image is the appearance of the skin, which tends to look mottled and spotty. Disorders of the skin that tend to cause the most physical discomfort are pruritus, keratoses, epitheliomas, malignant melanomas, herpes zoster, psoriasis, and pressure sores.

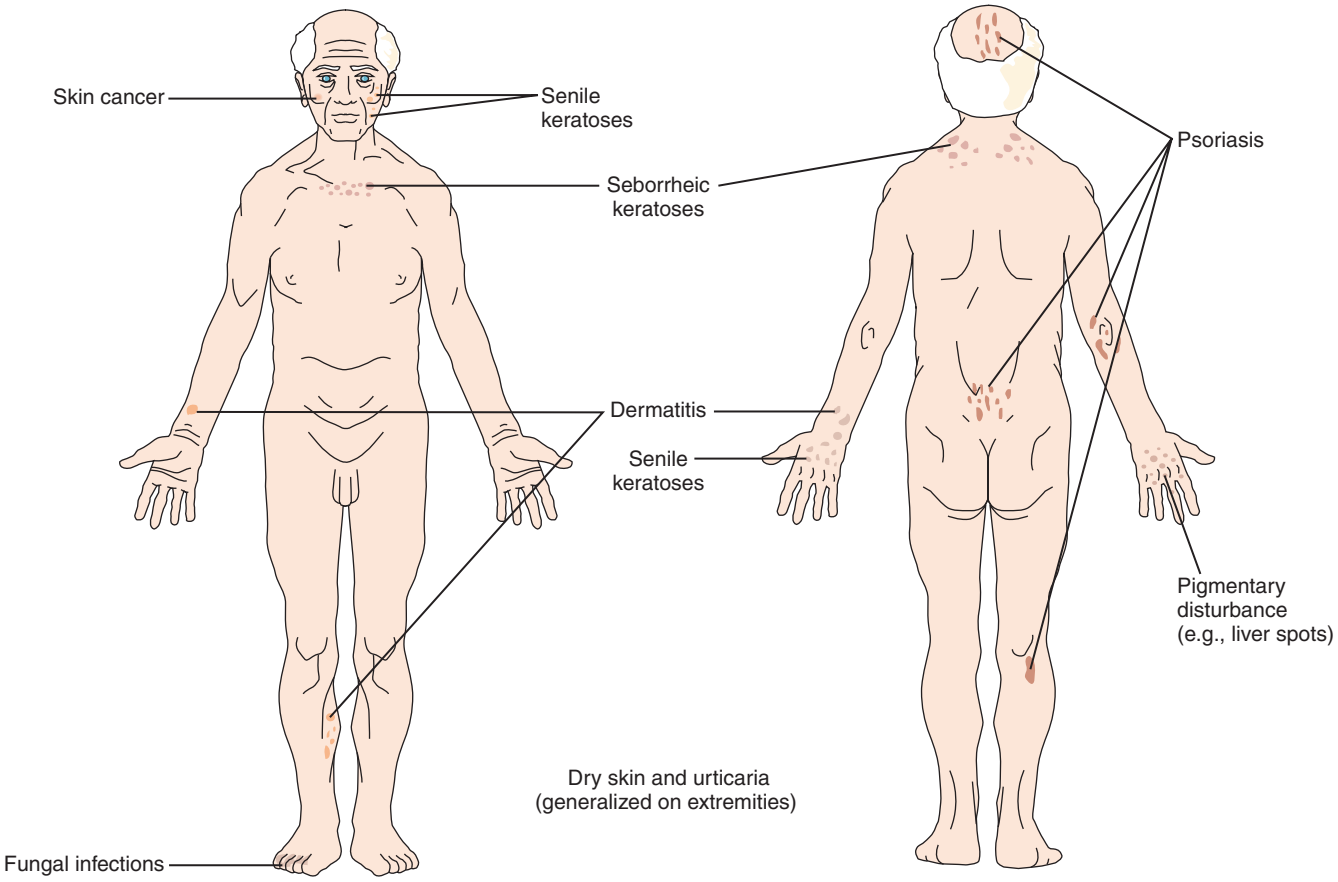


FIG 53.41 Sites of common dermatoses of the elderly.

FIG 53.42 Common skin lesions associated with aging. **A,** Cherry angioma. **B,** Acrochordons (skin tags). **C,** Senile lentigines (liver spots). (A, From Ferri FF, Studdiford JS, Tully A: *Ferri's fast facts in dermatology,* Philadelphia, 2011, Saunders. B, From Brinster NK et al: *Dermatology: high-yield pathology,* Philadelphia, 2011, Saunders. C, From Klatt EC: *Robbins and Cotran atlas of pathology,* ed 3, Philadelphia, 2015, Saunders.)

KEY POINTS

- Certain skin disorders are more common in particular age groups.
- Infants are prone to irritating lesions, including prickly heat, contact dermatitis, and cradle cap. Altered areas of pigmentation are first noticed in infancy, including Mongolian spots, hemangiomas, and nevi.
- Children are prone to skin injuries and communicable diseases. A number of viral infections, including rubella, roseola, measles, and chickenpox, are associated with characteristic skin rashes. Fever and malaise are usually present. Treatment is symptomatic. Vaccinations are available to prevent rubella, measles, and chickenpox. Scarlet fever is due to a bacterial infection and is managed with antibiotics. Children are often exposed to superficial infections and infestations, including head lice, ringworm, scabies, and impetigo.
- Acne is the most common skin disorder of adolescents.
- Elderly skin is prone to a number of problems, including psoriasis, angiomas, and skin tags. Cancerous and precancerous lesions are common and require careful screening examination.

SUMMARY

In systemic diseases, the color, texture, and composition of the skin mirror and participate in widespread pathophysiologic events. For example, internal disease states such as acquired immunodeficiency syndrome; collagen diseases such as scleroderma and dermatomyositis; diabetes; gout; malignancies; neurologic diseases; liver disease; muscle weakness; and vascular, inflammatory, and metabolic disorders all exhibit cutaneous manifestations. Because the skin mirrors the interior condition of the body, it is important in the diagnosis of disease. Cutaneous manifestations may be caused by bodily changes such as pregnancy or obesity. They may also be caused by external factors such as climate, industrial contamination, indoor heating systems, clothing, plant life, and toxic or allergic reactions to drugs and cosmetics.

A distinct advantage in treating individuals with skin disease is the ability to observe the pathology and the effects of treatment. In addition to a careful history, a culture, skin scraping, or biopsy provides good diagnostic information. A correct diagnosis can help determine the best approach to prevention and treatment.

RESOURCES

Ackerman AB, Cockerell CJ: Cutaneous lesions: correlations from microscopic to gross morphologic features. *Cutis* 37(2):137–138, 1986.

Arndt KA, et al: *Cutaneous medicine and surgery: an integrated program in dermatology*, (vol 1). Philadelphia, 1996, Saunders.

Arvin AM: Varicella-zoster virus. *Clin Microbiol Rev* 9(3):361–381, 1996.

Berardesca E, Maibach HI: Sensitive and ethnic skin: a need for special skin-care agents? *Dermatol Clin* 9(1):89–92, 1991.

Cather J, Crowley J: Use of biologic agents in combination with other therapies for the treatment of psoriasis. *Am J Clin Dermatol* 15(6):467–478, 2014.

Cerimele D, Celleno L, Serri F: Physiological changes in ageing skin. *Br J Dermatol* 122(Suppl 35):S13–S20, 1990.

Dunkle LM, Arvin AM, Whitley RJ, et al: A controlled trial of acyclovir for chickenpox in normal children. *N Engl J Med* 325(22):1539–1544, 1991.

Epstein JH: Phototherapy and photochemotherapy. *N Engl J Med* 322(16): 1149–1151, 1990.

Fanelli M, et al: Antibiotics, acne, and *Staphylococcus aureus* colonization. *Arch Dermatol* 147(8):917–921, 2011.

Gilchrest BA, Yaar M: Ageing and photoageing of the skin: observations at the cellular and molecular level. *Br J Dermatol* 127(Suppl 41):25–30, 1992.

Goldman L, Shafer A: *Goldman-Cecil medicine*, ed 25, Philadelphia, 2015, Saunders.

Gupta AK, Jain HC, Lynde CW, et al: Prevalence and epidemiology of onychomycosis in patients visiting physicians' offices: a multicenter Canadian survey of 15,000 patients. *J Am Acad Dermatol* 43(2 Pt 1):244–248, 2000.

Heyman WR: The herpes zoster vaccine. *J Am Acad Dermatol* 58(5):872–873, 2008.

James WD, et al, editors: *Andrew's diseases of the skin: clinical dermatology*, ed 12, Philadelphia, 2016.

Leyden JJ, Rawlings AV: *Skin moisturization*, New York, 2002, Marcel Dekker.

Martin S: Variants of normal skin in blacks. In Rosen T, Martin S, editors: *Atlas of black dermatology*, Boston, 1981, Little, Brown, pp 1–16.

Pieper B: Mechanical forces, pressure, shear, and friction. In Bryant RA, Nix DP, editors: *Acute and chronic wounds*, ed 3, St Louis, 2007, Mosby, pp 205–234.

Sherman CD, et al: Malignant melanomas. In Rubin P, editor: *Clinical oncology: a multidisciplinary approach for physicians and students*, ed 7, Philadelphia, 1993, Saunders, pp 667–675.

Sidbury R, Davis DM, Cohen DE, et al: Guidelines of care for the management of atopic dermatitis. *J Am Acad Dermatol* 71(2):327–349, 2014.

Smith L: Histopathologic characteristics and ultrastructure of aging skin. *Cutis* 43(5):414–424, 1989.

Spanos NP, Williams V, Gwynn MI: Effects of hypnotic, placebo, and salicylic acid treatments on wart regression. *Psychosom Med* 52(1): 109–114, 1990.

Strauss JS: Biology of the sebaceous gland and the pathophysiology of acne vulgaris. In Soter NA, Baden HP, editors: *Pathophysiology of dermatologic diseases*, New York, 1991, McGraw-Hill, pp 195–210.

Volker R: FDA expands age range for shingles vaccine. *JAMA* 305(15):1526, 2011.

Yeung-Yue KA, Brentjens MH, Lee PC, et al: Herpes simplex viruses 1 and 2. *Dermatol Clin* 20(2):249–266, 2002.

54

Burn Injuries

Sarah Ogle and Nirav Patel

http://evolve.elsevier.com/Banasik/pathophysiology/

- Review Questions and Answers
- Glossary (with audio pronunciations for selected terms)
- Animations
- Case Studies
- Key Points Review

KEY QUESTIONS

- What are the most common causes of burn injuries?
- How are burn degree and severity determined?
- What are the principles that guide the management of burn injuries?
- What are the potential complications associated with burn injuries?
- What are the outcomes after burn injuries?

CHAPTER OUTLINE

Thermal Injury, 1094
Etiology, Incidence, and Mortality, 1094
Risk Factors, 1094
Integument Effects, 1094
Depth Classification, 1096
Extent of Injury, 1097
Severity Classification, 1097
Acute Management, 1097
Assessment, 1099
Burn Shock and Acute Resuscitation, 1099
Organ Dysfunction, 1101
Cardiovascular Dysfunction, 1101
Respiratory Dysfunction, 1101
Renal Dysfunction, 1102
Metabolic Changes, 1102
Cellular Changes, 1102
Immune Response, 1103
Elements of Burn Injury Survival, 1103
Management of Wounds, 1103
Burn Surgery, 1103
Excision and Grafting, 1104
Skin Substitutes, 1105
Nutritional Support, 1105
Rehabilitation Phase, 1106
Wound Healing, 1106
Electrical Injury, 1107
Incidence and Mortality, 1107
Pathophysiology, 1107
Management and Complications, 1108
Chemical Injury, 1109
Management and Complications, 1109
Common Agents and Treatment, 1109
Hydrofluoric Acid, 1109
Anhydrous Ammonia, 1109
Cement Burns, 1110
Chemicals Associated With Automobile Airbag Burns, 1110
Tar and Asphalt, 1110
Special Populations
Introduction, 1110
Geriatric, 1110
Pediatric, 1110
Obese, 1110

Burns are injuries to tissues caused by contact with dry heat (flame or hot surfaces), moist heat (steam or hot liquids), electricity (current or lightning), chemicals (corrosive substances), friction, or radiant and electromagnetic energy. Approximately 486,000 individuals seek medical attention for burn injuries annually in the United States, with 8% necessitating acute hospitalization. In recent decades, associated mortality rates have decreased significantly, with most patients achieving excellent functional and cosmetic outcomes. Improved outcomes have been attributed to the delineation of pathophysiologic mechanism of burns, advances in management, and the development of a comprehensive, treatment-oriented approach to care. The American Burn Association in conjunction with the American College of Surgeons Committee on Trauma has been instrumental in promoting the development of burn centers and developing guidelines for trauma centers that care for burn patients. The advanced burn life support prehospital and provider courses have also been developed to provide evidence-based guidelines for the assessment and treatment of burn victims. To facilitate appropriate treatment decisions and recognize potential complications, an

understanding of the pathophysiologic processes associated with the burn injury is vital.

THERMAL INJURY

Etiology, Incidence, and Mortality

Thermal injuries are burns caused by contact with or exposure to extremes of temperature. Over the past 6 decades the incidence of burn injuries in the United States has steadily decreased, with approximately 486,000 currently occurring annually, resulting in approximately 40,000 hospitalizations and 3240 deaths. The fire loss record in the United States is one of the worst in the industrialized world, with the fire-related mortality double that of most countries on a per capita basis. In 2013 residential fires resulted in nearly $11.5 billion per year in property damage.

A half-century ago, burns over 50% of the total body surface area (TBSA) resulted in a greater than 50% mortality in pediatric patients. Currently, most children survive burns of this size, and more than half survive burns of more than 90% of their TBSA. Early mortality generally occurred as a result of inadequate initial resuscitation. As vigorous resuscitation protocols were developed, a significant reduction in mortality occurred. Additionally, many advances have contributed to decreasing thermal burn–related mortality, including a better understanding of the pathophysiologic mechanisms of burns, implementation of a multidisciplinary team–oriented approach, improvement of resuscitation strategies and infection control measures, use of early surgical excision and skin-grafting techniques, application of advances in skin substitute development, and employment of improved rehabilitation strategies.

The reduction in mortality, however, has been associated with an increased incidence of burn wound–related sepsis. The development of improved topical and systemic antimicrobial agents, recognition of the importance of maintaining proper nutrition, adoption of early wound excision, and grafting techniques have resulted in a significant reduction in wound sepsis–related complications and associated mortality. Mortality now occurs primarily in the subacute setting, primarily as a result of respiratory system–associated infectious complications. The vast majority of patients sustaining thermal burns have involvement of less than 20% of their TBSA and are treated on an outpatient basis. Despite burn degree, the associated physical and emotional sequelae are often extensive and prolonged. As a result, this unique population requires vigilant follow-up care.

Risk Factors

The U.S. Centers for Disease Control and Prevention has identified the following groups as being at high risk for fire-related injuries and deaths: children younger than 4 years, adults 65 years and older, Native Americans and African Americans, economically challenged individuals, people living in rural areas, and those residing in manufactured homes or substandard housing. Children younger than 16 years old account for one-third of all admissions to burn units and one-third of all deaths from burns and burn-related injuries. Scald injuries, primarily secondary to accidental spills, or in up to 20% from an element of abuse or neglect, account for up to 33% of all burn injuries. See Fig. 54.1 illustrating an abuse burn.

The population older than 65 years of age continues to account for an increasing proportion of all burn/fire-related deaths in the United States (~48%). Burns in the elderly carry a high mortality rate as a result of preinjury disability, age-related immunosuppression, and impaired healing responses (Table 54.1). Environmental and lifestyle factors influence the frequency and magnitude of thermal burn injuries. Alcohol and drug abuse contributes to approximately 14% of all residential fire-related deaths. Neurologic and psychiatric disorders have also been found to increase the risk of accidental burn injury, with one study finding 3% of patients admitted to the burn unit had thermal injuries associated with neurologic disorders.

FIG 54.1 All burns in children must be carefully evaluated with a consideration for nonaccidental or intentional injury. This 2-year-old boy was immersed by his father in a bathtub of hot water and sustained burns over 55% of his total body surface area. One feature that characterizes abuse burns is a clear demarcation between burned and unburned skin and the absence of drip, spill, or splatter marks. (Courtesy Michael Peck, MD, University of North Carolina Burn Center, Chapel Hill.)

Integument Effects

The skin is the largest organ of the body and constitutes approximately 20% of the total body weight. It consists of two layers, the epidermis and dermis, which rest on the hypodermis (or subcutaneous layer). The epidermis contains two main cell types, melanocytes and keratinocytes, as well as multiple appendages such as hair, nails, and glands (sweat and sebaceous). The appendages, although originating from the epidermal layer, are anatomically located in both the epidermis and the dermis. Keratinocytes synthesize keratin, and melanocytes scattered throughout the basal layer (stratum germinativum) produce melanin, a pigment that shields deeper structures of the skin from sunlight. Apocrine glands, the large sweat glands, are rudimentary structures with no known useful purpose. They respond to autonomic nerve stimulation rather than thermal stimulation to produce an odorless, viscous, milklike droplet from the hair shaft. Apocrine glands are more numerous in women and are located in the axilla, areola of nipples, groin, perineum, and perianal and periumbilical regions. Eccrine glands are small sweat glands distributed over the body that act as true secretory glands and produce the sweat responsible for heat regulation. At environmental temperatures greater than 31° or 32°C (90°F), sweating occurs over the entire body; at lower temperatures, microscopically visible droplets are secreted periodically as part of the total insensible water loss from the body. Sweat normally provides skin with an acid mantle (average pH, 5.7 to 6.4) that retards growth of the many bacteria that reside in the keratin layer, glands, and hair follicles.

Sebaceous glands secrete sebum, a complex mixture of lipids that is emptied into the hair shaft. The rate of production of sebum and its location depend on androgens, which initiate and continue production. During the hypermetabolic state that follows major thermal injury, production of sebum is decreased, leading to the dry skin conditions commonly found after recovery.

Thermal injury to the integument occurs in two phases: (1) immediate, which is a result of direct cellular injury; and (2) delayed, which occurs as a result of progressive dermal ischemia.

TABLE 54.1 Physiologic Changes Associated With Age

Body System	Pediatric	Elderly
Cardiovascular	**Symptoms of shock:**	Increased chronic disease processes
	Increased heart rate	Decreased vascular elasticity results in increased systolic blood pressure
	Decreased blood pressure	
	Decreased urine output	Decreased cardiac output
	Cardiac output dependent on heart rate	Decreased β-adrenergic responsiveness
	Decreased myocardial compliance	Decreased cardiac stress response
	Stroke volume plateaus at lower filling pressures	Decreased intrinsic heart rate
	Peripheral cyanosis:	Decreased blood flow
	Neonates: a normal finding	Decreased vascular permeability
	Children: decreased cardiac output	Increased myocardial irritability
		Decreased myocardial perfusion
		Increased dysrhythmias
		Conduction system changes
Pulmonary	Small trachea is easily obstructed	Increased need for ventilatory support
	Neck hyperextension leads to epiglottal or tracheal obstruction	Increased incidence of inhalation injury
	At <8 yr, cricoid cartilage is narrowest airway point	Increased pneumonia
	At <8 yr, no cuff is needed on endotracheal tube	Decreased lung elasticity
	Hypoxemia leads to decreased heart rate in neonates	Decreased chest wall muscle
	Hypoxemia leads to decreased heart rate in children	Decreased oxygen saturation
	Diaphragmatic breathing	Decreased tidal volume
	Lower airways are easily obstructed	Decreased vital capacity
	Decreased O_2 reserve	Decreased pulmonary capillary circulation
Thermoregulation and metabolism	Increased resting metabolic rate	Increased burn wound infection
	Increased resting O_2 consumption	Increased sepsis
	Increased BSA in relation to body weight	Poor or delayed wound healing
	Increased hypothermia	Increased preexisting malnutrition
	Increased heat loss from evaporation and convection	Decreased febrile response
	At <6 mo, inability to shiver to increase body heat	Increased hypothermia
	Stress leads to hypoglycemia	
	Glycosuria is a sign of infection	
	At <2 yr, buffering capacity is decreased	
	Increased metabolic demands of growth	
Gastrointestinal and renal	Decreased endogenous calorie stores	Increased volume sensitivity
	Increased diarrhea with fluid and calorie deficits	Decreased nutritional status
	At <2 yr, gastric emptying is delayed	Increased likelihood of hypotensive or hypertensive renal damage
	At <2 yr, increased gastric distention	
	At <30% TBSA burn, can take sufficient caloric supplementation	Increased incidence of type 2 diabetes mellitus
	At >30% TBSA burn, requires caloric supplementation	Decreased creatinine clearance
	At <12 mo, poor renal filtration and absorption	
Neurocognitive	Increased incidence of cerebral edema with fluid resuscitation	Decreased brain mass
	Increased irritability	Decreased cerebral nerve cells
	Increased regressive behavior	Decreased brain cortex layer
	Increased risk-taking behavior	Decreased cerebellar cortical cells
	Immature judgment	Decreased nerve conduction velocity
		Decreased memory
		Decreased electroencephalographic activity
		Slower reaction time
		Decreased taste
		Decreased smell
		Decreased vision
		Decreased hearing
		Increased pain threshold
		Decreased judgment and cognitive abilities
Immune	Immature immune system	Decreased number of leukocytes
	Decreased immunocompetence	Decreased immunocompetence
		Decreased T-cell response

From Carrougher GJ, editor: *Burn care and therapy,* St Louis, 1998, Mosby, p 100.
BSA, Body surface area; *mo,* month(s); *TBSA,* total body surface area; *yr,* year(s).

The duration of exposure and the temperature, or the amount of energy, to which the skin is exposed primarily determine the degree of tissue/cellular injury, which is characterized by three zones (from inner to outer): (1) zone of necrosis; (2) zone of stasis, which is a region with decreased blood flow that either can be returned to its normal state with appropriate resuscitation or can be converted to a necrotic state in the case of dehydration, infection, or decreased perfusion; and (3) zone of hyperemia, which is minimally injured tissue that usually recovers normal function within 1 week.

In response to thermal injury, keratinocytes develop from cells in the basal layer of the epidermis and progress upward from the stratum germinativum to the stratum corneum over a 14-day period. During this time the wound develops a light pink or reddish coloration, with normal skin color restored in a delayed fashion by the melanocytes. The regeneration of hair and nails is dependent on viability of the hair follicle and nail matrix. With intact follicles, hair generally regrows at approximately 1 cm per month. New nail formation is often irregular and of abnormal thickness as regrowth occurs.

Depth Classification

The depth of burn injury is divided into four classifications: first-degree burns, also known as *superficial burns;* second-degree burns, also known as *superficial and deep partial-thickness;* third-degree burns (full-thickness); and fourth-degree burns (full-thickness with bone or muscle involvement), based on criteria established by the American Burn Association (Table 54.2).

First-degree burns involve only superficial tissue destruction in the outermost layers of the epidermis, with no associated compromise of the function of the skin (see Table 54.2). These burns are often associated with local discomfort, erythema, and mild systemic responses such as headache, chills, nausea, and vomiting. Erythema, a thermovascular response that occurs in first-degree burns in the absence of direct trauma to the dermis, is probably related to the release of tissue contents into the superficial circulation. First-degree burns are generally self-limiting, require no fluid resuscitation, and are therefore not included in estimates of the percentage of TBSA burned. Therapy generally includes simple analgesia. These injuries typically heal in 3 to 6 days without scarring or pigmentation changes.

Second-degree (superficial partial-thickness) burns involve the epidermis to the level of the dermis and appear red to pale ivory. Moist, thin-walled blisters often form within minutes of the injury (Fig. 54.2). Pain is a major clinical feature of this depth of injury because tactile and pain sensors remain intact (see Table 54.2). Injuries typically heal in 7 to 21 days in the absence of wound infection. The amount of scarring that

FIG 54.2 This young child sustained a hot water scald burn on the heel of the right foot that resulted in a superficial second-degree burn. A superficial second-degree (partial-thickness) burn will reepithelialize within 3 weeks. These burns are characterized by loss of epidermis (blistering) and by a shiny, sensate, vascularized dermis. (Courtesy Michael Peck, MD, University of North Carolina Burn Center, Chapel Hill.)

TABLE 54.2 Burn Wound Classification

Degree of Burn	Cause of Injury	Depth of Injury	Wound Characteristics	Treatment Course
First-degree burn	Prolonged ultraviolet light exposure, brief exposure to hot liquids	Limited damage to epithelium, skin intact	Erythematous, hypersensitive, no blister formation	Complete healing within 3–5 days without scarring
Superficial partial-thickness burn: second degree	Brief exposure to flash, flame, or hot liquids	Epidermis destroyed, minimal damage to superficial layers of dermis, epidermal appendages remain intact	Moist and weepy, pink or red, blisters, blanching, hypersensitive	Complete healing within 21 days with minimal or no scarring
Deep partial-thickness burn: second degree	Intense radiant energy, scalding liquids or hot semiliquids (e.g., tar) or solids, flame	Epidermis destroyed, underlying dermis damaged, some epidermal appendages remain intact	Pale, decreased moistness, blanching absent or prolonged; intact sensation to deep pressure but not to pinprick	Prolonged healing (often longer than 21 days), may require skin grafting to achieve complete healing with better functional outcome
Full-thickness burn: third degree	Prolonged contact with flame, scalding liquids, steam; hot objects; chemicals; electrical current	Epidermis, dermis, and epidermal appendages destroyed; injury through dermis	Dry, leatherlike; pale, mottled brown, or red; thrombosed vessels visible; insensate	Requires skin grafting
Full-thickness burn: fourth degree	Electrical current, prolonged contact with flame (e.g., unconscious victim)	Epidermis, dermis, and epidermal appendages destroyed; injury involves connective tissue, muscle, and possibly bone	Dry; charred, mottled brown, white, or red; no sensation; limited or no movement of involved extremities or digits	Requires skin grafting, amputation of involved extremities or digits likely

Data From Carrougher GJ, editor: *Burn care therapy,* St Louis, 1998, Mosby, p 138., Bittner EA, et. al. Acute and perioperative care of the burn-injured patient. *Anesthes.* 2015; 122(2):448-464. *doi: 10.1097/ALN.0000000000000559*

follows is a genetically determined trait, with some groups of people tending to scar excessively (African Americans and Caucasians with red hair) or minimally (Native American and Asian groups). Hair follicles remain intact and will regrow hair in the area of injury. Hair usually reappears 7 to 10 days after injury.

Second-degree (deep partial-thickness) burns may involve the entire dermis and leave only the epidermal skin appendages located in the hair follicles. The area of injury has a mottled appearance, with large areas of waxy-white tissue surrounded by light pink or red tissue. The surface is generally dry, and blisters tend to resemble flat, dry tissue paper rather than the fluid-filled raised areas seen with superficial partial-thickness injury. Tactile and pain sensors are either absent or greatly diminished in the area of deepest tissue destruction, but this area is usually surrounded by margins of lesser depth of injury in which pain and tactile sensors remain intact. Deep partial-thickness injury is visually and clinically indistinguishable from full-thickness injury at the time of injury. These wounds heal spontaneously in previously healthy individuals in about 4 weeks in the absence of secondary infection. As the length of healing time increases, so does the degree of scarring and depigmentation. As a result, these burns are often excised early and subsequently treated with a skin graft in an effort to diminish scarring and achieve early wound closure.

Third-degree (full-thickness) burns may also involve the entire epidermis, the dermis, and the underlying subcutaneous tissue. Immediately after injury, these areas appear white, cherry red, or black. Deep blisters may be present under a dry layer of dehydrated skin. Superficial blood vessels coagulated by the heat of injury may be visible through the skin as thrombosed veins. One of the physiologic characteristics of the skin that is lost (Box 54.1) is the elasticity of the dermis, resulting in a wound with a dry, hard, leathery texture. The massive edema that accompanies major burn injury combined with the loss of elasticity may result in a tourniquet-like effect when the injury occurs circumferentially around a limb or torso. This often necessitates escharotomies or, rarely, fasciotomies to restore distal circulation.

These burns are painless to touch, because all superficial nerve endings in the skin have been destroyed. However, as with partial-thickness injuries, rarely are burn injuries totally uniform, and an area of lesser injury in which pain and tactile sensors are intact is usually located on the periphery. The areas of full-thickness injury require skin grafting with the patient's own skin because all dermal elements have been destroyed, leaving no residual tissue for regeneration. However, small injuries often heal by secondary intention as a result of ingrowth of dermal elements from the margins of the wound.

Fourth-degree (full-thickness) injuries that extend beyond the dermis to involve muscle, bone, or both are often classified as fourth-degree. These injuries often occur in victims of high-voltage electrical injury or in persons who have had prolonged exposure to intense heat, such as unconscious fire victims.

Extent of Injury

Extent of injury refers to the percentage of TBSA burned. Estimates can be calculated with the rule of nines (Fig. 54.3) or the Lund and Browder chart (Fig. 54.4). The rule of nines is commonly used in prehospital settings and emergency departments and provides a rough estimate of TBSA involved.

Severity Classification

The severity of a burn injury is determined by the extent to which the physiologic functions of the skin are disrupted beyond the body's normal ability to respond with compensatory mechanisms. The American Burn Association classifies burn injury as minor, moderate, and major (Table 54.3). The severity of the burn injury and the eventual morbidity and mortality associated with it are related to a combination of factors: the patient's medical history, the extent and depth of the burn, the body area involved, the presence of concomitant trauma sustained at the time of the burn, and the patient's age.

BOX 54.1 Normal Physiologic Functions of the Skin Altered or Lost After Thermal Injury

Protection
- Barrier between the internal organs and the external environment
- Continuous with the mucous membrane at the external openings of organs of the digestive, respiratory, and urogenital systems
- Acidic skin (pH 4.2 to 5.6) and perspiration protect against bacterial invasion
- Thickened skin of palms and soles provides padding

Percutaneous Absorption
- Epidermis is relatively impermeable to most chemical substances; some may be absorbed through the epidermis or the orifices of hair follicles

Sensory Processing
- Receptor skin nerve endings allow constant monitoring of the environment by sensing warm and cold temperature, pain, touch, and pressure

Production
- Endogenous production of vitamin D_3, which is necessary for synthesis of vitamin D

Barrier
- Skin prevents water and electrolyte loss, maintains moist subcutaneous tissues, and prevents water absorption during immersion

Thermoregulation
- Body continuously produces heat as a by-product of cellular metabolism; heat is dissipated through skin
- Internal body temperature is regulated by radiation, conduction, or convection
- Rate of heat loss depends primarily on the surface temperature of skin, which is a function of skin blood flow

Immunologic
- Major location of immune complexes is at the dermal-epidermal junction and the dermal vessels of skin
- Monocyte/macrophage system is mobilized by local tissue mediators

Circulatory
- Skin temperature depends on the rate of blood flow through the skin
- Circulatory system distributes pharmacologic agents to local tissues

Esthetic
- Provides the individual identity of a person

Acute Management

The first priority in burn management is the elimination of the source; however, extreme caution must be exercised to ensure that the rescuer does not become a victim. The fire should preferably be extinguished with water, because it not only eliminates the source of the heat but also enables cooling of the underlying skin. In the absence of water, flames may be smothered with a blanket, coat, or any other nonflammable covering that will aid in deprivation of the oxygen required for combustion. Once the flames are eliminated, the cover should be promptly removed to enable dissipation of the underlying heat, thereby minimizing

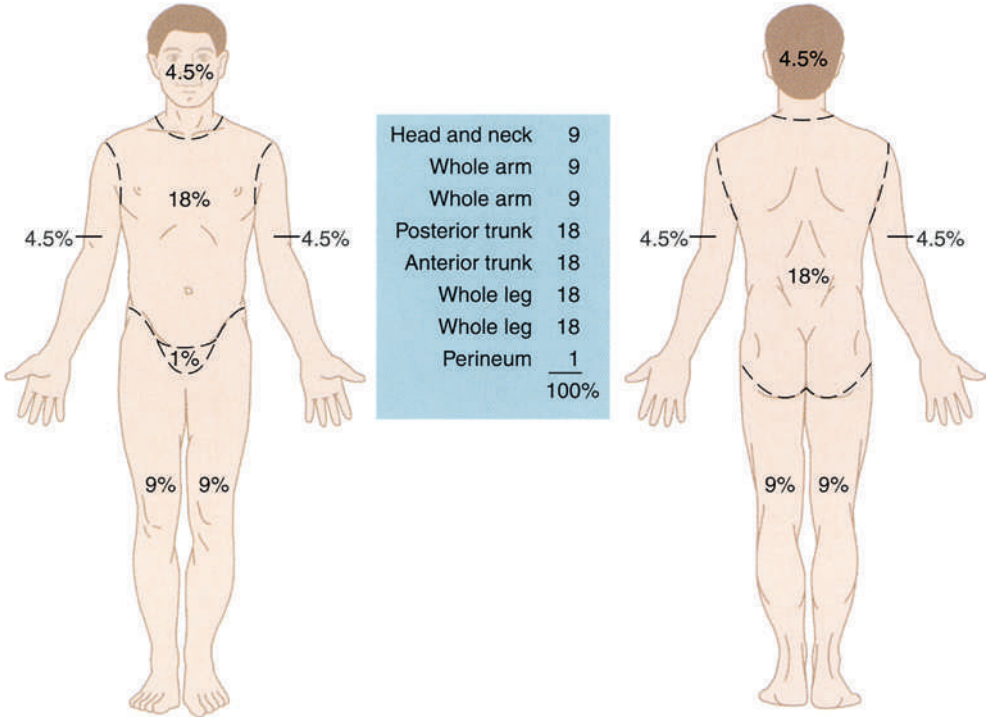
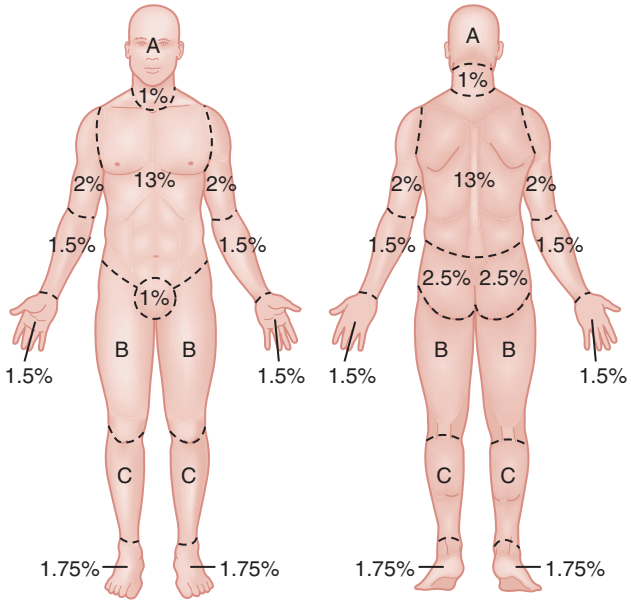


FIG 54.3 The rule of nines is a commonly used assessment tool that permits a timely and useful estimate of the percentage of total body surface area burned.



Relative percentages of areas affected by growth

Age	Half of head (A)	Half of one thigh (B)	Half of one leg (C)
Infant	9.5	2.75	2.5
1 yr	8.5	3.25	2.5
5 yr	6.5	4	2.75
10 yr	5.5	4.25	3
15 yr	4.5	4.25	3.25
Adult	3.5	4.75	3.5

FIG 54.4 The Lund and Browder chart for estimation of burn surface area for children and adults. (From Walls RM et al: *Rosen's emergency medicine: concepts and clinical practice,* ed 9, Philadelphia, 2018, Elsevier.)

TABLE 54.3 American Burn Association Burn Severity Classification Schedule

Classification	Assessment Criteria
Minor burn injury	<15% TBSA burn in adults <40 yr old <10% TBSA burn in adults >40 yr old <10% TBSA burn in children <10 yr old *and* <2% TBSA full-thickness burn without risk of cosmetic or functional impairment or disability
Moderate burn injury	12%–25% TBSA burn in adults <40 yr old 10%–20% TBSA burn in adults >40 yr old 10%–20% TBSA burn in children <10 yr old *and* 10% TBSA full-thickness burn without cosmetic or functional risk to burned area involving face, eyes, ears, hands, feet, or perineum
Major burn injury	>25% TBSA burn in adults <40 yr old >20% TBSA burn in adults >40 yr old >20% TBSA burn in children <10 yr old *or* >10% TBSA full-thickness burn (any age) *or* Injuries involving face, eyes, ears, hands, feet, or perineum likely to result in functional or cosmetic disability *or* High-voltage electrical burn injury *or* All burn injuries with concomitant inhalation injury or major trauma

Data from American Burn Association guidelines. In Carrougher GJ, editor: *Burn care and therapy,* St Louis, 1998, Mosby, p 94; Harvin KR, Norrie TE. Anesthetic management of patients with major burn injury. *AANA Journal,* 2012;(80):6.
TBSA, Total body surface area.

BOX 54.2 Burn Unit Referral Criteria

- Partial-thickness burns greater than 10% TBSA
- Burns that involve the face, hands, feet, genitalia, perineum, or major joints
- Third-degree burns in any age group
- Electrical burns, including lightning injury
- Chemical burns
- Inhalation injury
- Burn injury in patients with preexisting medical disorders that could complicate management, prolong recovery, or affect mortality
- Any patients with burns and concomitant trauma (such as fractures) in which the burn injury poses the greatest risk of morbidity or mortality; in such cases, if the trauma poses the greater immediate risk, the patient may be initially stabilized in a trauma center before being transferred to a burn unit; physician judgment will be necessary in such situations and should be in concert with the regional medical control plan and triage protocols
- Burns in children; children with burns should be transferred to a burn center verified to treat children. In the absence of a regional pediatric burn center, an adult burn center may serve as a second option for the management of pediatric burns
- Burn injury in patients who will require special social, emotional, or long-term rehabilitative intervention

From American College of Surgeons Committee on Trauma: *Resources for optimal care of the injured patient,* Chicago, 2014, The College, p 101.
TBSA, Total body surface area.

injury depth. Scald injuries are best treated initially with cool water, which allows cooling of the scalding liquid as well as the underlying skin.

Assessment

Initial management of patients with thermal burn injuries should focus on stabilizing the ABCs (airway–breathing–circulation). Common signs of inhalation injury include cough, stridor, hoarseness, singed nasal hair, carbonaceous sputum, oropharyngeal edema, and blisters. These injuries generally evolve over time, often progressing to complete airway obstruction secondary to edema. Soot in the oral cavity or facial burns should raise the suspicion of inhalation injury; therefore consideration should be given early to bronchoscopy and endotracheal intubation. Compromise of breathing in burn patients is often attributed to underlying inhalation injury or circumferential full-thickness burns resulting in impaired chest excursion; however, breathing difficulties may also arise from associated pneumothorax, hemothorax, or chest wall instability from multiple segmental fractures (flail chest). Vascular access is preferentially obtained peripherally through unburned tissue if possible. If no such sites are available, access may be established through burned skin. The patient is then evaluated by a complete head-to-toe examination for any other associated traumatic injuries, which are identified and managed as required. TBSA is then determined using a standardized chart such as the Lund and Browder chart (see Fig. 54.4). Fluid resuscitation requirements are subsequently determined and initiated while the patient is placed on clean sheets and the burned areas covered with dry clean sheets or dressings. Coverage with cool wet sheets should be avoided, because these quickly become cold wet sheets, and with the loss of the burned skin's ability to regulate body temperature, hypothermia can quickly ensue. Application of topical agents in the acute setting outside of a burn center is not recommended. A Foley catheter is also placed to monitor urine output during resuscitation, and in patients with burns greater than 20% of the TBSA, if possible, a nasogastric tube should be placed to allow for gastric decompression and to minimize the risk of abdominal bloating and aspiration. Transfer of patients to burn units or other facilities with appropriate resources is initiated during the course of the initial assessment. The American College of Surgeons Committee on Trauma has developed a set of burn unit referral criteria to assist initial evaluators in triage (Box 54.2).

Burn Shock and Acute Resuscitation

Two different but simultaneous mechanisms occur in cases of major burns: local wound pathophysiologic processes related to the loss of skin integrity and systemic pathophysiologic processes related to sequelae of the burn injury.

Within minutes of a burn injury, the cardiovascular system, which is normally a closed, semipermeable system, becomes an open system through which the patient's circulating volume leaves the circulatory system. This phenomenon, known as *capillary leak,* occurs within a few minutes of injury and persists for 24 hours. Burn shock is not confined to the burn area, but rather is a systemic process. While the capillary system throughout the body becomes leaky, fluid lost in the area of the burn leaks through the burn into the environment in an evaporative fashion, whereas fluid loss internally collects in the nearby soft tissues, producing extensive interstitial edema (Fig. 54.5).

Restoration of the patient's circulating volume is an essential part of acute burn management. The rate and volume of fluids lost are

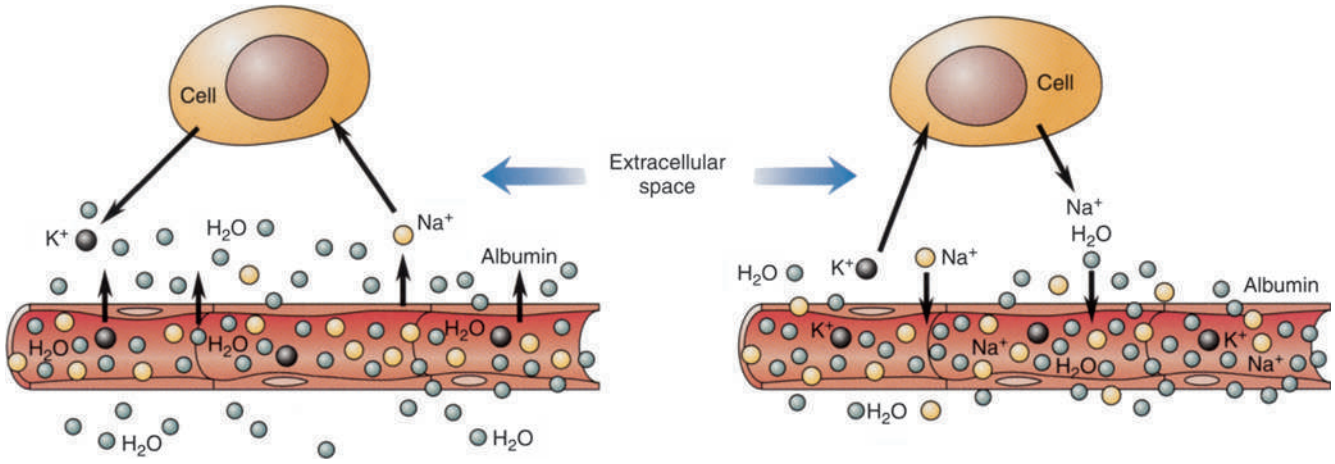


FIG 54.5 Direction of fluid and electrolyte shifts associated with burn shock. During burn shock, K^+ is exiting the cell, and Na^+ and H_2O are moving into the cell. After burn shock, K^+ enters the cell and Na^+ and H_2O exit it.

BOX 54.3 Parkland Formula for Fluid Resuscitation in Burn Shock

During the first 24 hours after a burn, administer intravenous LRS at the following rate:

4 mL LRS/% TBA burn/kg body weight

where

- Time is calculated from the time of burn injury.
- TBSA is total body surface area.
- Half of the total fluid is administered in the first 8 hr after burn.
- One-fourth of the total is administered in the second 8 hr.
- One-fourth of the total is administered in the third 8 hr or in quantities to maintain adult urine output at 30 mL/hr or child urine output at 1 mL/kg/hr.

Example of formula calculation in a 70-kg patient with a 50% TBSA burn:

4 mL × 70 kg × 50% TBA burn = 14,000 mL (14 L) LRS in 24 hr

- Administer 7000 mL in the first 8 hr at 875 mL/hr.
- Administer 3500 mL in the second 8 hr at 437 mL/hr.
- Administer 3500 mL in the third 8 hr at 437 mL/hr.

Data from Guilabert P et al: Fluid resuscitation management in patients with burns: update, *Br J Anaesth* 117(3):284–296, 2016.
LRS, Lactated Ringer solution; *TBSA,* total body surface area.

BOX 54.4 Plasma Requirements and Evaporative Water Loss After Burn Injury

Colloid Replacement

Twenty percent of blood volume given as fresh-frozen plasma or plasma expander
Adult males: 20% = 5 mL/kg body weight
Adult females and children: 20% = 8 mL/kg*

Maintenance Fluids Until Wound Closure Is Achieved

Basal fluid requirements:
1500 mL of fluid/m^2 TBSA = 24-hr requirement
Evaporative water loss from burn wound until healed:
Adults: (25 + % TBSA burn) × m^2 BSA = mL/hr requirement
Children: (35 + % TBSA burn) × m^2 BSA = mL/hr requirement
Maintenance fluids equal basal fluid requirements plus evaporative water loss and may be administered intravenously, orally, or by nasogastric or jejunal tube, according to patient need.

(*Data from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3038406/.)
BSA, Body surface area; *TBSA,* total body surface area burned.

related directly to the severity of the burn. Therefore the extent and depth of the burn injury must be ascertained during the initial clinical assessment.

The most widely used formula to guide fluid resuscitation within the first 24 hours of burn injury is the Parkland formula (Box 54.3). The formula utilizes lactated Ringer solution as the resuscitation fluid because it most closely approximates the fluid it is replacing and thereby minimizes the profound electrolyte imbalances often seen with large-volume resuscitation (Box 54.4). The standardized formulas provide an excellent guideline for initiating fluid resuscitation, but adequacy of resuscitation is determined by the global response of the patient to the fluid administration and not by one single variable. Recent reviews of practices have demonstrated that starting with the Parkland formula may lead to overresuscitation. It is now recommended to start with 2 mL/kg/%TBSA/24 hr (Brook formula) and increase to 4 mL/kg/%TBSA/24 hr as needed. Markers commonly used to reflect adequacy of resuscitation include normalization of mental status, blood pressure, pulse rate, capillary refill time, arterial pH, and base deficit and maintenance of urine output at 0.5 to 1 mL/kg per hour for adults. In addition to the fluid volumes calculated by the standardized formulas, the patient should receive appropriate maintenance fluid over the first 24-hour period. Proposed biomarkers for acute kidney injury in burn patients include neutrophil gelatinase–associated lipocalin, insulin-like growth factor binding protein 7, and tissue inhibitor of metalloproteinase 2.

Approximately 24 hours after the acute burn injury, the capillary leak syndrome begins to resolve as cardiovascular integrity is restored. At this time, a colloid solution such as albumin may be administered according to an appropriate formula, which differs by gender and age (see Box 54.4), in an effort to replace the protein lost during the acute burn shock phase. Resuscitation in the acute setting with colloid has been shown to decrease net volume of fluid administered and

complications associated with large-volume resuscitation such as abdominal compartment syndrome.

One of the major functions of intact skin is to serve as a barrier to water evaporation. With major burn injury, this ability of the skin to regulate evaporative loss is disrupted. In a classic study conducted in 1962, Moncrief and Mason attempted to determine the magnitude of such a loss and found that daily evaporative loss was in the range of 20 times normal during the early phase of burn injury, with gradual decreases as wound closure was achieved. Further studies revealed that the insensible water loss through burned skin is not caused by evaporation of water from sweat glands, but rather by water vapor formed within the body and lost through the skin. As the patient progresses to the subacute phase of resuscitation, it is imperative that attention be paid to these unaccountable losses.

Organ Dysfunction

Cardiovascular Dysfunction

Burn shock is often accompanied by a precipitous drop in cardiac output that does not parallel the gradual reduction in blood volume and is refractory to restoration of the circulating volume. This finding of low cardiac output in the presence of vigorous intravenous fluid resuscitation and massive catecholamine release has led to the suggestion of a specific myocardial depressant factor. However, there appears to be no simple, specific myocardial depressant factor, but rather a cascade of events involving multiple factors (Fig. 54.6). Abu-Sittah et al (2012) reviewed the factors contributing to cardiovascular dysfunction, which included intravascular volume loss, myocyte dysfunction and toxicity due to circulating cytokines, influence of hormones on cardiac function, paradoxical function of cardiac beta receptors, and presence of inhalation injury.

Respiratory Dysfunction

Respiratory dysfunction after burn injury generally is the result of obstruction, interstitial alterations, and metabolic changes. Obstruction may occur as a result of edema of the upper airway secondary to direct injury, or more often is attributable to the generalized edema that occurs after fluid resuscitation in the face of an ongoing capillary leak syndrome that accompanies burn injuries. On occasion, burns of the oral cavity and upper airway occur as superheated air is inhaled or scalding water enters the mouth. The pulmonary system is extremely efficient at dissipating heat and prevents the inhalation of superheated air beyond the bronchi, but steam may permeate further into the lung parenchyma (see Fig. 54.6). Airway obstruction secondary to edema generally has its onset within the first few hours after burn injury, but tends to manifest clinically 2 to 4 hours later as resuscitation is undertaken. Endotracheal intubation is recommended prophylactically when impending airway obstruction is identified. The endotracheal tube is preferably secured with methods/devices that enable readjustment to allow for the increase in head and face circumference related to increasing facial and soft tissue edema.

Smoke or fume inhalation often leads to acute hypoxia that is refractory to oxygen administration. Inhalation injury is reported to occur in 5% to 35% of patients. The gold standard for diagnosis is bronchoscopy; however, lung involvement on computed tomography may be helpful. Inhalation injury directly results in the chemical denaturing of pulmonary tissue, and the subsequent edema results in increased distances over which oxygen must diffuse to the capillaries. This may progress to acute respiratory distress syndrome (ARDS), which usually occurs within the first week. Whether the actual inhalation injury causes susceptibility toward ARDS is still a subject of debate.

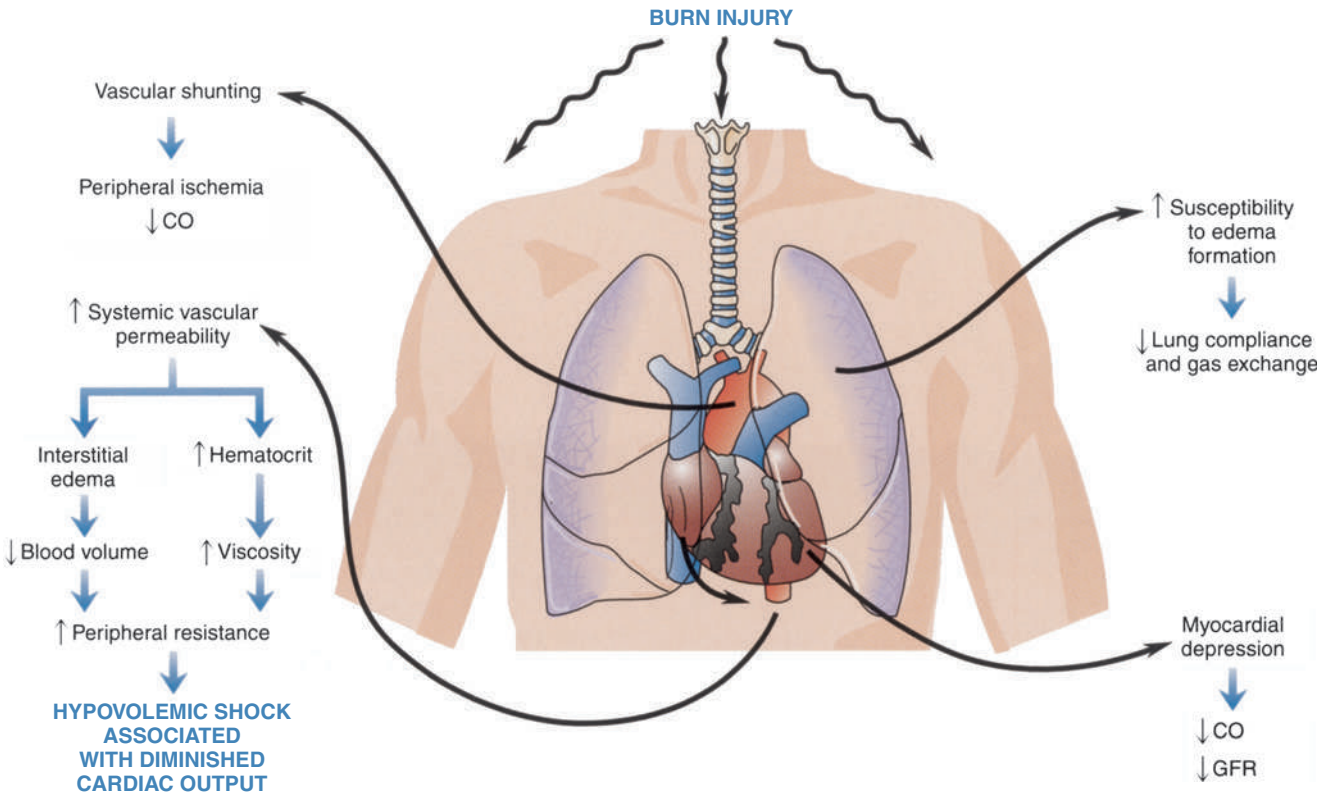


FIG 54.6 Cardiovascular and pulmonary effects of major burn injury within the first 24 hours after burn injury, during burn shock. *CO,* Cardiac output; *GFR,* glomerular filtration rate.

However, it is clear that many factors play a role in this condition because of the intense inflammatory response to the burn in addition to contributions from sepsis, shock, pneumonia, multiple organ failure, and increasing duration of mechanical support. Current ventilatory strategies emphasize use of low tidal volumes and increased frequency of breaths, which have resulted in decreased mortality.

Smoke inhalation is also responsible for the majority of cases of inadvertent carbon monoxide poisoning. Carbon monoxide is in direct competition with oxygen for hemoglobin-binding sites and has much greater affinity for the oxygen-binding sites on hemoglobin. Once carbon monoxide has bound to hemoglobin, it becomes carboxyhemoglobin, which has decreased ability to off-load oxygen to peripheral tissues. It is also responsible for inactivation of cytochrome oxidase, an enzyme that aids in oxygen utilization. The vast range of clinical symptoms can include vague constitutional symptoms such as headache, nausea, and fatigue, or may be much more severe and lead to lactic acidosis, seizures, and coma. The degree of carboxyhemoglobinemia is largely dependent on the amounts of carbon monoxide and oxygen in the native environment, the duration of exposure, and the minute ventilation of the patient.

Carbon monoxide poisoning is diagnosed with a compatible history and physical in conjunction with the presence of elevated carboxyhemoglobin levels. Carboxyhemoglobin can be directly measured on arterial blood gas samples, although these levels may not correlate with the development of neurologic sequelae. In addition, pulse oximetry plays no role in monitoring patients with carbon monoxide poisoning because it cannot differentiate between oxyhemoglobin and carboxyhemoglobin.

Treatment involves removal of carbon monoxide from hemoglobin via its direct competition with oxygen. The normal half-life of carbon monoxide in a patient breathing room air is 300 minutes. If high-flow oxygen is delivered via face mask, the half-life can be decreased to 90 minutes. Therefore treatment involves removal of the patient from the source of inhalation and immediate institution of high-flow 100% oxygen via a nonrebreather face mask. In cases of severe poisoning manifested by a carboxyhemoglobin level greater than 25%, loss of consciousness, severe metabolic acidosis, or evidence of ischemic change in any organ system, the treatment should involve institution of hyperbaric oxygen therapy. This will reduce the half-life of carbon monoxide even further to approximately 30 minutes, and it also increases the amount of oxygen dissolved in the blood (unbound to hemoglobin) from 0.3 to 6 mL/dL. In addition to assisting with elimination of carbon monoxide, the addition of hyperbaric oxygen improves delivery of oxygen to peripheral tissues, even in the presence of elevated carboxyhemoglobin levels.

Renal Dysfunction

Acute renal failure (ARF) after burn injury can be classified as early or late. Early renal failure occurs during the first 5 days postburn, secondary to low intravascular volumes and rhabdomyolysis-associated myoglobinemia. The etiology of renal failure in rhabdomyolysis is twofold: first, the shift of fluid into damaged muscle from the intravascular space (i.e., third spacing) causes reduced blood volume and less flow to the kidneys, with direct nephrotoxic effects of myoglobin by-products and possible renal failure. Second, the by-products of myoglobin can be toxic to the kidneys and cause damage by precipitation into the renal tubules. Development of early renal failure can be minimized by adequate fluid resuscitation and decompressive fasciotomy for compartment syndrome that may contribute to further muscle damage. Late ARF, defined as occurring after 5 days postburn, is less of a result of the initial burn and is generally attributable to infectious or direct nephrotoxic factors such as medications. Most patients will require renal supportive therapy. Continuous venovenous hemofiltration has been used with success in treating ARF in burns and is associated with fewer vascular complications than arteriovenous hemodialysis. The prognosis is generally poor for the burn patient who develops ARF. Mortality rates from 73% to 100% have been described. Survival is dismal for patients with early-onset ARF associated with rhabdomyolysis. However, even in the face of severe ARF, renal function will generally recover over time in those patients who survive.

Metabolic Changes

The metabolic changes associated with major burn injuries primarily involve the sympathetic nervous system and are manifested as sustained tachycardia (120 to 140 beats/min range) and increased oxygen consumption (approximately 150% of normal). The hypermetabolic state generally persists until the burn wound is reduced to less than 20% of TBSA and poses a major challenge in the treatment of these patients. Modalities to attenuate the hypermetabolic response have included early excision and grafting, increasing ambient temperatures, use of propranolol, administration of counterregulatory hormones such as insulin and insulin-like growth factor, and stimulation of anabolism using growth hormones and steroids (natural or synthetic). A recent meta-analysis by Flores et al in 2015 revealed benefit from propranolol with regard to cardiac function, liver function, metabolic status, and wound healing without any adverse side effects.

The muscle protein wasting appears to be caused primarily by accelerated protein breakdown. Although protein synthesis is also increased, it fails to keep pace with proteolysis and amino acid mobilization. Thomas and colleagues evaluated the effect of continuous infusions of insulin (to maintain blood glucose levels between 100 and 140 mg/dL) on preservation of muscle mass in a randomized controlled clinical trial in 18 children with major burns. Evaluation of each patient when the burn wounds were 95% healed showed that insulin-treated patients had improved lean body mass, diminished muscle wasting, and reduced length of hospital stay in comparison with controls.

Gianotti and colleagues studied temporal fluctuations in insulin-like growth factor type I and its binding protein in a group of burn patients and demonstrated that they both decline for the first 14 days after the burn, paralleling decreases in prealbumin and transferrin levels. However, plasma levels of growth hormone remained unchanged. In a small study of burned children, administration of insulin-like growth factor type I and insulin growth factor binding protein appeared to exert widespread effects in ameliorating acute inflammation by reducing the synthesis of type 1 and type 2 acute-phase proteins and interleukin 6 (IL-6) while increasing the synthesis of constituted protein such as prealbumin and transferrin. Anabolic steroids, including testosterone and its analogs, such as oxandrolone, are widely used to ameliorate muscle wasting in patients with cancer and acquired immunodeficiency syndrome; however, they have had limited benefits in burn patients.

Cellular Changes

Major burn injury affects the entire body, but survival ultimately depends on its effects at the cellular level. The cellular response to burn injury occurs as a metabolic and an immunologic pathophysiologic process. The basic pathologic condition, named the *sick cell syndrome* by Welt in 1967, is a cell membrane transport defect related to an alteration in the steady-state composition characterized by high intracellular concentrations of sodium. Entry of sodium into the intracellular space occurs simultaneously with entry of water, which leads to cellular edema and possibly rupture of the cell wall. An associated decrease in resting membrane potential occurs as the transmembrane potential is disrupted and results in a decrease in amplitude of the action potential and prolongation of the repolarization and depolarization times. As sodium and water enter the cell, the sodium–potassium pump is disrupted and potassium moves out, thus further exacerbating the electrolyte imbalance

intracellularly. Calcium channel transport is disrupted, along with a loss of intracellular magnesium and phosphate and an increase in serum lactate dehydrogenase levels. The cascade of events that occurs at the level of the cell membrane suggests impairments of basic cellular function as the underlying cause of the diminished membrane potentials. Although the pathophysiologic mechanism has not been completely described, data suggest a decrease in the efficiency of the sodium–potassium pump, a change that can be reversed over time with adequate fluid resuscitation.

Evidence suggests that the burn wound itself at least partially mediates the physiologic response to burn injury at both local and systemic levels. Burn tissue inflammation can lead to vasodilation, increased capillary permeability, and edema, which are normal conditions that promote wound healing. Massive injury results in increased metabolic demands and consumption of inflammatory mediators by the wound, when the priority for survival should be transport of these mediators to healthy tissue.

The extensive evaporative water loss that accompanies burn injury is a heat-consuming process, with the energy need met in part by increased visceral heat production. This hypermetabolic state persists during rest, sleep, and external cooling. The increased oxygen consumption cannot be accounted for on the basis of elevated body temperature alone; thus an increased basal metabolic rate, not a thermoregulatory drive, is responsible for the increased heat production.

Immune Response

Local and systemic physiologic changes are primarily mediated by the release of cytokines from burn wounds. Cytokines act directly on the burn wound and activate other agents, including those that release oxidants, arachidonic acid metabolites, and proteases, thereby contributing to further local and systemic inflammation and, potentially, multisystem organ dysfunction. Recent studies have shown a relationship between decreased cellular cytotoxicity in burned patients and increased production of the cytokines IL-4 and IL-10. Administration of immunopotentiators such as IL-12 has also demonstrated increased survival and resistance to bacterial infection.

A host of chemicals found in altered concentrations in burn plasma may also play a role in burn shock. These substances include vasoactive amines (histamine, serotonin), products of complement activation (C3a, C5a), prostaglandins, kinins, endotoxins, and metabolic hormones (catecholamines, glucocorticoids). A decrease in the complement components C3a and C5a in the circulation after burn injury suggests nonspecific activation of the complement system. Activation of the complement system in injured tissue results in an inflammatory response caused by the release of histamine and serotonin by C3a and C5a. Because both histamine and serotonin alter capillary permeability, some investigators propose this mechanism as a cause for burn shock because these vasoactive amines initiate the inflammatory response along with kinin polypeptides and other chemical mediators. As a result of these vascular changes, fluid and fibrinogen leave the dilated, permeable vessels.

Elements of Burn Injury Survival

The emergent phase of burn care refers to the time between the end of burn shock and the closure of the burn wound to less than 20% of the TBSA. Three elements are essential for survival after a major burn injury: meticulous wound management, adequate nutritional support to establish positive nitrogen balance, and timely surgical excision and grafting of full-thickness wounds.

Management of Wounds

Optimal wound management strategies center around measures to limit bacterial proliferation on the wound and adjacent tissue after loss of the first line of defense—the skin. Burn wounds are sterile initially as a result of thermal decontamination; however, bacterial flora soon reestablish colonies on the burned skin, or eschar. This medium is favorable for pathogenic growth because of the necrotic tissue and warm environment that exist within the burn wound dressing. Benign microorganisms normally found on skin, in the gastrointestinal tract, and in the pulmonary system become lethal as they colonize the burn wound, often resulting in burn progression, invasive sepsis, and death.

The most common source of burn wound bacteria is the patient's own hair follicles, sweat glands, pulmonary tract, and gastrointestinal system, although poor handwashing technique by staff members can contribute to infection through cross-contamination from other patients.

The goals of wound care are to cleanse and debride the wound of necrotic tissue and debris that promote bacterial growth, minimize further destruction of viable tissue, prevent cross-contamination, preserve body heat and energy, and promote patient comfort. The excision of blisters is controversial and varies among institutions. Some argue that the blister provides a moist and protective environment for healing, whereas others state that large tense blisters apply pressure on the underlying wound bed (which impedes healing) and the contained fluid may serve as a nidus for infection. The decision to debride a blister should take into consideration the size and type of the blister. Blisters less than 6 mm in diameter are unlikely to rupture or place pressure on underlying tissue and should not be debrided. Thick-walled blisters attributable to thick skin, particularly on the hands and feet, should be left intact because they are not likely to become infected, and debridement often leads to limited mobility of the affected extremity. Large tense blisters or those with thin walls that are likely to rupture and become infected can impede wound healing, and debridement of these lesions should be considered. This debridement may also allow for further evaluation of the depth of injury in the underlying wound bed. Additional wound care requires daily observation and management, which includes bathing the patient at least once each day with mild soap and water. Burn wounds are washed to remove accumulated bacteria and previously applied ointments and to debride necrotic tissue. Cleansing of wounds is the most stressful and painful experience that burn patients endure. Pain medication diminishes the pain only marginally because the most effective analgesics work best on visceral or deep pain rather than pain at superficial skin nerve endings. Benzodiazepines are often added to decrease anxiety and provide a degree of amnesia.

After the wound is clean, topical antibacterial agents are applied and covered with a light dressing (Table 54.4). Systemic antibiotics are not helpful in controlling burn wound flora because the burn eschar has no blood supply, limiting local antibiotic bioavailability. Topical burn agents penetrate the eschar, thereby inhibiting bacterial invasion of the wound. Systemic antibiotics are administered when the patient demonstrates signs of systemic infection and are used prophylactically at times of surgical procedures. Appropriate antibiotic selection is based on laboratory cultures of the patient's wound tissue to identify and deliver antibiotics to which the bacteria are sensitive.

The healing of burn wounds begins when white blood cells have surrounded the burn wound and phagocytosis begins. Necrotic tissue begins to slough. Fibroblasts begin to build matrices of the collagen precursors that eventually form granulation tissue. Kept free from infection, a partial-thickness burn will heal from the edges and from below in a process that occurs over a 14- to 21-day period. Full-thickness burns require autografting to achieve wound closure because no dermal elements are available to form new skin.

Burn Surgery

The third element essential to survival after major burn injury is surgical excision of dead skin, or burn eschar, followed by skin grafting with

TABLE 54.4 Topical Antibiotic Therapy for Thermal Injury Wounds

Drug	Indications for Use	Advantages	Disadvantages	Method of Use
Bacitracin	Bland ointment with minimal antibiotic properties used to promote comfort in patients with minor injury (<25% TBSA)	Prevents drying of wounds; keeps eschar soft and pliable; economical; works well on facial burns to promote healing and patient comfort without facial dressings; painless upon application	No major antibiotic properties; oil based, so it is difficult to remove	1. Apply to cleansed wound twice daily; cover with Adaptic and Kerlix 2. Apply to facial burns twice daily 3. Apply to recently grafted or healed areas twice daily; wrap with Adaptic and Kerlix
Silver sulfadiazine (Silvadene)	Partial- and/or full-thickness thermal injury (>25% TBSA); small wounds that require topical antibiotic therapy such as frostbite	Wide-spectrum bacteriostatic action and painless on application; organisms resistant to silver nitrate are usually sensitive to silver sulfadiazine; eschar remains soft and pliable; water-miscible base promotes ease of removal	Not effective against fungal organisms; can cause leukopenia and is expensive; sulfa component can produce allergic reactions in sensitive patients; resistance can emerge with prolonged use	Apply to cleansed wound 1–3 times daily; may leave wound open or cover with light dressing
Silver nitrate	Partial- and/or full-thickness burns (>25% TBSA), fungal infections, patients with sulfa allergy	Wide-spectrum bacteriostatic action; effective against fungal infections; comfortable and economical; no sensitivity reported; painless on application; no resistant organisms	Can cause severe electrolyte imbalances (hyponatremia and hypochloremia), which are corrected with oral and intravenous NaCl; poor penetration into wound; requires bulky dressing, thereby severely limiting motion; messy and time consuming to use	0.5% solution in distilled water applied to wet dressing every 2 hr; dressing changes twice daily
Mafenide acetate (Sulfamylon)	Electrical injury, ear burns, wounds colonized with organisms resistant to other topical agents because it penetrates eschar more deeply	Wide-spectrum bacteriostatic action; active penetration allows delayed therapy to be effective; requires no dressing, thereby promoting motion; resistant organisms do not develop with prolonged use; drug of choice for all burns	Causes severe metabolic alterations within 72 hr when used on >20% TBSA wounds; carbonic anhydrase inhibition with HCO_3^- excretion and chloride retention; compensation is by hyperventilation with subsequent CO_2 decreased or depletion	Apply to cleansed wound 1–2 times daily; leave open because wrapping produces maceration

From Kravitz M: Thermal injuries. In Cardona VD et al, editors: *Trauma nursing: from resuscitation through rehabilitation,* Philadelphia, 1988, Saunders, p 723.
TBSA, Total body surface area.

the patient's own skin (autograft). Areas of full-thickness burn are excised. In some cases, it is difficult to assess the depth of the burn. In these cases, it may be appropriate to wait 7 to 10 days to allow the area to declare its depth. This is especially true in young children with scald burns. In burns involving the face, scalp, and ears, it may be appropriate to wait as long as 3 weeks for the area to declare itself. These sites are particularly dense areas of dermal appendages and may heal without grafting if given time. Current surgical management of burn wounds has evolved from daily bathing and mechanical debridement of necrotic tissue for months, to early surgical excision and grafting, with significant associated decreases in attendant morbidity and mortality.

Excision and Grafting

Significant blood loss associated with burn excision traditionally resulted in local wound management until eschar separation, at which time grafting was undertaken. This often resulted in prolonged hospital stays of approximately 3 to 5 weeks, increased bacterial colonization of wounds, and subsequent higher incidences of sepsis and multiple organ failure. Current practice now favors early excision and grafting of wounds 2 to 7 days after injury. Although blood loss remains a concern with early excision, improvements in excision and hemostasis techniques have enabled a significant decrease in the degree of magnitude.

In tangential excision, eschar is removed in thin layers with an instrument called a *dermatome* until viable tissue is visible. Full-thickness excision using a surgical scalpel removes eschar to the level of fascia. Full-thickness excision often leaves an uneven contour postexcision, which is difficult to graft, with resultant poor cosmetic effects. After bleeding has been controlled in the area of excision, application of an autograft, skin substitute, or dressing follows. In some cases, the area may need to be covered with wet dressings soaked in antibiotic solutions for 24 hours with delayed autografting. Some studies have shown a benefit with use of a vacuum-assisted closure (VAC) device to promote blood circulation in the wound bed, allowing for earlier placement of an epidermal graft.

Ong and colleagues conducted a meta-analysis of data from six randomized controlled trials published from 1966 through 2004 that

compared early excision of burns with wound dressing and grafting after eschar separation. No difference in overall mortality was observed, except for a lower mortality in the subgroup without inhalation injury that underwent early excision. Early excision patients overall had higher blood transfusion requirements and shorter lengths of stay. No evidence of reduced sepsis or a better cosmetic or functional outcome was noted with early excision.

Skin Substitutes

In patients with burns that do not permit initial autografting, temporary coverage may be achieved by use of any of the following skin substitutes: biological (homograft—skin harvested from cadavers, xenograft—skin harvested from pigs); synthetic; and amnion (amniotic lining of human placenta harvested from afterbirth after delivery). Application often enhances patient comfort while partially restoring the water vapor barrier, thereby minimizing evaporative losses. Grafts often consist of a bilayer membrane that provides a dermal matrix of bovine collagen and an epidermal or silicone layer to prevent desiccation (drying), or the dermal and epidermal components may be replaced separately. Bioactive dermal components added to the graft help stimulate healing and include elastin, fibronectin, growth factors, glycosaminoglycans, and hyaluronic acid. The composition of a permanent skin substitute often includes the patient's own cells in either the dermal or the epidermal layer. The dermal matrix allows ingrowth of capillaries and fibroblasts. This matrix is slowly degraded as a neodermis develops. The silicone layer acting as the epidermal barrier is then removed and autograft is applied over the neodermis.

The unburned area of the patient from which skin is harvested in a paper-thin sheet is referred to as the *donor site.* Donor sites heal in about 5 to 7 days in the presence of adequate nutritional support and the absence of infection and can be reharvested at that time. Donor sites can be repeatedly harvested depending on graft thickness, enabling increased wound coverage, which thereby permits survival in some patients with TBSA injury as large as 90%. To expand the surface area that a sheet of autograft will cover, harvested skin is cut in a manner that resembles a net or mesh by using an instrument called a *skin mesher.* The skin may then be expanded, depending on the size of the mesh, to cover two, three, four, or more times its original size. This combination of repeated harvesting and meshing allows autografting of massive burn injuries over a period of a few weeks. After grafting, the areas must be protected from infection, pressure, shearing, and trauma that produce bruising or bleeding under the graft. Major causes of graft loss include infection, blisters, or hematoma underneath the graft that interferes with revascularization. Negative-pressure dressings or VAC devices placed over a graft improve the contact surface of the graft as well as prevent underlying accumulation of serum or blood. Compared with standard bulky dressings, use of these devices has demonstrated a decrease in loss of graft, a decreased need for secondary grafting procedures on the same wound, and a resulting decrease in hospital stay for these procedures.

The greatest risk of infection is after postoperative day 3, when the bacteria begin to recolonize the area. Grafts are usually stable by postoperative day 4, at which time physical and occupational therapy may begin.

Nutritional Support

One of the most significant advances in recent burn management is recognition of the critical importance of early nutrition to the wound-healing process. The magnitude of nutritional support required by burn patients depends on two factors: the patient's preburn nutritional status and the extent of the TBSA burn. Patients with minor burns require no nutritional support beyond a regular diet, whereas those with moderate and large burns require additional carbohydrate and protein supplementation. Patients with poor preburn nutritional status are classified as having a critical injury, regardless of the burn size, because of the associated immune deficiencies and limited metabolic reserves. The most easily recognized and documented finding in the absence of adequate nutritional support after burn injury is massive loss of body weight. Maintenance of body protein is critical for healing, minimizing complications, and survival. A 10% loss of total body mass leads to immune dysfunction; 20%, to decreased wound healing; 30%, to severe infections; and 40%, to death.

Patients with a greater than 40% TBSA burn demonstrate the maximal stress response within predictable ranges of body mass. In these hypermetabolic patients, providing early protein and caloric support of at least the predicted energy requirement is necessary for optimal outcome and may be essential for survival. Providing early nutrition in enteral form may also help by blunting the hypermetabolic response to thermal injury. Weight loss after thermal injury is not an obligatory component of the response to trauma, but rather a reflection of the difference between the total energy requirements and the ability to supply them in the form of adequate caloric intake. Kao and colleagues demonstrated that enteral feeding should begin within 18 hours of admission with a Dobhoff feeding tube. These feedings should be continuous and should not be stopped when the patient goes to the operating room. With vigorous nutritional support, erosion of total body mass and subsequent starvation leading to immunologic alteration are not inevitable in a massively burned patient.

General formulas are used to estimate the caloric requirements of burn patients, all of which are based on either preburn body weight and %TBSA burn or square meters of body surface area and % of TBSA burn. Caloric requirements in adult burn patients could be expressed by the following formula:

$$(25) \times (\text{body weight in kg}) + (40) \times (\%\ \text{TBSA burn}) = \text{Ideal 24-hour caloric needs}$$

If the ideal daily caloric intake can be maintained, then the amount of postburn weight loss could be minimized. It is important to emphasize that these formulas represent more than just total caloric intake; they are used to predict positive nitrogen balance for each patient. Thus if the patient is losing tremendous amounts of nitrogen or is not absorbing glucose, the net caloric utilization will be much less than the intake, even though the adult patient may be receiving as much as 5000 kcal/day. Monitoring of daily nitrogen balance by indirect calorimetry is essential throughout the course of burn treatment to ensure a positive nitrogen balance (nonprotein kilocalorie to nitrogen ratio of 100:1 and at least 2 grams of protein per kilogram per day). The serum prealbumin concentration is a useful indicator of nutritional progress.

The routes for initiating caloric support after major burn injury are either enteral or parenteral. Any patient with a functioning gastrointestinal tract should receive enteral nutrition orally, by tube feeding, or by a combination of both. Early enteral nutrition has been shown to effectively deliver caloric requirements (resting energy expenditure) by postburn day 3; diminish the hypermetabolic response; decrease circulating levels of catecholamines, cortisol, and glucagon; and preserve gut mucosal integrity, motility, and intestinal blood flow, which serves to decrease bacterial translocation and lower the incidence of intestinal ischemia. Postburn ileus primarily affects the stomach and colon. Patients with severe burn injuries may be fed through enteral tubes to the small bowel (duodenum or jejunum) as early as 6 hours postburn, independent of total gastroduodenal function. In a large meta-analysis conducted in critically ill patients that included a small number of burned patients, a grade B evidence-based recommendation was made for the use of

parenteral nutrition for patients in whom enteral nutrition cannot be started within the first 24 hours of hospital admission. A subgroup analysis attributed a mortality risk reduction to parenteral nutrition versus delayed (>24 hours) enteral nutrition, despite an association with increased infectious complications with parenteral nutrition.

Catabolic response may persist for up to 12 months in adults and 24 months in children. Given this information, the nutritional status and dietary habits of burn patients should be continually evaluated for many months after their discharge from the burn unit.

Rehabilitation Phase

The rehabilitation phase begins when the burn size is reduced to less than 20% of the TBSA and the patient is capable of assuming some self-care. This phase may occur as early as 2 weeks or as long as 2 to 3 months after the burn and, in the case of a major debilitating or disfiguring injury, may last many years. Goals for this period are to assist the patient in resuming a functional role in society and to accomplish functional and cosmetic reconstruction.

Wound Healing

During the rehabilitation phase, the pathophysiologic mechanism of hypermetabolism and the impaired immune function have begun to normalize, although some changes will persist beyond discharge from the hospital. The major pathophysiologic process of this phase is related to the dysfunctional results of wounds healing in a manner that causes flexor contractures, excessive scarring, and keloid formation. The burn wounds have healed either by primary intention or by autografting. Layers of epithelialization begin rebuilding the tissue structure destroyed by the burn injury. Collagen fibers present in the new scar tissue help healing and add strength to weakened areas. After healing, the new skin appears flat and pink, even in dark-skinned people.

In approximately 4 to 6 weeks the area becomes raised and hyperemic. If adequate range-of-motion exercises are not instituted early in the hospital course, the new tissue will shorten and a contracture will result (Fig. 54.7). Mature healing is reached in 6 to 12 months, when suppleness has returned and the pink or red color has faded to a slightly lighter hue than the surrounding unburned tissue. It takes longer for darker skin to regain its color because many of the melanocytes were destroyed, and often the skin never returns to its original color. The mesh pattern in meshed autograft fades with time, but in larger expansions such as 4:1 or greater the pattern may persist.

Scarring has two components: discoloration and contour. The discoloration of scars fades with time and can be covered with makeup on visible body surface areas. However, scar tissue tends to develop altered contours; that is, the skin is no longer flat but becomes raised above the contour of the surrounding area (also known as *hypertrophic scarring*). Areas of the face tend to scar in an even plane—a process that distorts the natural contours around the nose, chin, and mouth and thus greatly alters a patient's appearance. Scarring on the cheeks can contract and pull the lower eyelid down sufficiently to prevent closure and protection of the eye normally afforded by the eyelid—a condition called *ectropion.* Burns on the eyelid can also result in ectropion and must be corrected by reconstructive surgery. Pressure can help keep a scar flat if the pressure is slightly greater than capillary pressure and is continuous during the healing process. This knowledge led to the development of burn garments, which are custom-made for each patient to contour with pressure over the area of the burn for about 12 to 18 months after burn injury. Except for bath times, the garments must be worn continuously; patient compliance often becomes an issue (Fig. 54.8).

Excessive and sometimes debilitating discomfort from pruritus occurs in the healing burn wound and persists for many months. The exact pathophysiologic process is not known, but is related to the absence of sebaceous glands in the area and to the hyperactivity of sweat glands. Topical lotions and orally administered antihistamines provide partial relief of symptoms, but tolerance to the drugs develops, and patients often require a series of different medications over time. The newly formed skin is extremely sensitive to trauma, and blisters form after very slight pressure or friction. The newly healed areas may be hypersensitive or hyposensitive to cold, heat, or touch. Ward and colleagues studied loss of cutaneous sensibility after grafting in 60 patients and found that 97% demonstrated markedly diminished or absent responses to sharp/dull, hot/cold, and light touch stimuli over the grafted areas. Grafted

FIG 54.7 Physical and occupational therapy is necessary from the time of injury. This 8-year-old girl was burned 4 years previously in a house fire. Inadequate follow-up care because of parental neglect led to severe scar contractures and hand disability. (Courtesy Michael Peck, MD, University of North Carolina Burn Center, Chapel Hill.)

FIG 54.8 The custom-fitted antiscar support garment modeled here effectively provides pressure therapy over wounds, which helps minimize the development of hypertrophic scarring. (Courtesy Medical Z, San Antonio, TX. In Black JM, Hawks JH, editors: *Medical-surgical nursing: clinical management for positive outcomes,* ed 7, Philadelphia, 2005, Saunders.)

areas are more likely to be hyposensitive until peripheral nerve regeneration occurs, although donor sites harvested several times will show all the same healing pathologic process as healed burn wounds.

Scarring is a genetically inherited trait. Some people will have minimal scarring, whereas others, especially African Americans and Caucasians with red hair, tend to have significant scarring and keloid formation in which the scar tissue actually outgrows the boundary of the original wound. Healed burn wounds must be protected from direct sunlight for 1 year to prevent hyperpigmentation.

The most common complications during the rehabilitation phase are related to the formation of skin and joint contractures. Because of pain associated with movement, the patient will want to assume the position of comfort, which is with all extremities flexed, but this position predisposes to contracture formation. To minimize contracture formation, positioning in extension, splinting in the position of function, and performing active range-of-motion exercises are initiated on admission and continue throughout the course of treatment. The areas that are most at risk to contracture formation include the anterior and lateral neck areas, axillae, antecubital fossae, fingers, groin areas, popliteal fossae, and ankles. Not only do contractures develop in the skin, but also the underlying tissues such as ligaments and tendons have a tendency to shorten during the healing process. Therapy is aimed at extension of body parts to ensure that the flexors are longer than the extensors.

FIG 54.9 High-voltage electrical injuries produce devastating consequences, such as the damage to the right hand of this electrician who inadvertently contacted a 17,000-V line. The underlying muscle damage is often greater than that found in a thermal skin burn. Myoglobinuria, if inadequately managed, can lead to acute tubular necrosis. Early fasciotomies are mandatory, and amputation may be necessary to control rhabdomyolysis. (Courtesy Michael Peck, MD, University of North Carolina Burn Center, Chapel Hill.)

KEY POINTS

- The emergent phase is the time between the end of burn shock and closure of the wound to less than 20% of the TBSA. Wound management, nutritional support, and surgical grafting of full-thickness wounds are the priorities of treatment during the emergent phase.
- Wound management is necessary to prevent bacterial colonization of the wound and subsequent septicemia. Early surgical wound management is essential. Topical antibiotics are used because systemic antibiotics cannot reach the wound because of a lack of blood supply.
- Nutritional requirements after burn injury are high. A high-calorie, high-protein diet is needed. Persons with major burns usually cannot ingest sufficient nutrients and require parenteral and enteral supplementation. A positive nitrogen balance is essential for healing.
- Early surgical excision and skin grafting are the treatments of choice for deep burns. Excision procedures result in significant blood loss requiring blood transfusions. Skin grafts are taken from a healthy portion of the patient's skin. Temporary grafts (e.g., cadaver skin, synthetics, porcine skin) may be used to cover the wound until an autograft can be obtained.
- The rehabilitation phase begins when the burn is reduced to less than 20% of the TBSA. Problems during this phase include skin contracture and excessive scarring. Healing is complete at 6 to 12 months. Positioning in extension and performing range-of-motion exercises are important to prevent contracture.

ELECTRICAL INJURY

Incidence and Mortality

Electrical injury accounts for fewer than 4% of admissions to burn facilities; however, their incidence has been increasing in the United States. Based on data from the National Institute for Occupational Safety and Health National Traumatic Occupational Fatalities surveillance system, electrocutions were the fifth-leading cause of death from 1980 through 1992.

Electrical injuries are classified as high voltage (1000 volts or greater) or low voltage. Household currents of 120 and 220 volts typically cause low-voltage electrical injury. High-voltage injuries are frequently due to high-tension sources, which commonly carry from 7200 to 19,000 volts (Fig. 54.9) but may involve 100,000 to 1 million volts.

According to the National Weather Service Storm Data, between 2010 and 2015, the United States has averaged 27 *reported* lightning fatalities per year. Of all lightning strikes, only 10% are fatal, leaving 90% with various degrees of disability. Lightning injuries kill between 20 and 30 people per year in the United States. Lightning carries a direct current of 100 million or more volts and up to 200,000 amperes, and it can injure either by a direct strike or by a side flash as a result of the flow of current between the victim's body and a nearby object struck by lightning.

Arnoldo and colleagues reviewed electrical injuries at a single institution over a 20-year period and reported that the highest mortality resulted from lightning strikes (17.6%) followed by high-voltage (5.3%) and low-voltage (2.8%) injuries, with the lowest electrical injury mortality resulting from electric arc injuries without passage of current through the patient (1.1%). Complication rates, mean length of hospital stay (18.9 ± 1.4 days), and number of operative procedures (3.0 ± 0.2) were increased in the high-voltage group. Work-related activity was responsible for the majority of these high-voltage injuries, with the most common occupations being linemen and electricians.

Pathophysiology

The pathophysiologic mechanism of electrical injury is related to the subsequent tissue damage as electrical energy is converted to heat. Workplace electrocutions account for 4% of all worker deaths. In children, electrical burns account for 2% to 3% of all burns, and 23% of these result from biting extension cords.

Arcing electricity produces surface heat, which may ignite clothing and destroy superficial tissue, but internal damage is absent; this injury is actually a flame or thermal injury and not electrical. These injuries are properly classified as heat injuries, for which the treatment plan is identical to that for other heat injuries. True electrical injury occurs as electrical current enters the body, traverses a portion of the body, and exits at another body site. Electrical injuries are usually deeper than full-thickness skin injury and are often classified as fourth-degree injury.

Voltage, type of current (direct or alternating), and length of contact all influence the extent of damage. Alternating current (AC) produces prolonged tetanic muscle contraction. At low voltages it can cause ventricular fibrillation, tetanic contraction of the respiratory muscles, superficial burns, and rhabdomyolysis. At lower voltage, AC is associated with low mortality. High-voltage AC or direct current causes ventricular fibrillation, a single sustained contraction, rhabdomyolysis, and higher overall mortality.

Each true electrical injury produces an entrance wound and at least one exit wound, with the most extensive damage commonly occurring at the exit point. Electrical current follows the path of least resistance: in humans, this path is through blood vessels, nerves, tendons, and bone. Skin has high resistance; thus the current enters through the skin but goes deeper to travel the path of least resistance until it exits the body. The current rarely produces direct visceral damage, but severe injuries to the extremities are common. The amputation rate after electrical injury ranges from 24% to 43%. The pathophysiologic process, in addition to direct tissue destruction, involves heat coagulation of blood vessels, which leaves distal areas without blood supply. Electrical injuries produce both systemic and local alterations. The systemic changes produce three common complications during the acute period: dysrhythmias or cardiac arrest, metabolic acidosis, and myoglobinuria. Electrical injury may also cause direct myocardial necrosis. Dysrhythmias are exacerbated by any given voltage of AC. Higher voltage may also cause asystole. Locally, electrical injury produces direct cellular denaturation; areas of healthy tissue are devascularized as a result of heat coagulation of arteries and veins. These events are followed 48 to 72 hours after injury by gross tissue necrosis and subsequent gangrene resulting from lack of blood flow. Amputation is required early in electrical injury to prevent the development of deep soft tissue infections and sepsis, leading to death.

Management and Complications

Once the patient is admitted, airway management is the primary focus; patients with major electrical injury often require endotracheal intubation to ensure a patent airway. A condition similar to burn shock develops within a few minutes of major electrical injury and requires similar fluid resuscitation measures; however, there is no standardized formula to predict fluid requirements because often the only apparent damage is the entrance and exit wounds, and no assessment of internal damage is possible. An adult patient is given a 1-L bolus of lactated Ringer solution intravenously within the first 15 minutes after intravenous line placement and frequently requires 1 to 2 L of fluid per hour to support the cardiovascular system.

Cardiovascular complications associated with lightning injury range from benign electrocardiogram (ECG) changes to sudden cardiac death. Atrial or ventricular fibrillation and T-wave inversions and prolonged QT intervals are the more common dysrhythmias and ECG findings. Traditionally, cardiac monitoring has been performed on these patients for the first 24 hours after injury. Bailey and colleagues determined that this is unnecessary, however, if the initial ECG is normal, there is no loss of consciousness at the scene, and the patient is an adult; 24-hour monitoring is indicated in adult patients with an abnormal initial ECG, a history of cardiac disease, positive loss of consciousness at the scene, and/or exposure to a voltage greater than 240 volts. Measurement of the levels of cardiac enzymes initially reveals elevated values, also suggesting acute myocardial damage, but in such patients these findings are not indicative of a cardiac pathologic process.

Electrical injury also produces a profound, potentially lethal metabolic acidosis. These patients often have initial serum pH values of 6.8 to 7.2 on admission. Treatment consists of intravenous administration of sodium bicarbonate in amounts to correct the values toward normal. Metabolic acidosis is a recurring problem requiring ongoing treatment until it has been resolved for 24 to 48 hours after injury. The pathophysiologic mechanism is related both to the release of intracellular contents into the general circulation from areas of tissue damage and to the development of lactic acidosis that accompanies hypotensive shock states.

Myoglobinuria follows electrical injury as myoglobin, a component of muscle tissue, is released from muscles damaged by electrical current and enters the systemic circulation. Myoglobin is a large protein that precipitates in the renal tubules and leads to cast formation. Subsequently, the tubules become obstructed and renal tubular acidosis develops. This accumulation is prevented by maintaining urine output at 100 to 200 mL/hr in adults and 2 mL/kg/hr in children until the urine clears. Mannitol, an osmotic diuretic, is administered along with large volumes of intravenous fluids to prevent the development of acute tubular necrosis, a totally preventable sequela of electrical injury with proper management. Sodium bicarbonate is often administered to alkalinize the urine, thereby increasing the solubility of myoglobin.

Local effects of electrical injury are related to alterations in tissue perfusion. Surgical decompression of areas of electrical burn by fasciotomy is performed for the purpose of releasing any increased compartment pressures that may be compromising blood flow. Amputation may be required during the initial surgery for devascularized areas. Because of the continued presence of necrotic tissue, areas of surgical decompression or initial amputation are not closed surgically.

Central nervous system alterations can be noted in patients with major electrical injury. Memory deficits can occur for several weeks. This condition improves gradually and usually resolves within 4 to 6 weeks after injury. Other central nervous system deficits after electrical injury include ataxia and gait alterations accompanied by sensory deficits. These alterations may or may not improve over time. Electrically injured patients experience all the challenges of rehabilitation plus possible adjustments to amputation and gait instability related to central nervous system impairment. Skin grafting in areas adjacent to amputation presents challenging prosthetic problems that may delay independent ambulation and restoration of self-care abilities. In general, patients with major electrical injury experience longer rehabilitation periods than do thermally injured patients.

A unique complication of electrical injury is the formation of corneal cataracts, which can be detected as early as 1 month post injury. Ophthalmic examinations should be performed monthly for the first year and every 3 months for 1 year thereafter to enable early identification. Patients will usually complain of blurring vision, but young children may not report this visual change because they do not recognize the concept. Treatment consists of corneal transplantation.

KEY POINTS

- Broadly speaking, persons wounded by major electrical burns have longer periods of rehabilitation than do thermally injured patients. Systemic involvement is complex. Each true electrical injury produces an entrance wound and at least one exit wound, with the most extensive damage commonly occurring at the exit point.
- Management for electrical injury includes core practices of maintenance of airway, breathing, and circulation. Once the patient is admitted to the hospital for a serious electrical burn, airway management is the primary focus.
- A condition similar to burn shock develops within a few minutes of major electrical injury and requires similar fluid resuscitation measures; however, there is no standardized formula to predict fluid requirements because often

the only apparent damage is the entrance and exit wounds and no assessment of internal damage is possible.
- Cardiovascular complications associated with electrical (lightning) injury range from benign ECG changes to sudden cardiac death. Twenty-four-hour monitoring is indicated in adult patients with an abnormal initial ECG, a history of cardiac disease, positive loss of consciousness at the scene, and/or exposure to a voltage greater than 240 volts.
- Electrical injury also produces a profound, potentially lethal metabolic acidosis. These patients often have initial serum pH values of 6.8 to 7.2 on admission.
- Treatment for metabolic acidosis consists of IV administration of sodium bicarbonate. Metabolic acidosis is a recurring problem requiring ongoing treatment until resolved for 24 to 48 hours after injury. The pathophysiologic mechanisms are related both to the release of intracellular contents into the general circulation from areas of tissue damage and to the development of lactic acidosis that accompanies hypotensive shock states.
- Myoglobinuria follows electrical injury as myoglobin, a component of muscle tissue, is released from muscles damaged by electrical current and enters the systemic circulation. Myoglobin is a large protein that precipitates in the renal tubules and leads to cast formation. Subsequently, the tubules become obstructed and renal tubular acidosis develops.
- Accumulation of myoglobin is prevented by maintaining urine output at 100 to 200 mL/hr in adults and 2 mL/kg/hr in children until the urine clears. Mannitol, an osmotic diuretic, is administered along with large volumes of IV fluids to prevent the development of acute tubular necrosis. With proper treatment, acute tubular necrosis is totally preventable. Sodium bicarbonate is often administered to alkalinize the urine, thereby increasing the solubility of myoglobin.
- Local effects of electrical injury are related to alterations in tissue perfusion. Because of the continued presence of necrotic tissue, areas of surgical decompression or initial amputation are not closed surgically.
- Electrically injured patients experience all the challenges of rehabilitation plus possible adjustments to amputation and gait instability related to central nervous system impairment.
- A unique complication of electrical injury is the formation of corneal cataracts, which can be detected as early as 1 month post injury. Treatment consists of corneal transplantation.

CHEMICAL INJURY

Management and Complications

Chemicals are a mainstay in our everyday lives and encompass a diverse spectrum, ranging from occupational to household exposures. Chemical burns accounted for 3% of admissions to burn centers in the United States. Management protocols for these injuries share some basic core principles, yet have unique aspects depending on chemical composition, making accurate knowledge of involved chemical(s) imperative.

Acids predominantly produce a coagulation necrosis by denaturing proteins and forming a coagulum (eschar) that limits the penetration of the acid. In contrast, alkalis produce a liquefaction necrosis that, in addition to involving denaturing proteins, results in fat saponification, which does not limit tissue penetration and results in more severe injuries. The severity of a chemical burn is related to a number of factors, including the agent's pH, concentration, volume, and physical form as well as the length of contact time. To minimize local and systemic toxic effects, it is critical that treatment be initiated immediately.

Initial management strategies include removing any contaminated clothing, brushing off dry agents, and irrigating with copious amounts of water. Care providers must ensure that they have taken the appropriate protection measures to avoid self-exposure as established by the Occupational Safety and Health Administration. Water irrigation is contraindicated for certain chemicals and metals such as lime, phenol, muriatic acid, concentrated sulfuric acid (52% to 100%), elemental potassium, and sodium, because it may result in an exothermic reaction or release of hazardous by-products. Antidotes have a minor role in the management of most chemical burns except for those secondary to hydrofluoric acid.

The true extent of chemical burns is often difficult to appreciate in the acute setting given the nature of these injuries to continue to evolve over time after exposure. As a result, a lack of appreciation for unrecognized deep tissue involvement may result in inadequate initial management. Systemic toxicity associated with these injuries occurs primarily as a result of absorption through the skin and inhalation.

Comprehensive management guidelines for clinicians for specific chemical exposures are readily accessible online at https://www.osha.gov/dts/osta/bestpractices/firstreceivers_hospital.pdf or by calling regional poison control centers. The World Health Organization provides a listing of international poison centers online at http://www.who.int/ipcs/poisons/centre/en/.

Common Agents and Treatment

Hydrofluoric Acid

Hydrofluoric acid is a highly corrosive inorganic acid used in glass etching, electronic industries, and cleaning solutions. Symptomatic manifestation after exposure is dependent on solution concentration, with concentrations greater than 15% manifesting more acutely. Due to the ability of hydrofluoric acid to penetrate tissue, poisoning can occur readily through exposure of skin or eyes or when inhaled or swallowed. Symptoms of exposure to hydrofluoric acid may not be immediately evident. Hydrofluoric acid interferes with nerve function, meaning that burns may not initially be painful. Accidental exposures can go unnoticed, delaying treatment and increasing the extent and seriousness of the injury.

Hydrofluoric acid penetrates quickly into the dermis and deeper structures, where the fluoride ion forms complexes with calcium and magnesium. This complex formation in conjunction with the direct cardiotoxic effects of fluoride ions may contribute to the development of cardiac dysrhythmias, which are the primary cause of death in these patients.

The management of hydrofluoric acid injuries consists of copious water irrigation and calcium administration. Calcium ions form complexes with free fluoride ions, thereby minimizing toxicity while also helping correct any associated hypocalcemia. Calcium gluconate (2.5%) gel is massaged into affected areas for 30 to 60 minutes. If discomfort persists, 5% calcium gluconate (0.5 mL/cm^2 wound area) may be injected intradermally directly into and around the affected areas. Injection directly into digits is not recommended. Burns refractory to these initial measures may necessitate intraarterial administration of calcium gluconate. This should only be undertaken in conjunction with the appropriate toxicology expert. In cases where systemic toxicity is suspected (prolongation of QT interval, dysrhythmias) intravenous calcium and magnesium replacement should be considered.

Anhydrous Ammonia

Anhydrous ammonia is a colorless, pungent gas usually stored as a pressurized liquid at −33° C (−28° F) and used extensively as a fertilizer and in the manufacture of synthetic fibers and methamphetamine. Exposure generally results in a combination of alkali burns and cold injury. It is extremely water soluble, and immediate treatment consists of copious water irrigation after all clothing has been removed. Irrigation should be repeated every 4 to 6 hours for the first 24 hours. Inhalation injury is concentration dependent and may range from minor airway

irritation to laryngospasm, glottic edema, and pulmonary edema. Early intubation in patients with suspected or clinical evidence of significant inhalation exposure is critical. There is unfortunately no specific therapy for these inhalation injuries except supportive care.

Cement Burns

Wet cement is highly alkalotic, with the pH increasing as the cement sets. Burn symptoms are generally delayed, with partial- or full-thickness burns not becoming evident for up to 48 hours. Immediate treatment with copious water irrigation is highly effective in preventing injury progression. Management of burns is similar to that used for secondary thermal injuries. Prevention education that promotes appropriate use of protective gear can be highly successful in minimizing these injuries.

Chemicals Associated With Automobile Airbag Burns

Airbag perforations during deployment have been reported to cause both thermal and alkaline burns (sodium azide or sodium hydroxide). Lack of recognition of associated airbag rupture often results in failure to treat the potential alkali component of the burn and therefore concomitant injury progression. Management is simply irrigation with copious amounts of water.

Tar and Asphalt

In construction, both tar and asphalt are heated to high temperatures (approximately 140° C for paving; approximately 245° C for roofing); however, they cool rapidly. Initial treatment of injuries sustained by these substances consists of accelerating cooling by application of cold water. Subsequent removal of the substances, which may require multiple applications, can be facilitated by the application of several organic solvents (polymyxin-neomycin-bacitracin, petrolatum, sunflower oil, olive oil, butter, and baby oil). Associated burns are treated in a fashion similar to that used in thermal burn treatment.

KEY POINTS

- Burn injuries can be caused by thermal, electrical, or chemical agents. Depending on the cause, burn injuries influence a variety of complex systemic, circulatory, and metabolic changes.
- Management protocols for chemical burns share some basic core principles with electrical and thermal wounds, yet chemical injuries have unique treatment aspects depending on the composition of the etiologic agent, making accurate knowledge of involved chemical(s) imperative.
- To assist the clinician in the treatment of burns, comprehensive current management guidelines for specific exposures are also accessible online.
- The true extent of chemical burns is often difficult to appreciate, given that the nature of these injuries continues to evolve over the time after exposure.
- Systemic toxicity associated with chemical injuries occurs primarily as a result of absorption through the skin and by inhalation.

SPECIAL POPULATIONS

Introduction

Pediatric, geriatric, and obese patients consist of a large proportion of burned patients. Children less than 16 years old and adults older than 60 years old represent approximately half of all burned patients, and currently 35% of adults and 17% of kids age 2 to 19 are obese.

These populations require special considerations in patient management due to their unique physiologies and are at higher risk for complications and mortality and have increased hospital stay lengths. Table 54.1 summarizes the physiologic changes associated with age that increase the vulnerability of children and elderly persons.

Geriatric

Adults older than 60 years of age represent 13% of burn patients each year with an 11% mortality. Neglect or abuse is estimated to account for 20% of elderly burns; however, this is often difficult to prove due to some degree of memory impairment in 50% of these patients.

Increasing age and comorbidities pose potential challenges in patient management and operative management (Tables 54.1 and 54.5). Even first-degree burns may lead to dehydration, requiring resuscitation, due to the thin dermal and epidermal layers. Surgical wound management of elderly burn patients is determined by the philosophy of the burn center. Elderly patients often lack the physiologic reserve to tolerate surgical procedures as well, and as a result conservative, nonsurgical wound management has been employed for weeks after injury with acceptable survival rates. On the contrary there have also been reports demonstrating increased survival and decreased length of stay after early excision of eschar and wound closure.

Pediatric

Pediatric patients represent 30% of all burn victims, of which 6% to 10% require hospital admission. Eight percent to 10% of admitted patients will then go on to require operative interventions. Children less than 6 years of age are at increased risk for burns, even due to short exposure time, due to their thinner dermis. Predictors of mortality include inhalation injury, male gender, increased TBSA%, and age less than 4 years old. Neglect or abuse is estimated to occur in 10% to 20% of pediatric burn victims. Providers should have a high index of suspicion when the patient has bilateral burns, cigarette burns, sharp demarcation of wounds, or other signs of abuse. Abusers are more likely to be female, have lower educational level, be unemployed, and have a history of drug or alcohol abuse.

Younger pediatric patients have an altered distribution of body surface area in comparison to adults. The percent area of burn involvement for children less than 10 years old can be determined by using a modified Lund and Browder chart (see Fig. 54.4).

Table 54.1 outlines physiologic differences of pediatric patients compared with adults that may affect management. For example, due to a child's disproportional weight to body surface area, fluid losses are proportionally greater. Targeting a urine output of 1 to 2 mL/kg/hr is commonly used to measure effectiveness of resuscitation. Younger children are at greater risk of thermodysregulation. With regard to nutrition, caloric needs for children are most commonly calculated using the Polk formula:

$$(60) \times (\text{body weight in kg}) + (36) \times (\%\ \text{TBSA burn}) = \text{Ideal 24-hour caloric needs}$$

Obese

The high prevalence of obesity in adults requires providers to be aware of physiologic differences that may affect management. Obese patients have twice the length of stay and twice the rate of mortality, which is estimated at 36%. The increased rate of mortality is due to comorbidities, higher propensity to be entrapped at scene, and higher risk of postoperative complications. Intubation and ventilation (including bag-valve-mask ventilation) may be more difficult due to altered anatomy. Initial resuscitation should be based on a modified Lund and Browder chart, as the "rule of nines" does not adequately reflect the surface area in obese patients due to altered anatomy of the obese patient.

Common comorbidities include coronary artery disease, hypertension, pulmonary disease, altered immune function, poor wound healing,

TABLE 54.5 Physiologic Changes Related to the Aging Process That Can Affect Surgery

Physiologic Changes	Effects	Potential Postoperative Complication
Cardiovascular		
↓ Elasticity of blood vessels ↓ Cardiac output ↓ Peripheral circulation	↓ Circulation to vital organs Slower blood flow	Shock (hypotension), thrombosis with pulmonary emboli, delayed wound healing, postoperative confusion, hypervolemia, decreased response to stress
Respiratory		
↓ Elasticity of lungs and chest wall ↑ Residual lung volume ↓ Forced expiratory volume ↓ Ciliary action Fewer alveolar capillaries	↓ Vital capacity ↓ Alveolar volume ↓ Gas exchange ↓ Cough reflex	Atelectasis, pneumonia, postoperative confusion
Urinary		
↓ Glomerular filtration rate ↓ Bladder muscle tone Weakened perineal muscles	↓ Kidney function Stasis of urine in bladder Loss of urinary control	Prolonged response to anesthesia and drugs, overhydration with intravenous fluids, hyperkalemia, urinary tract infection, urinary retention
Musculoskeletal		
↓ Muscle strength Limitation of motion	↓ Activity	Atelectasis, pneumonia, thrombophlebitis, constipation or fecal impaction
Gastrointestinal		
↓ Intestinal motility	Retention of feces	Constipation or fecal impaction
Metabolic		
↓ γ-Globulin level ↓ Plasma proteins	↓ Inflammatory response	Delayed wound healing, wound dehiscence, or evisceration
Immune		
Fewer killer T cells ↓ Response to foreign antigens	↓ Ability to protect against invasion by pathogenic microorganisms	Wound infection, wound dehiscence, pneumonia, urinary tract infection

From Keeling AW, Muro GA, Long BC: Preoperative nursing. In Phipps WJ et al, editors: *Medical-surgical nursing: concepts and clinical practice,* ed 5, St Louis, 1995, Mosby.

increased wound infections secondary to poorer perfusion of adipose tissue leading to decreased neutrophil phagocytosis, and cell-mediated immunity. Obese patients are also at higher risk for deep vein thrombosis and pressure ulcerations.

Surgical management needs to be carefully planned due to increased size of soft tissue mass. Patient positioning and operative length of time may need to be altered in the obese population. Additional assistance or lifting equipment may be necessary for wound debridement and dressing changes. Postoperatively, skin grafts are more likely to be dislodged secondary to difficulty securing during placement and higher propensity to undergo sheer stress. Finally, caloric calculations for nutrition should be based on ideal body weight using the Curreri formula for adults.

KEY POINTS

- Familiarity of the distinct physiologies of extremes of age and obese patients is critical to appropriate resuscitation and medical management.
- Thin dermal layers make pediatric and geriatric patients at increased risk for more severe burns and may become dehydrated with superficial burns.
- Obese and pediatric patients have altered body surface area percentages. Referring to the modified Lund and Browder charts is necessary for calculating %TBSA for resuscitative purposes.
- Comorbidities of geriatric and obese patients may alter surgical management and outcomes.

SUMMARY

In recent decades, burn mortality rates have decreased significantly, with most patients achieving excellent functional and cosmetic outcomes. The improved outcomes have clearly been related to an improved understanding of the pathophysiologic mechanism of burns, advances in burn care management, and the development of a comprehensive, patient-centered, treatment-oriented approach.

Priorities for assessment and treatment of burn victims are no different from those of other trauma patients. However, accurate wound management depends on precise identification of the cause of the burn injury.

As a result, patient care is extremely complex. In order to make appropriate treatment decisions and recognize potential complications, a clear understanding of the etiologic agent of the burn injury, the associated pathophysiologic processes, and the importance of supportive care, treatment, and rehabilitation for the involved person are essential.

RESOURCES

Abdullahi A, Jeschke MG: Nutrition and anabolic pharmacotherapies in the care of burn patients. *Nutr Clin Pract* 29:621–630, 2014.

Abu-Sittah GS, Sarhane KA, Dibo SA, Ibrahim A: Cardiovascular dysfunction in burns: review of the literature. *Ann Burns Fire Disasters* 25:26–37, 2012.

Achora S, Muliira JK, Thanka AN: Strategies to promote healing of split thickness skin grafts: an integrative review. *J Wound Ostomy Continence Nurs* 41:335–339, quiz E331-332, 2014. doi:10.1097/WON.0000000000000035.

Ahrens M: *Home Structure Fires. National Fire Protection Association, Fire Analysis and Research Division* file:///C:/Users/docwa/Downloads/fireloss2014.pdf (2014).

American Burn Association: *National Burn Repository*. http://www.ameriburn.org/2015NBRAnnualReport.pdf. (Accessed 6 June 2016) (2015).

American College of Surgeons Committee on Trauma: *Resources for optimal care of the injured patient. American College of Surgeons Committee on Trauma* **Chicago** (2014).

Antonio AC, Castro PS, Freire LO: Smoke inhalation injury during enclosed-space fires: an update. *J Bras Pneumol* 39:373–381, 2013. doi:10.1590/S1806-37132013000300016.

Arnoldo BD, et al: Electrical injuries: a 20-year review. *J Burn Care Rehabil* 25:479–484, 2004.

Arnoldo B, Klein M, Gibran NS: Practice guidelines for the management of electrical injuries. *J Burn Care Res* 27:439–447, 2006. doi:10.1097/01.BCR.0000226250.26567.4C.

Arturson G: Forty years in burns research—the postburn inflammatory response. *Burns* 26:599–604, doi, 2000. doi:10.1016/S0305-4179(00)00069-3.

Association AB: *Burn incidence and treatment in the United States: 2015 fact sheet*. Available at www.ameriburn.org/resources_factsheet.php. (Accessed 30 September 2015) (2015).

Bailey B, Gaudreault P, Thivierge RL: Experience with guidelines for cardiac monitoring after electrical injury in children. *Am J Emerg Med* 18:671–675, 2000. doi:10.1053/ajem.2000.16307.

Berkowitz Z, Orr MF, Kaye WE, Haugh GS: Hazardous substances emergency events in the agriculture industry and related services in four mid-western states. *J Occup Environ Med* 44:714–723, 2002.

Bloom GR, Suhail F, Hopkins-Price P, Sood A: Acute anhydrous ammonia injury from accidents during illicit methamphetamine production. *Burns* 34:713–718, 2008. doi:10.1016/j.burns.2007.08.015.

Bozkurt A, O'Dey D, Pallua N: Treatment of hot bitumen-contact-burn injuries. *Burns* 34:1053–1054, 2008. doi:10.1016/j.burns.2008.02.006.

Brigham PA, Dimick AR: The evolution of burn care facilities in the United States. *J Burn Care Res* 29:248–256, 2008. doi:10.1097/BCR.0b013e31815f366c.

Brunicardia FC, editor: *Schwart'z principles of surgery*, ed 10, New York, 2015, McGraw-Hill.

Clardy PF, M S: Carbon monoxide poisoning. *Up To Date Online 20.2* (October 19, 2011).

Control NC *f. I. P. a. Fire deaths and injuries*. Available at www.cdc.gov/HomeandRecreationalSafety/Fire-Prevention/fires-factsheet.html. (Accessed 24 March 2012).

Cunningham JN, Shires T, Wagner Y: Changes in intracellular sodium and potassium content of red blood cells in trauma and shock. *Am J Surg* 122:650, 1971. doi:10.1016/0002-9610(71)90293-5.

Curreri PW, et al: Dietary requirements of patient with major burns. *J Am Diet Assoc* 65:415–417, 1974.

David WS: Myoglobinuria. *Neurol Clin* 18:215–243, 2000.

Deets DK, G. V: Plasma and cardiac lactic dehydrogenase activity in burn shock. *Proc Soc Exp Biol Med* 142:412–416, 1973.

Dega S, Gnaneswar SG, Rao PR, et al: Electrical burn injuries. Some unusual clinical situations and management. *Burns* 33:653–665, 2007. doi:10.1016/j.burns.2006.09.008.

DeSanti L, Lincoln L, Egan F, Demling R: Development of a burn rehabilitation unit: impact on burn center length of stay and functional outcome. *J Burn Care Rehabil* 19:414–419, 1998.

Diaz EC, et al: Effects of pharmacologic interventions on muscle protein synthesis and breakdown in recovery from burns. *Burns* 41:649–657, 2015. doi:10.1016/j.burns.2014.10.010.

Electrical Safety Foundation International: *Occupational Injury and Fatality Statistics*. http://www.esfi.org/workplace-injury-and-fatality-statistics (Accessed 6 June 2015).

Emara SS, Alzaylai AA: Renal failure in burn patients: a review. *Ann Burns Fire Disasters* 26:12–15, 2013.

Endorf FW, Ahrenholz D: Burn management. *Curr Opin Crit Care* 17:601–605, 2011. doi:10.1097/MCC.0b013e32834c563f.

Endorf FW, Dries DJ: Burn resuscitation. *Scand J Trauma Resusc Emerg Med* 19:69, 2011. doi:10.1186/1757-7241-19-69.

Fazeli S, et al: Predictive factors of mortality in burn patients. *Trauma Mon* 19:e14480, 2014. doi:10.5812/traumamon.14480.

Flores O, Stockton K, Roberts JA, et al: The efficacy and safety of adrenergic blockade after burn injury: A systematic review and meta-analysis. *J Trauma Acute Care* 80:146–155, 2016.

Garcia CT, Smith GA, Cohen DM, Fernandez K: Electrical injuries in a pediatric emergency department. *Ann Emerg Med* 26:604–608, 1995.

Gianotti L, et al: Activity of GH/IGF-1 axis in burn patients: Comparison with normal subjects and patients with GH deficiency. *J Endocrinol Invest* 25:116–124, 2002.

Gibran NS, Heimbach DM: Current status of burn wound pathophysiology. *Clin Plast Surg* 27:11, 2000.

Gnaneswaran N, Perera E, Perera M, Sawhney R: Cutaneous chemical burns: assessment and early management. *Aust Fam Physician* 44:135–139, 2015.

Gonzales R, Shanti C: Overview of current pediatric burn care. *Semin Pediatr Surg* 24:47–49, 2015.

Haberal M, Sakallioglu Abali AE, Karakayali H: Fluid management in major burn injuries. *Indian J Plast Surg* 43:S29–S36, 2010. doi:10.4103/0970-0358.70715.

Hager DN: Recent advances in the management of the acute respiratory distress syndrome. *Clin Chest Med* 36:481–496, 2015. doi:10.1016/j.ccm.2015.05.002.

Hall KL, Shahrokhi S, Jeschke MG: Enteral nutrition support in burn care: a review of current recommendations as instituted in the Ross Tilley Burn Centre. *Nutrients* 4:1554–1565, 2012. doi:10.3390/nu4111554.

Hamza T, Barnett JB, Li B: Interleukin 12 a key immunoregulatory cytokine in infection applications. *Int J Mol Sci* 11:789–806, 2010. doi:10.3390/ijms11030789.

Haynes BW, Jr: The management of burns in children. *Journal of Trauma* 5:267–277, 1965.

Holm C, Melcer B, Horbrand F, et al: The relationship between oxygen delivery and oxygen consumption during fluid resuscitation of burn-related shock. *J Burn Care Rehabil* 21:147–154, 2000. doi:10.1097/00004630-200021020-00011.

Ibrahim AE, Sarhane KA, Fagan SP, Goverman J: Renal dysfunction in burns: a review. *Ann Burns Fire Disasters* 26:16–25, 2013.

Jain S, Bandi V: Electrical and lightning injuries. *Crit Care Clin* 15:319, 1999. doi:10.1016/S0749-0704(05)70057-9.

Jeschke MG, et al: Pathophysiologic response to burns in the elderly. *Ebiomedicine* 2:1536–1548, 2015. doi:10.1016/j.ebiom.2015.07.040.

Karter M: *Fire loss in the United States during 2011. National Fire Protection Association, Fire Analysis and Research Division* http://www.nfpa.org/research/reports-and-statistics/demographics-and-victim-patterns/human-factors-contributing-to-fatal-injury (2011).

Koumbourlis AC: Electrical injuries. *Crit Care Med* 30:S424–S430, 2002.

Latenser BA, K.-V A: Pediatric burn rehabilitation (review). *Pediatr Rehabil* 5:2002.

Luan YY, Yao YM, Sheng ZY: Update on the immunological pathway of negative regulation in acute insults and sepsis. *J Interferon Cytokine Res* 32:288–298, 2012. doi:10.1089/jir.2011.0117.

Luce EA, Gottlieb SE: True" high-tension electrical injuries. *Ann Plast Surg* 12:321–326, 1984.

Lundy JB, et al: Update on severe burn management for the intensivist. *J Intensive Care Med* 2015. doi:10.1177/0885066615592346.

Mandell SP, Gibran NS: Early enteral nutrition for burn injury. *Adv Wound Care (New Rochelle)* 3:64–70, 2014. doi:10.1089/wound.2012.0382.

Matt SE, Shupp JW, Carter EA, et al: Comparing a single institution's experience with electrical injuries to the data recorded in the National Burn Repository. *J Burn Care Res* 33:606–611, 2012. doi:10.1097/BCR.0b013e318241b13d.

McClave SA, et al: Guidelines for the Provision and Assessment of Nutrition Support Therapy in the Adult Critically Ill Patient: Society of Critical Care Medicine (SCCM) and American Society for Parenteral and Enteral Nutrition (A.S.P.E.N.). *JPEN J Parenter Enteral Nutr* 40:159–211, 2016. doi:10.1177/0148607115621863.

Mendonca Machado N, Gragnani A, Masako Ferreira L: Burns, metabolism and nutritional requirements. *Nutr Hosp* 26:692–700, 2011. doi:10.1590/S0212-16112011000400005.

Miller AC, Rashid MR, Elamin EM: The "T" in trauma: the helper T-cell response and the role of immunomodulation in trauma and burn patients. *J Trauma* 63:1407–1417, 2007.

Moncrief JA, Mason AD: Water vapor loss in the burned patient. *Surg Forum* 13:38–41, 1962.

Muller MJ, Pegg SP, Rule MR: Determinants of death following burn injury. *Brit J Surg* 88:583–587, doi, 2001. doi:10.1046/j.1365-2168.2001.01726.x.

Murphy F, Amblum J: Treatment for burn blisters: debride or leave intact? *Emerg Nurse* 22:24–27, 2014. doi:10.7748/en2014.04.22.2.24.e1300.

Nagy EM, Juhasz I: Negative pressure wound therapy—an effective, minimally invasive therapeutic modality in burn wound management. *International Journal of Clinical Medicine* 6:301–306, 2015.

Namdar T, et al: Transdermal fluid loss in severely burned patients. *Ger Med Sci* 8:Doc28, 2010. doi:10.3205/000117.

Navickis RJ, Greenhalgh DG, Wilkes MM: Albumin in burn shock resuscitation: a meta-analysis of controlled clinical studies. *J Burn Care Res* 37:e268–e278, 2016. doi:10.1097/BCR.0000000000000201.

Nyame TT, Chiang HA, Orgill DP: Clinical applications of skin substitutes. *Surg Clin North Am* 94:839–850, 2014. doi:10.1016/j.suc.2014.05.013.

Ogden CL, Carroll MD, Kit BK, Flegal KM: Prevalence of childhood and adult obesity in the United States, 2011-2012. *JAMA* 311:806–814, 2014. doi:10.1001/jama.2014.732.

Ong YS, Samuel M, Song C: Meta-analysis of early excision of burns. *Burns* 32:145–150, 2006. doi:10.1016/j.burns.2005.09.005.

Roderique JD, Josef CS, Feldman MJ, Spiess BD: A modern literature review of carbon monoxide poisoning theories, therapies, and potential targets for therapy advancement. *Toxicology* 334:45–58, 2015. doi:10.1016/j.tox.2015.05.004.

Rodriguez NA, Jeschke MG, Williams FN, et al: Nutrition in burns: Galveston contributions. *JPEN J Parenter Enteral Nutr* 35:704–714, 2011. doi:10.1177/0148607111417446.

Rosanova MT, Stamboulian D, Lede R: Risk factors for mortality in burn children. *Braz J Infect Dis* 18:144–149, 2014. doi:10.1016/j.bjid.2013.08.004.

Rosenberg M, et al: Effects of a hospital based Wellness and Exercise program on quality of life of children with severe burns. *Burns* 39:599–609, 2013. doi:10.1016/j.burns.2012.08.019.

Rosenthal SM, Tabor H: Electrolyte changes and chemotherapy in experimental burn and traumatic shock and hemorrhage. *Arch Surg* 51:244–252, 1945.

Rousseau A, Massion PB, Langani A, et al: Toward targeted early burn care: lessons from a European Survey. *J Burn Care Res* 35:234–239, 2014.

Rowan MP, et al: Burn wound healing and treatment: review and advancements. *Crit Care* 19:243, 2015. doi:10.1186/s13054-015-0961-2.

Serghiou MA, et al: One world one burn rehabilitation standard. *Burns* 2016. doi:10.1016/j.burns.2016.04.002.

Service NW: *Lightning Safety*. Available at www.lightningsafety.noaa.gov/medical/htm (Accessed 15 November 2015).

Shakirov B: Treatment of acute deep burns in lower extremities of the elderly. *Journal of Acute Disease* 117–119, 2015.

Sheridan RL, Schnitzer JJ: Management of the high-risk pediatric burn patient. *J Pediatr Surg* 36:1308–1312, 2001. doi:10.1053/jpsu.2001.25805.

Sheridan RL: Burns. *Crit Care Med* 30:S500–S514, 2002. doi:10.1097/01.Ccm.0000034128.77577.30.

Shevchenko RV, James SL, James SE: A review of tissue-engineered skin bioconstructs available for skin reconstruction. *J R Soc Interface* 7:229–258, 2010. doi:10.1098/rsif.2009.0403.

Shrivastava P, Goel A: Pre-hospital care in burn injury. *Indian J Plast Surg* 43:S15–S22, 2010. doi:10.4103/0970-0358.70720.

Simpson F, Doig GS: Parenteral vs. enteral nutrition in the critically ill patient: a meta-analysis of trials using the intention to treat principle. *Intensive Care Med* 31:12–23, 2005. doi:10.1007/s00134-004-2511-2.

Snell JA, Loh NH, Mahambrey T, Shokrollahi K: Clinical review: the critical care management of the burn patient. *Crit Care* 17:241, 2013. doi:10.1186/cc12706.

Soleimani T, et al: Pediatric burns: Kids' Inpatient Database vs the National Burn Repository. *J Surg Res* 201:455–463, 2016. doi:10.1016/j.jss.2015.11.034.

Spoo J, Elsner P: Cement burns: a review 1960-2000. *Contact Dermatitis* 45:68–71, 2001.

Suhr M, Kreusch T: Burn injuries resulting from (accidental) airbag inflation. *J Craniomaxillofac Surg* 32:35–37, 2004.

Sun IF, Lee SS, Lin SD, Lai CS: Continuous arteriovenous hemodialysis and continuous venovenous hemofiltration in burn patients with acute renal failure. *Kaohsiung J Med Sci* 23:344–351, 2007. doi:10.1016/S1607-551X(09)70420-6.

Tarim A, Ezer A: Electrical burn is still a major risk factor for amputations. *Burns* 39:354–357, 2013. doi:10.1016/j.burns.2012.06.012.

Thomas SJ, et al: The effect of prolonged euglycemic hyperinsulinemia on lean body mass after severe burn. *Surgery* 132(2):341–347, 2002.

Turinsky J, Gonnerman WA, Loose LD: Impaired mineral metabolism in post-burn muscle. *Journal of Trauma* 21:417–423, 1981.

Turnage RH, et al: Mechanisms of pulmonary microvascular dysfunction during severe burn injury. *World J Surg* 26:848–853, 2002. doi:10.1007/s00268-002-4063-3.

USA SK: *Burns and Fire Safety Fact Sheet* http://www.safekids.org/sites/default/files/documents/skw_burns_fact_sheet_feb_2015.pdf (Accessed 16 May 2016) (2015).

Wang X, Han C: Re-emphasizing the role of copious water irrigation in the first aid treatment of chemical burns. *Burns* 40:779–780, 2014. doi:10.1016/j.burns.2013.12.029.

Wang X, et al: A review of treatment strategies for hydrofluoric acid burns: current status and future prospects. *Burns* 40:1447–1457, 2014. doi:10.1016/j.burns.2014.04.009.

Ward RS, et al: Prosthetic use in patients with burns and associated limb amputations. *J Burn Care Rehabil* 11:361–364, 1990.

Ward RS, et al: Sensory loss over grafted areas in patients with burns. *J Burn Care Rehabil* 10:536–538, 1989.

Welt LG: Membrane transport defect: the sick cell. *Trans Assoc Am Physicians* 80:1967.

White DJ, Maass DL, Sanders B, Horton JW: Cardiomyocyte intracellular calcium and cardiac dysfunction after burn trauma. *Crit Care Med* 30:14–22, 2002.

Wibbenmeyer LA, et al: Predicting survival in an elderly burn patient population. *Burns* 27:583–590, 2001.

Williams FN, Branski LK, Jeschke MG, Herndon DN: What, how, and how much should patients with burns be fed? *Surg Clin N Am* 91:609, 2011. doi:10.1016/j.suc.2011.03.002.

Williams FN, Herndon DN, Jeschke MG: The hypermetabolic response to burn injury and interventions to modify this response. *Clin Plast Surg* 36:583, 2009. doi:10.1016/j.cps.2009.05.001.

Williams RY, Wohlgemuth SD: Does the "rule of nines" apply to morbidly obese burn victims? *J Burn Care Res* 34:447–452, 2013. doi:10.1097/BCR.0b013e31827217bd.

Yuanhai Z, et al: Clinical arterial infusion of calcium gluconate: the preferred method for treating hydrofluoric acid burns of distal human limbs. *Int J Occup Med Environ Health* 27:104–113, 2014. doi:10.2478/s13382-014-0225-4.

Zhang Y, et al: The clinical effectiveness of the intravenous infusion of calcium gluconate for treatment of hydrofluoric acid burn of distal limbs. *Burns* 40:e26–e30, 2014. doi:10.1016/j.burns.2013.12.003.

GLOSSARY

A

A band A dark band corresponding to an area where actin and myosin filaments overlap in skeletal or cardiac muscle.

Abruptio placentae Premature separation of the placenta before delivery; the separation may be partial or complete and may result in overt or concealed hemorrhage.

Absolute anemia Anemia involving a decrease in the number of red blood cells (as opposed to a decrease in the percent of red blood cells).

Acalculous cholecystitis An important subgroup of acute cholecystitis. It tends to occur in the setting of major surgery, critical illness, trauma, or burn-related injury and does not occur in association with gallstones.

Accelerated (malignant) high blood pressure Rapidly progressing, potentially fatal form of hypertension in which the diastolic blood pressure exceeds 120 mm Hg.

Acclimatization A normal adaptive response to environmental changes, such as changes in altitude. For example, the red blood cell count increases when a person moves to a high altitude.

Accommodative capacity Ability of the eye to adjust to see objects at changing distances. This is a function of the ciliary muscle's ability to flatten or thicken the lens, thereby focusing the image on the retina.

Achalasia A disorder of esophageal smooth muscle function resulting in difficulty in swallowing both liquids and solids.

Acid A substance that releases hydrogen ions in solution and from which hydrogen may be displaced by a metal to form a salt. An increase in acid concentration produces a decrease in pH.

Acidemia The state in which the blood is overly acidic; usually defined as a pH <7.35.

Acidosis Presence of a condition that tends to make body fluids overly acidic.

Acne vulgaris A common disease of the skin in which sebaceous glands are numerous (face, upper back, and chest). Characteristic lesions include open (blackhead) and closed (whitehead) comedones, inflammatory papules, pustules, nodules, and cysts.

Acquired immunodeficiency syndrome (AIDS) A syndrome caused by the human immunodeficiency virus (HIV) in which the CD4 lymphocyte count is <200 cells/μL or an AIDS-indicator condition is present. Also called *stage 3 HIV disease.*

Acquired or secondary immunodeficiency An immunodeficiency that develops after birth and is the result of an illness rather than a genetic defect. Examples include impaired immune function secondary to poor nutrition or medication. This type of immunodeficiency may be reversible.

Acromegaly A chronic metabolic condition in adults characterized by excessive growth of bone, soft tissues, and organs due to abnormally high levels of growth hormone.

Acrosome Covering on the head of the sperm that contains large quantities of hydrolytic (water-splitting) enzymes that are released during capacitation.

Actin A cytoskeletal protein that comprises the thin filament of the muscle sarcomere in skeletal and cardiac muscle. It is also present in nonmuscle cells and is an important component of cell movement.

Actinic keratosis A horny premalignancy of skin epithelium caused by excessive exposure to sunlight.

Action potential An electrical impulse consisting of a self-propagating series of depolarizations and repolarizations transmitted across the plasma membranes of excitable cells that have voltage-gated ion channels, such as nerve and muscle.

Active immunity A form of long-term, acquired immunity that protects the body against a new infection as the result of antibodies that develop naturally after an initial infection or artificially after a vaccination.

Active transport pumps The membrane proteins that move polar substances across lipid membranes against a concentration gradient.

Acute Relatively severe but running a short course.

Acute coronary syndrome (ACS) Sudden onset of cardiac ischemia from occlusion of coronary vessels resulting in unstable angina or myocardial infarction.

Acute HAV infection A viral hepatitis infection caused by the hepatitis A virus (HAV) characterized by jaundice and fatigue.

Acute renal failure An abrupt reduction of renal function that is potentially reversible.

Acute rheumatic fever An inflammatory disease following a group A β-hemolytic streptococcal pharyngeal infection.

Acute tubular necrosis Sudden onset of renal dysfunction from death of nephron tubule cells (tubular epithelium), usually resulting from nephrotoxicity, ischemia after major surgery, trauma, severe hypovolemia, sepsis, or burns.

Acute viral hepatitis Inflammatory liver disease usually caused by hepatitis A virus, hepatitis B virus, and hepatitis C virus.

Adaptation An alteration in structure or function in response to a changed environment, which enhances or promotes survival.

Adapting Making an adjustment to a change in internal or external conditions or circumstances.

Addison disease Primary adrenocortical insufficiency thought to be autoimmune in etiology.

Adenosine triphosphate (ATP) A nucleoside with three phosphate groups and an adenine base; it functions as the principal source of energy in cells.

Adherence The quality of clinging or being closely attached. The degree to which a patient complies with therapeutic interventions such as taking medications.

Adhesive capsulitis A shoulder condition characterized by stiffness, pain, and limited range of motion.

Adrenocortical insufficiency Abnormally diminished secretion of corticosteroids by the adrenal cortex, as in Addison disease.

Adrenocorticotropic hormone (ACTH) A hormone that stimulates growth of the adrenal cortex and the synthesis and secretion of corticosteroids.

Affect The outward expression of emotion associated with a mental state or in response to a stimulus.

Affective disorders Mood disorders consisting of a variety of conditions characterized by a disturbance in mood.

Afferent neuron A neuron that transmits impulses from the periphery (sensory receptors) to the central nervous system.

Affinity The "tightness" of a ligand–receptor bond; the tendency of ligand and receptor to remain bound at low ligand concentrations.

Afterload The impedance or resistance that must be overcome in order to eject blood from a cardiac chamber. Systemic vascular resistance is the primary determinant of left ventricular afterload.

Age-related macular degeneration A progressive deterioration of the retina associated with abnormal retinal pigment epithelium that can progress to blindness.

Aggregates A total considered with reference to its constituent parts.

Agnosia Total or partial loss of the ability to recognize familiar objects or persons through sensory stimuli.

Agonist A substance with affinity for binding to receptors and mimicking the effect of the normal receptor–ligand interaction.

Agoraphobia Irrational fear of open spaces. In panic disorder, agoraphobia is a fear of any place or situation in which assistance would be unavailable in case of an unexpected panic attack. Agoraphobia is also known as *phobic avoidance*.

Airway resistance Relationship between pressure and flow of gas as determined by the radius of the airway.

Alagille syndrome Also called *arteriohepatic dysplasia;* this autosomal-dominant condition is associated with typical bony and vascular malformations and paucity of intrahepatic bile ducts.

Alarm The initial response to stress. The major features of the alarm reaction are attributable to activation of the sympathetic nervous system.

Albinism Partial or total absence of pigment in skin, hair, and eyes.

Alcohol abuse Overingestion of alcohol to the point of a person's being dependent on the substance.

Alcoholic fatty liver An accumulation of fat in the liver cells resulting from chronic alcohol consumption; also called *steatosis.*

Alcoholic hepatitis An active inflammation, especially of the centrilobular region of the liver, resulting from acute or chronic alcohol consumption.

Alcoholic liver disease Manifested by fatty liver, hepatitis, and cirrhosis. One or more of these manifestations may be found in alcoholic patients.

Aldosterone A mineralocorticoid synthesized by the adrenal cortex in response to angiotensin II that conserves sodium, producing increased water retention and consequently increased blood volume.

Alkalemia The state in which the blood is overly alkaline; usually defined as a pH >7.45.

Alkalosis Presence of a condition in which body fluids are overly alkaline.

Allele One of two or more alternative forms of a gene located at the same site on homologous chromosomes.

Allergic contact dermatitis Indicates delayed acquired hypersensitivity to a specific allergen on the skin. Chromates, nickel, ethylenediamine, *para*-phenylenediamine, neomycin, formaldehyde, and lanolin components may cause allergic contact dermatitis.

Allergic purpura A chronic disorder of the skin associated with urticaria, erythema, asthma, and rheumatic joint swellings. Platelet counts, bleeding times, and blood clotting are normal.

Allergy Type I hypersensitivity of the immune system to environmental agents. Antigens that trigger an allergic response are often called *allergens.*

Allodynia Perception of pain in response to normally nonpainful sensory stimuli.

Allogeneic Referring to transplanted tissue that was obtained from a closely matched donor, usually a sibling, parent, or child.

All-or-none response In a skeletal muscle, all of the muscle fibers innervated by a motor unit will respond as a single entity to its maximum or they will not contract at all. In a nerve, depolarization will result in either a full amplitude action potential or none at all.

Allostasis The process of achieving stability, or homeostasis, through physiologic or behavioral change.

Allostatic load A term coined as a more precise alternative to the term *stress;* used to refer to environmental challenges that cause an organism to begin efforts to maintain stability.

Alopecia Loss of hair, usually referring to the scalp.

α_1-Antitrypsin A plasma protein produced primarily in the liver; it is an acute-phase reactant that inhibits the activity of elastase, cathepsin G, trypsin, and other proteolytic enzymes.

Alveolar period The last stage in fetal lung development when alveolar ducts form from terminal sacs and alveoli mature by increasing both in size and in number.

Amblyopia Reduced vision in an eye not correctable by a fraction adjustment.

Amenorrhea Absence or suppression of menstrual bleeding, usually attributable to an altered pattern of hormonal functioning that interrupts the normal sequence of endometrial proliferation and sloughing.

Amniocentesis A procedure in which fluid is obtained from the amniotic cavity by an ultrasound-guided needle. The fluid contains fetal cells that can be used to screen for chromosomal and other defects.

Amniotic cavity The space between the amniotic sac and the developing embryo. It is filled with a clear amniotic fluid that keeps the embryo moist and provides a measure of protection against mechanical injury.

Amphipathic Having different characteristics. For example, membrane lipids are partly hydrophobic and partly hydrophilic, and hence are amphipathic. Also called *amphiphilic.*

Ampulla A flasklike cavity or dilatation of a tubular structure.

Amyloid plaque A microscopic lesion in the cerebral cortex composed of fragmented axon terminals and dendrites surrounding a core of β-amyloid, as found in Alzheimer disease.

Amyloid precursor protein A member of a large family of proteins that is associated with cell membranes and a precursor to β-amyloid, a component of brain plaques in Alzheimer disease.

Amyotrophic lateral sclerosis A progressive degenerative disease affecting both the upper and lower motor neurons characterized by muscle wasting and atrophy of the hands, arms, and legs; also called *Lou Gehrig disease.*

Anabolism The energy-requiring phase of metabolism through which molecules, cells, and tissues are created.

Anagen The growing phase of the hair cycle.

Anaphylactic shock A severe and sometimes fatal systemic allergic reaction to an allergen.

Anaplasia A lack of differentiated features in a tumor cell as evidenced by variations in cell size and shape and presence of abnormal nuclei.

Androgenic Producing masculine characteristics such as the androgenic hormone testosterone.

Anemia A decrease in the quantity of hemoglobin, hematocrit, and/or red blood cells.

Anergy Diminished immune responsiveness to antigens.

Aneuploidy An abnormal number of chromosomes—either too few (hypoploidy) or too many (hyperploidy, polyploidy).

Aneurysm Local dilation of an arterial wall or muscular chamber (e.g., cardiac ventricle).

Angina pectoris A paroxysmal chest pain most often due to cardiac ischemia associated with atherosclerotic coronary artery disease.

Angiogenesis The physiologic process involving the growth of new blood vessels from preexisting vessels.

Angiomyolipoma The most common benign tumor of the kidney; composed of blood vessels, smooth muscle cells, and fat cells.

Angiotensinogen A serum glycoprotein produced in the liver that is the precursor of angiotensin I.

Anhedonia Loss of interest in and withdrawal from all regular and pleasurable activities, often associated with depression.

Ankylosis The fusion of a joint, often in an abnormal position, usually resulting from destruction of articular cartilage and subchondral bone, as occurs in rheumatoid arthritis.

Ankylosing spondylitis An arthritis of the axial skeleton, including the sacroiliac joints, spine, hips, and shoulders. Marked limitation of motion develops, and a flexed spinal posture with flexed hips and knees may predominate.

Anorexia Loss of appetite.

Anorexia nervosa A refusal to eat or an aberration in eating patterns to the point of danger. The clinical syndrome may be due to an intense fear of becoming obese or to emotional states such as anxiety, irritation, or anger. Affected individuals become obsessed with the desire to become thin, and food intake is restricted even as weight falls well below the minimal normal value for age and height. Periods of fasting may alternate with periods of bingeing.

Anosognosia Lack of insight or denial of a neurologic defect, or illness in general (especially paralysis), on one side of the body.

Antagonist or blocking agent A substance that has an affinity to bind to a cellular receptor and blocks the activity of the normal receptor–ligand interaction.

Anthropometric Pertaining to measurements of the body or body parts such as height and weight for the purposes of understanding human physical variation.

Antibody Protein produced and secreted by B cells that binds to a specific antigen.

Anticipatory anxiety Anxious anticipation of an anxiety-provoking event.

Anticodon Sequence of three nucleotides in a transfer RNA molecule that is complementary to the messenger RNA codon.

Antidiuretic hormone (ADH) A posterior pituitary hormone that induces renal collecting duct cells to become permeable to water, thus decreasing the production of urine and reducing the osmolality of the blood. Also called *vasopressin.*

Antigen Macromolecule that provokes an immune system response.

Antimicrobial A chemical or agent that inhibits microbial activity.

Antisocial personality disorder A mental disorder characterized by failure to acquire the conditioned responses that are necessary for the learning of avoidance behaviors, conventional morality, and socialized positive responses to others. Also known as *antisocial reaction.*

Anuria Severe decrease or lack of urine output of less than 100 mL per day.

Anxiety disorders General group that comprises three major diagnoses: panic disorder, generalized anxiety disorder, and obsessive-compulsive disorder. Anxiety disorders are characterized by irrational fears and have great potential to cause disability in affected persons.

Aortic valve The cardiac valve that lies between the left ventricle and the aorta. It is open during ventricular systole and closed during ventricular diastole. Aortic valve closure contributes to heart sound S_2.

Aphasia A global disorder of language involving impaired speech (expressive aphasia) and impaired ability to understand the spoken word (receptive aphasia).

Aplastic anemia A deficiency of the formed elements of blood (specifically erythrocytes, leukocytes, and platelets) because of failure of the bone marrow to produce them.

Apocrine sweat gland A sweat gland that becomes functional only after puberty and discharges its products onto the skin through the hair follicle (hair pore).

Apoprotein A polypeptide chain not yet bound to its specific prosthetic group.

Apoptosis Programmed cell death, a process that requires cellular energy; characterized by DNA degradation and cell dissolution, but without necrosis.

Appendicitis Inflammation of the vermiform appendix due to an obstruction. This inflammation may lead to necrosis of the appendix, with subsequent abscess formation and peritonitis.

Apraxia An inability to execute previously learned skills, usually after a stroke.

Arachnoidal villi Fingerlike projections in the delicate membrane between the dura mater and the pia mater of the brain.

Archaea A group of single-celled microorganisms, distinct from bacteria.

ARDS Acute respiratory distress syndrome; severe pulmonary dysfunction characterized by diffuse inflammatory injury to alveolar–capillary membranes.

Arnold–Chiari II malformation A congenital anomaly associated with meningomyelocele and hydrocephalus in which the cerebellum and medulla oblongata protrude into the cervical spinal canal through the foramen magnum.

Arterial pulse pressure The difference between systolic and diastolic blood pressures.

Arteriosclerosis Generalized term for pathologic conditions resulting in decreased distensibility of arteries; also known as *hardening of the arteries*.

Arteriovenous fistula Abnormal communication between an artery and a vein.

Arteriovenous malformation (AVM) A congenital disorder of the connections between veins and arteries that have no intervening capillary between them.

Arteritis Inflammation of an artery. May be associated with an autoimmune reaction.

Arthralgia Joint pain.

Arthritis Any inflammatory condition of the joints; characterized by pain, swelling, heat, redness, and limitation of movement.

Articular (hyaline) cartilage Connective tissue that forms a smooth, resilient, low-friction surface for articulation of two bones. It is without nerves, is avascular in adults, and derives nourishment from synovial fluid. It tolerates extreme compression stress.

Articulation A point of contact between bones. Also called *joint.*

Ascites Abnormal accumulation of fluid in the peritoneal cavity. Causes include liver disease, heart failure, constrictive pericarditis, infection, malnutrition, pancreatitis, lymphatic obstruction or leakage, renal disease, hypothyroidism, collagen vascular diseases, and malignancy.

Aspiration Inadvertent entry of food substances, liquids, or gastric contents into the respiratory system. This potentially life-threatening occurrence is normally prevented by the coordinated set of actions performed by the muscles in the pharynx during swallowing.

Asthenia The lack or loss of strength or energy; weakness.

Asthma A respiratory condition characterized by increased responsiveness of the trachea and bronchi to various stimuli and manifested by widespread narrowing of the airways and inflammation.

Astigmatism An abnormal condition of the eye in which the light rays cannot be focused clearly in a point on the retina because the spherical curve of the cornea or lens is not equal in all meridians. Vision is typically blurred.

Ataxia Failure of muscular coordination, resulting in incoordination and disturbances in posture and gait.

Atelectasis Full or partial collapse of the lung alveoli.

Atherosclerosis A type of arteriosclerosis characterized by proliferation of smooth muscle cells and lipid collection within the walls of arteries, resulting in narrowed lumina and impaired ability to dilate.

Atopic Pertaining to a hereditary tendency to experience immediate allergic reactions because of the presence of an antibody.

Atopic dermatitis An intensely pruritic, often excoriated inflammation of skin in allergy-prone individuals.

Atopy A genetic predisposition to allergies.

Atresia Congenital failure to develop (absence) or abnormal closure of a normally open passage.

Atrial fibrillation A completely disorganized and irregular atrial rhythm accompanied by an irregular ventricular rhythm of variable rate.

Atrophy A reduction in size and function of a cell or tissue; wasting.

Attention-deficit/hyperactivity disorder (ADHD) A mental disorder involving impaired or diminished attention, impulsivity, and hyperactivity.

Aura A peculiar sensation preceding the appearance of more definite symptoms, as in migraines and seizures.

Auscultatory gap The time during cuff deflation after systolic blood pressure when the Korotkoff sounds disappear and then reappear.

Autism A mental disorder primarily characterized by abnormal development of social interaction and communicative skills. Affected individuals may manifest an inability to perceive or understand others' feelings or to express their own feelings and may adhere to rigid, nonfunctional behaviors or rituals.

Autocrine Relating to hormonelike chemicals in which the target cell is the same cell that secretes the chemical.

Autocrine signaling The secretion of factors that feed back onto the cell that secreted them. Usually used in reference to growth factors.

Autografting Surgical procedure to move skin from one area of the body to an area of injury. The purpose is to provide permanent skin coverage to the injured area.

Autoimmune liver disease Hepatic injury from self-reactive antibodies produced by errant B lymphocytes.

Autoimmunity An inappropriate and excessive response of the immune system to self antigens causing disease. Disorders that result from an autoimmune response are called *autoimmune diseases.*

Autologous Pertaining to a tissue or structure occurring naturally and derived from the same individual, such as blood donated by a patient before surgery to be returned to the patient.

Automaticity A property of specialized excitable tissue that allows self-activation through spontaneous development of an action potential, as in the pacemaker cells of the heart.

Autonomic dysreflexia Hyperreflexia; an uninhibited and exaggerated reflex of the autonomic nervous system in response to stimulation in patients with spinal cord injuries.

Autoregulation The intrinsic tendency of an organ or tissue to maintain adequate blood flow despite changes in metabolism or blood pressure.

Autosomal-dominant polycystic kidney disease Hereditary disorder associated with defects on chromosome 16 (95% of cases) or chromosome 4 (5% of cases), resulting in dilation of, and cyst formation in, collecting ducts and impaired renal function.

Autosomal-recessive polycystic kidney disease Congenital disorder linked to a defect on chromosome 6 that results in dilations of the renal collecting ducts and hepatic fibrosis.

Autosome Any ordinary paired chromosome (1 through 22), as distinguished from a sex chromosome.

Avoidance Refers to conscious or subconscious defensive reactions used to increase feelings of control and decrease the risk of anxiety.

Avulsion fracture A separation of a small fragment of bone at the site of attachment of a ligament or tendon.

Axoneme 1. Central core of a cilium or flagellum, consisting of two central fibrils surrounded by nine peripheral fibrils. 2. Motor apparatus of the sperm's tail.

Azotemia Increased levels of nitrogenous waste products, especially urea nitrogen, in the blood indicative of impaired renal clearance.

B

B cell A type of lymphocyte that has receptors for binding antigen and can present antigens to T cells; responsible for humoral immunity. B cells that are actively producing antibodies are called *plasma cells.*

Bacillus A genus of aerobic or facultatively anaerobic, gram-positive, spore-bearing, rod-shaped bacteria that may or may not be motile.

Bacteria A domain of life existing as small unicellular microorganisms without nuclei, called *prokaryotic.*

Bacterial enzyme An enzyme that aids in the microorganism's ability to spread or invade tissues; examples include fibrinolysin, coagulase, and hyaluronidase.

Ball-and-socket joint Formed by a ball-like surface fitting into a concave socket. Ball-and-socket joints permit flexion-extension, adduction-abduction, and rotational movements, such as those of the hip and shoulder.

Baroreceptors One of the pressure-sensitive sensory nerve endings in the walls of the atria of the heart, the aortic arch, and the carotid sinuses.

Barrett esophagus A complication of chronic gastroesophageal reflux disease that represents replacement of the normal squamous epithelium of the distal esophagus by columnar tissue. Considered to be a preneoplastic condition.

Bartholinitis An inflammatory condition of one or both Bartholin glands; caused by bacterial infection.

Basal energy expenditure A term used to describe the calculated basal metabolic rate—the metabolic rate at rest.

Basal ganglia Groups of cell bodies (nuclei) located deep within the cerebral hemispheres that help plan and execute motor activities, including the caudate, putamen, globus pallidus, substantia nigra, and subthalamus.

Basal metabolic rate The amount of energy required for an individual to maintain vital processes such as respiration, digestion, and circulation at rest.

Base 1. The nonacid part of a salt. 2. A substance that accepts hydrogen ions in solution to form salts and increases pH.

Basophil/basophilic granulocyte A leukocyte that is functionally and chemically related to the mast cell; it has a kidney-shaped nucleus and large, deep basophilic granules, which contain vasoactive amine and heparin and are important in IgE binding.

Beau line Transverse furrow in the nail that indicates a disturbance in nail growth.

Becker dystrophy A milder form of inherited muscle degeneration than the Duchenne type and somewhat less common, with an annual incidence of 5 per 100,000. The genetic mutation leads to production of a reduced amount of an abnormal dystrophin protein and slower muscular degeneration.

Bell palsy An acute idiopathic unilateral paresis or paralysis of the facial nerve (cranial nerve VII) involving an inflammatory reaction at or near the stylomastoid foramen or in the bony facial canal.

Bence Jones proteins Proteins found in the urine of patients with plasma cell (multiple) myeloma. They are derived from overproduction of light chain fragments of antibodies by malignant plasma cells. Bence Jones proteins are nephrotoxic and may contribute to the development of kidney disease.

Benign breast disorders A group of lesions affecting the breast, which are usually divided into two categories: fibrocystic breast disease and benign neoplasms of the breast.

Benign prostatic hyperplasia or hypertrophy (BPH) A noncancerous enlargement of the prostate gland.

Benign tumor A type of tumor that is strictly local, is usually well differentiated, and does not metastasize.

β-Amyloid Protein fragment snipped from a larger molecule—called *amyloid precursor protein*—during metabolism. Abnormal β-amyloid is a component of neuritic plaques found in Alzheimer disease.

Biaxial joint A joint that has two axes of movement and permits movement in two planes.

Bile A substance produced by hepatocytes in the liver and stored in the gallbladder. It is composed primarily of water, electrolytes, bile salts, cholesterol, and phospholipids. The major functions of bile are to aid in the digestion of dietary lipids through emulsification and to transport waste products, particularly bilirubin, into the intestine for disposal or reabsorption.

Biliary atresia Also called *extrahepatic ductopenia* or *progressive obliterative cholangiopathy;* biliary atresia can be either congenital or acquired. The latter occurs in the setting of certain autoimmune illnesses and is one of the principal forms of chronic rejection of a transplanted liver allograft. Biliary atresia is a rather common birth defect, occurring in 1 of 10,000 to 15,000 live births.

Biliary cirrhosis A disease initiated by damage to the bile ducts, which may be due to macroscopic or microscopic biliary obstruction. Persistent biliary obstruction results in inflammation and scarring of the liver, with obliteration of the bile ductules.

Biliary colic Persistent epigastric pain related to intermittent obstruction of the cystic duct, usually by a gallstone. A typical episode lasts several hours.

Bilirubin A substance formed from the degradation of erythrocytic hemoglobin (porphyrin component) by reticuloendothelial cells.

Biliverdin A greenish bile pigment formed in the breakdown of hemoglobin and converted to bilirubin.

Bipennate Pertaining to a muscle with a central tendon toward which the fibers converge on either side like the barbs of a feather.

Bipolar disorder A mood disorder characterized by alternating periods of mania and depression.

Bladder calculus A solid mass (stone) formed from debris within the bladder.

Blast An immature precursor form of a lymphoid or myeloid white blood cell. Blasts are not normally found in the peripheral blood because they are retained in the marrow until mature. The presence of blasts in the peripheral blood indicates leukemia.

Blood pressure The pressure exerted by the circulating volume of blood on the walls of the arteries.

Blood urea nitrogen (BUN) Urea is an end product of amino acid metabolism, measured in the blood as BUN and excreted primarily by the kidney.

Blunted affect A severe reduction in the intensity of externalized feelings.

Body fluid The water contained in the body plus the substances dissolved in it.

Body mass index (BMI) A weight reference standard. The formula for BMI is weight (kg) divided by height squared (m^2).

Body water All of the water contained in the body.

Bolus 1. A round mass of food that has been softened and formed into an appropriate size for swallowing by the action of chewing. 2. A concentrated mass of pharmaceutical preparation or fluid generally administered over a short period.

Bone and joint tuberculosis An extrapulmonary form of tuberculosis that occurs after lymphohematogenous spread from a primary lung lesion.

Bone marrow suppression Suppression of bone marrow activity, resulting in reduction in the number of platelets, red blood cells, and white blood cells, such as in aplastic anemia. Also called *myelosuppression*.

Borderline personality disorder Personality disorder that represents a pervasive and persistent disturbance in ways of handling events and situations. Personalities influenced by this disorder are unstable, unpredictable, impulsive, and often moody and self-deprecating. Some overlap with depression is suggested.

Brainstem Portion of the brain consisting of the midbrain, pons, and medulla oblongata and mesencephalon.

Branched-chain amino acids A group of amino acids that includes valine, leucine, and isoleucine; they are mainly metabolized in the muscle for energy.

Bronchiectasis A disorder characterized by destruction of the elastic and muscular structures; results in dilation of the bronchi.

Bronchiolitis Inflammation of small bronchi.

Bronchitis Widespread inflammation of bronchi and bronchioles attributable to infectious agents or allergic reactions.

Bronchospasm Narrowing of the bronchi and bronchioles because of abnormal contraction of the smooth muscles of the bronchial walls.

Bruit Sound generated by turbulent blood flow auscultated over a blood vessel.

Brush border Covering of the microvilli projecting from some types of epithelial cells, such as proximal renal tubule cells and the intestinal villi. This fuzzy coating contains many enzymes and transporters.

Buck fascia or fascia of Buck Thick fibrous envelope surrounding the tunica albuginea, which encloses each of the erectile bodies of the penis.

Buckle fracture A fracture in children whereby the bone buckles and eventually cracks as a result of a compression injury to cancellous bone of the metaphysis of a long bone.

Buffer A chemical that releases hydrogen ions when a fluid is too alkaline and takes up hydrogen ions when a fluid is too acidic.

Bulbourethral glands Also called *Cowper glands,* these two glands produce viscous fluid that is secreted into the urethra near the base of the penis.

Bulbous urethra The proximal portion of the penile urethra. The bulbous urethra is surrounded by the bulb of the urethra and the bulbospongiosus muscle.

Bulimia nervosa Recurrent episodes of binge eating followed by self-induced vomiting or diarrhea, excessive exercise, strict dieting, or fasting; person has an exaggerated concern about body shape and weight.

Bulla Large, thin-walled cyst. Commonly used in reference to lung or skin.

Bursa Pocket of connective tissue lined with liquid-containing synovium; located between muscles or between muscle or tendon and bone.

Byler syndrome A rare autosomal-recessive disorder involving severe jaundice, pruritus, and malabsorption caused by an error in bile salt metabolism. Also called *progressive intrahepatic*

cholestasis and *progressive familial intrahepatic cholestasis.*

C

Cachexia A combination of symptoms, including anorexia, weight loss, muscle wasting, and weakness, that is associated with the severe malnutrition of chronic diseases such as cancer.

Calcitonin A hormone produced by thyroid parafollicular cells, it influences the processing of calcium by bone cells.

Calculus A mass of solid mineral or metabolic substance. A stone.

Callus (bone) The bony deposit formed between and around the broken ends of a fractured bone during healing. Also called *keratoma*.

Calluses (skin) Common, usually painless thickenings of the stratum corneum at locations of external pressure or friction.

Cancellous bone Bone with a spongy or latticelike appearance; found in the interior of bones. Cancellous bone does not tolerate compression stress.

Cancer cachexia The severe nutritional effects of cancer. See *cachexia*.

Capacitation The multiple changes that activate sperm and enhance their ability to participate in the final process of fertilization.

Capillary hydrostatic pressure The outward push of the vascular fluid against the capillary walls that is caused by blood pressure.

Capillary osmotic pressure The inward pull of particles in the vascular fluid from dissolved proteins in the blood; also called *oncotic pressure.*

Carbohydrates The main energy source for the body; consists of simple or complex sugars. They must be supplied in a fairly constant manner to meet the energy requirements for normal body functioning. Provides 4 kcal/g of energy when metabolized.

Carbonic anhydrase The enzyme that catalyzes the reversible conversion of carbon dioxide and water to carbonic acid.

Carcinogen A substance that initiates or promotes the development of cancers. Most carcinogens cause cancer by damaging DNA to produce mutations.

Carcinoma in situ A premalignant neoplasm that has not invaded the basement membrane but shows cytologic characteristics of cancer.

Cardiac asthma Results from bronchospasm precipitated by congestive heart failure.

Cardiac catheterization A diagnostic procedure in which a catheter is introduced through an incision into a large vein or artery (cardiac angiography) and threaded through the circulatory system to the heart.

Cardiac cycle A cardiac cycle includes one diastolic and one systolic phase.

Cardiac index A measure of the heart's pumping ability, taking into account body surface area. The cardiac index is calculated by dividing cardiac output by body surface area. A cardiac index less than 2.0 L/min/m^2 is considered to be insufficient for adequate peripheral perfusion.

Cardiac output A measure of the amount of blood pumped by the heart in 1 minute; usually expressed in liters per minute.

Cardiac tamponade Abnormal external pressure on the heart that results in poor cardiac filling and decreased cardiac output.

Cardiogenic shock A condition of low cardiac output and inadequate perfusion of tissues associated with acute myocardial infarction and congestive heart failure.

Cardiomyopathy Diseases that primarily affect myocardial cells, often of unknown cause. Three common types of cardiomyopathy are dilated, hypertrophic, and restrictive.

Carina A ridgelike structure at the base of the trachea that projects from the area that separates the left and right bronchi.

Carpal joint A synovial joint between the carpal bones of the hands.

Carrier A person who harbors a recessive gene for a particular trait. A recessive heterozygote.

Carrier proteins Proteins located in lipid bilayers that transport ions and small molecules through the membrane by first binding on one side and then moving to the other side by changing conformation.

Cartilaginous joint A joint that connects bony segments by fibrocartilage or hyaline growth cartilage.

Casts Protein meshes that collect in a nephron tubule and conform to the shape of the tubule; when the casts contain white blood cells (WBCs) or red blood cells (RBCs), their presence indicates infection or inflammation of the kidney.

Catabolism The process of converting large molecules of carbohydrate, protein, and fat to smaller molecules to be utilized for energy.

Cataracts An abnormal progressive condition of the lens of the eye that is characterized by loss of transparency.

Catecholamine A hormone (e.g., epinephrine and norepinephrine) that stimulates glycogenolysis and gluconeogenesis. An amine neurotransmitter (e.g., norepinephrine, dopamine).

Catecholamine hypothesis A hypothesis that abnormally low catecholaminergic neurotransmission leads to depression and abnormally high catecholaminergic neurotransmission leads to mania.

Caudal Signifying a position toward the distal end of the body, or an inferior position.

Caudate nuclei Portion of each cerebral hemisphere that, together with the lentiform nuclei, forms the corpus striatum of the basal ganglia.

Celiac disease Also called *celiac sprue,* this disease is characterized by intolerance of gluten, a protein in wheat and wheat products; ingestion of gluten causes abdominal distention and malabsorption.

Cell cycle The phases through which a cell progresses during cellular reproduction, including gap 1, synthesis, gap 2, and mitosis.

Central venous pressure The blood pressure in the large veins of the body measured in the right atrium of the heart.

Centromere Constricted region that holds two sister chromatids together. The centromere is the site of attachment to the microtubules, which pull the chromatids apart during mitosis.

Centrosome A centrally located organelle that organizes microtubules in the cell. It acts as the spindle pole during mitosis.

Cerebellum Portion of the brain attached to the brainstem; it has an essential role in maintaining muscle tone and posture and coordinating normal movements.

Cerebral dysrhythmia An abnormality in an otherwise normal rhythmic pattern, as seen on electroencephalography.

Cerebral edema An accumulation of fluid in the brain tissues. Causes include infection, tumor, trauma, or exposure to certain toxins. Because the skull cannot expand to accommodate the fluid pressure, brain tissues are compressed.

Cerebral palsy Refers to a diverse group of crippling syndromes that appears during childhood and involves permanent, nonprogressive encephalopathic damage to the developing brain.

Cerebrospinal fluid Fluid found in the cavities and canals of the brain and spinal cord.

Cerebrotendinous xanthomatosis A steroid hydroxylase deficiency that leads to premature atherosclerosis and encephalopathy. Also known as *van Bogaert disease*.

Cerebrovascular accident (CVA) An abnormal condition of the brain characterized by occlusion by an embolus, thrombus, hemorrhage, or vasospasm, resulting in ischemia of the brain tissues. Also called *stroke,* or *brain attack*.

Cerebrum Portion of the brain that controls consciousness, memory, sensations, emotions, and voluntary movements. The largest part of the brain, it consists of two hemispheres.

Ceruminous gland A special variety or modification of an apocrine sweat gland. The mixed secretions of sebaceous and ceruminous glands form a brown waxy substance called *cerumen,* which protects the ear canal from dehydration.

Chagas disease Caused by *Trypanosoma cruzi* and transmitted to humans by bloodsucking insects. It is a common cause of acquired myocarditis and megacolon in Central and South America but is rarely seen in the United States.

Chancre Painless, ulcerative lesion arising at the original port of entry of the spirochete that causes syphilis.

Chancroid An ulcerative, infectious disease of the genital tract caused by the sexually transmitted bacillus *Haemophilus ducreyi.* Unlike the chancre in syphilis, the lesion in chancroid is painful, tender, and often multiple.

Channel proteins Proteins located in lipid bilayers; they form porelike structures that allow ions to pass through by diffusion when appropriately stimulated.

Chemodissolution Use of chemical substances, such as bile acids or organic solvents, to dissolve gallstones. Used as a nonoperative method to treat gallstones.

Chemokines Any of a group of low-molecular-weight cytokines. They function as regulators of the immune system that help immune cells localize to specific areas.

Chemotaxis The movement of cells according to chemical gradients (chemotaxins) that attract them.

Chemotherapy The treatment of cancer, infections, and other diseases with chemical agents. Chemotherapeutic agents are often cytotoxic.

Chest physiotherapy Use of percussion and postural drainage to mobilize secretions from specific segments of the lungs.

Chickenpox Also called *varicella.* Chickenpox is a common communicable childhood disease. It is caused by the varicella zoster virus, which is also the causative agent in shingles. The characteristic skin lesion occurs in three stages: macule, vesicle, and granular scab.

Chlamydia Genus of a microorganism that lives as an intracellular bacterium. *Chlamydia trachomatis* inhabits the epithelium of the urethra and cervix and is responsible for the highly contagious systemic infection *lymphogranuloma venereum*.

Chloride shift An exchange of chloride ions for bicarbonate ions (HCO_3^-) in red blood cells in peripheral tissues as a response to changes in the P_{CO_2} of blood.

Cholecalciferol Precursor substance of active vitamin D.

Cholecystectomy Surgical removal of the gallbladder.

Cholecystitis Inflammation of the gallbladder wall; may be acute or chronic and usually is associated with cholelithiasis.

Cholecystokinin Hormone secreted from the small intestinal mucosa; two of its chief functions are stimulation of the release of pancreatic enzymes during a meal and contraction of the gallbladder.

Cholelithiasis Formation of stones in the gallbladder.

Cholesterol A lipid-soluble compound that facilitates the absorption and transport of fatty acids in bile and provides the building blocks for steroid hormone production.

Cholinergic-noradrenergic imbalance hypothesis A hypothesis that suggests that a relative increase in the ratio of acetylcholine activity to norepinephrine activity produces depression and that mania is the result of a relative increase in the ratio of norepinephrine activity to acetylcholine activity.

Chondroblasts Any one of the cells that develop from the mesenchyme and form cartilage. They play an important role in endochondral ossification and especially in longitudinal bone growth.

Chondroma Also called *enchondroma;* a cartilage-forming tumor located within bone that accounts for about 15% of benign bone tumors.

Chondrosarcoma A malignant cartilage-forming tumor; chondrosarcomas tend to develop in the pelvic and shoulder girdles and the proximal ends of long bones.

Chordae tendineae Bands of fibrous connective tissue that anchor the atrioventricular valves to the papillary muscles of the ventricular chambers.

Chorionic villus sampling A procedure in which tissue is obtained from the placenta by ultrasound-guided biopsy. Chorionic villus sampling can be performed earlier in pregnancy (10 to 12 weeks) than amniocentesis (16 weeks).

Chromatid One copy of a chromosome formed by DNA replication that may be joined to the other copy (sister chromatid) at the centromere.

Chromosome A linear thread of nuclear DNA that becomes visible under the microscope during cell mitosis.

Chronic Refers to a condition that lasts for a long time, generally from 6 months to years.

Chronic active hepatitis A progressive, destructive inflammatory disease that extends beyond the portal triad to the hepatic lobule (piecemeal necrosis).

Chronic bronchitis A condition characterized by excessive secretion of bronchial mucus and manifested by productive cough for 3 or more months in at least 2 consecutive years in the absence of any other disease process that may cause this symptom. A component of chronic obstructive pulmonary disease (COPD).

Chronic hepatitis Ongoing inflammation of the liver, usually of more than 6 months' duration, after viral hepatitis or attributable to autoimmune disease.

Chronic persistent hepatitis Also called *triaditis* or *transaminitis.* A benign disease in which the inflammation is confined to the portal triads without destruction of normal liver functions despite elevated serum transaminase levels.

Chronic renal failure Gradual loss of renal function that is progressive and irreversible.

Chronic venous insufficiency Varicosity of the deep veins that prevents effective return of blood from the periphery. Usually manifests as edema.

Chylothorax An accumulation of chylous fluid attributable to leakage of chyle (lymph fluid) from the thoracic duct or to rheumatoid pleural effusion or tuberculous pleuritis. Also called *chylous pleural effusion.*

Chyme Viscous, semifluid contents of the stomach following the mixture of ingested nutrients with gastric secretions. Chyme then passes through the pylorus into the duodenum, where further digestion occurs.

Cilia Motile hairlike processes on the surface of some cells.

Circadian rhythm The regular recurrence of certain biological phenomena in approximately 24-hour cycles, regardless of constant darkness or other conditions of illumination.

Circumferential burn A burn injury that wraps completely around an extremity or the trunk. Loss of elasticity of skin results in a tourniquet effect, compromising circulation to distal tissues or respiratory expansion of the chest. Escharotomy or fasciotomy is necessary.

Cirrhosis A diffuse, irreversible scarring of the liver resulting in abnormal nodules of liver cells surrounded by fibrosis.

Citric acid cycle A sequence of enzymatic reactions in the mitochondrial matrix that produces carbon dioxide and high-energy electrons from acetyl coenzyme A. Also called *tricarboxylic acid cycle* and *Kreb cycle.*

Clang association Association of words similar in sound but not in meaning, or words having no logical connection; may include rhyming and punning.

Clinical dehydration The combination of extracellular fluid volume deficit and hypernatremia.

Clinical manifestations The functional consequences of the structural and associated alterations in cells or tissues that are either characteristic of the disease or diagnostic of the process.

Clonic Characterized by alternating periods of involuntary muscular contraction and relaxation in rapid succession.

Closed fracture A type of fracture that occurs when fragments of a fracture do not extend through mucous membranes or skin and skin is not broken.

Closing volume Lung volume at which airways in the lower lung zones collapse and ventilation ceases.

Clotting factors Proteins that circulate in inactive forms and can be triggered to initiate a clotting cascade to produce insoluble fibrin clots.

Clubbing A process characterized by flattening of the angle of the base of the nail. It may occur in association with cardiovascular disease, subacute bacterial endocarditis, and pulmonary disease.

CO_2 Carbon dioxide; this gas is produced by cells during metabolism, is carried in the blood as carbonic acid, and is excreted by the lungs.

Coagulation The process of blood clot formation.

Coagulopathy An abnormality in blood clot formation.

Cocci Round, nonmotile bacteria.

Codon Sequence of three nucleotides in DNA or messenger RNA that represents the instruction for a particular amino acid in a polypeptide chain.

Collagen Most abundant protein in the body. The major protein of the white fibers of connective tissue. Has tensile strength similar to that of steel and is responsible for functional integrity of connective tissue.

Colloid osmotic pressure Pressure produced by passage of fluid from an area of less concentration to an area of higher concentration of colloids (large charged molecules such as proteins).

Colonic or anal agenesis Imperforate anus.

Colonization Harmless inhabitation of the skin or mucous membranes by microorganisms.

Colostomy Establishment of an artificial opening of the colon on the abdominal wall; usually performed after removal of a diseased or injured bowel segment.

Comminuted fracture A fracture consisting of more than one fracture line and more than two bone fragments.

Compact bone Hard, dense bone that is usually found at the periphery of skeletal structures.

Compartment syndrome A syndrome resulting from trauma to soft tissue caused by swelling within the unyielding structure of a nonelastic tissue or device (e.g., a cast).

Compensation The counterbalancing of any defect of structure or function. For example, a process that tends to restore pH to normal by making other blood chemistry values abnormal.

Complement A protein that participates in a cascade of reactions resulting in inflammation and cell lysis. Complement activation can occur by the classical or the alternative pathways.

Complete fracture A fracture whose line disrupts bone continuity through the whole thickness of the bone, including the cortex.

Compliance A measure of the ease of elastic distensibility of a hollow organ.

Complication A new or separate process that may arise secondarily because of some change produced by the original entity. For example,

bacterial pneumonia may be a complication of viral infection of the respiratory tract.

Compression fracture Consistent with cancellous bone trauma. Also called a *crush fracture.*

Compulsion Repetitive ritualistic behavior that has a driven quality.

Concentric contraction The shortening contraction of a muscle when the muscle force generates sufficient tension to overcome the resistance of the limb. One example is lifting a cup of water to one's mouth.

Condyloid joint A joint that permits flexion and extension at one axis and adduction and abduction around another axis, such as the metacarpophalangeal joint of the hand.

Condyloma acuminatum; condylomata acuminata Genital wart(s) caused by papillomavirus forms.

Conformational change A movement or alteration in the three-dimensional formation of a protein without any change in amino acid structure.

Congenital adrenal hyperplasia Overproduction of adrenal androgens attributable to a lack of an enzyme needed for cortisol production. Symptoms include virilization of the female infant's genitalia.

Congenital ichthyosis An inherited disease characterized by an excessive growth of keratinocytes and keratin, which gives the skin a fish-scale appearance.

Congenital immunodeficiency Rare condition that results from improper development of immune system components before birth.

Congenital malformation A general term meaning a defect in form or function that is present at birth.

Congestive heart failure Dysfunctional cardiac pumping that results in congestion of blood behind the dysfunctional cardiac pump. Right-sided heart failure is associated with systemic venous congestion. Left-sided heart failure is associated with pulmonary congestion.

Consanguinity Mating of blood-related individuals.

Consolidation The process of tissues becoming firm and solid, such as when the lung alveoli become firm while air spaces are filled with exudate in pneumonia.

Constipation A condition of having small, infrequent, and difficult bowel movements. Authorities have established a norm of fewer than three stools per week as a guideline for defining constipation.

Contact dermatitis A cutaneous reaction to topical irritation or allergy. Irritant contact dermatitis can develop in any person exposed to a sufficiently high concentration of the irritating agent. Some of the more active irritants are acids, alkalis, and hydrocarbons.

Contact hypersensitivity Allergy to a substance that produced a reaction in a previous contact. Usually occurs on the skin and may take several hours to develop (type IV hypersensitivity).

Continent The ability to control bladder or bowel function.

Contractile tissue Tissues involved in the contraction of muscle, including not only the muscle belly but also the tendon and bony insertion.

Contractility The force and velocity of cardiac muscle shortening in response to stimuli that increase cytoplasmic free calcium ion levels.

Contraction time The time from initial tension development to peak tension.

Contralateral Referring to the opposite side of the body.

Convalescence The stage of recovery after a disease, injury, or surgical operation.

Coombs antiglobulin test The direct test is an assay for antibody that is attached to red blood cells; the indirect test is an assay for antibody circulating in serum.

Coping A measure of the individual's resourcefulness and ability to deal with stress and stressors.

Cor pulmonale Right ventricular hypertrophy secondary to pulmonary diseases that increase right ventricular afterload.

Corns Horny masses of condensed epithelial cells overlying bony prominences. Corns result from chronic friction and pressure.

Coronary angiography Radiographic visualization of the internal anatomy of the heart and blood vessels with the use of intravascular introduction of radiopaque contrast medium.

Corpora cavernosa Two paired erectile bodies that lie dorsally in the penis.

Corpus luteum Anatomic structure on the surface of the ovary that grows in the ruptured ovarian follicle after ovulation and acts as a temporary endocrine organ that secretes progesterone.

Corpus spongiosum Erectile body in the penis containing the urethra.

Correction A process whereby normal values are restored when the underlying cause is addressed. An example would be restoring pH to normal by addressing the underlying cause of an acid–base imbalance.

Cortical bone The dense cortex or outer shell of bone; designed to tolerate compression and shearing forces.

Corticosteroid A hormone produced by the adrenal gland that stimulates gluconeogenesis and contributes to insulin resistance (e.g., cortisol); or a drug that has similar effects.

Cortisol A glucocorticoid (steroid hormone) released by the adrenal gland that causes an increase in blood glucose level by promoting liver gluconeogenesis.

Coryza A head cold with profuse nasal drainage.

Costovertebral angle Area lateral to the sacrospinalis muscle and beneath the twelfth rib used as an external landmark for the kidneys.

Cowper glands Also called *bulbourethral glands,* these two glands produce viscous fluid that is secreted into the urethra near the base of the penis.

Crackles Rales (pronounced "rahls"); discontinuous fine crackling sounds, usually heard on inspiration, that are indicative of air moving through fluid.

Cradle cap A seborrheic condition in infants characterized by scaling of the scalp. Occurs as a result of infrequent or inadequate washing of the scalp.

Cranial nerve reflexes The 12 pairs of nerves emerging from the cranial cavity that carry impulses for such functions as the senses.

Creatine kinase An enzyme that catalyzes the transfer of a phosphate group between adenosine triphosphate and creatine. The isoenzyme found in cardiac muscle is called *CK-MB* (the MB fraction of creatine kinase).

Creatinine End product of muscle metabolism that is filtered freely through the glomeruli and excreted by the kidney only. Creatinine clearance is used as a measure of glomerular filtration rate.

Cretinism Extreme hypothyroidism during infancy and childhood that causes mental and physical abnormalities.

Cricothyroidotomy Incision through the site below the thyroid cartilage for emergency opening of the tracheal passageway.

Crigler–Najjar syndrome A rare autosomal-recessive disorder marked by severe unconjugated hyperbilirubinemia seen shortly after birth.

Crohn disease An inflammation of the gastrointestinal tract that extends through all layers of the intestinal wall, most commonly affecting the terminal ileum. It may affect multiple portions of the intestine, leaving intervening normal areas in between the affected regions. The manifestations of Crohn disease differ in some respects from those of *ulcerative colitis,* although some overlap may occur. In Crohn disease, abdominal pain is the predominant symptom.

Cross-bridge The interaction between thick and thin filaments of the contractile apparatus when myosin heads bind to actin.

Cross-bridge theory This theory of muscle contraction is suggested by the anatomic configuration of the sarcomere. Muscle shortening is accomplished by increasing the amount of overlap of actin and myosin filaments. Also called the *sliding filament theory.*

Croup An acute infection of the upper and lower respiratory tracts that occurs primarily in infants and young children; it is characterized by hoarseness and a distinctive harsh cough.

Crush fracture A fracture that is consistent with cancellous bone trauma. Also called *compression fracture.*

Cryptogenic cirrhosis Advanced liver disease in a small number of patients with neither a suggestive history nor any detectable markers that would place them in any of the four main groups of cirrhosis.

Cryptorchidism Undescended testes.

Culture An integrated pattern of customs, attitudes, values, and shared beliefs that bind people together to form a society.

Cushing disease Hyperfunctioning of the adrenal cortex with increased glucocorticoid (cortisol) secretion because of excessive secretion of adrenocorticotropic hormone (ACTH) from the anterior pituitary.

Cushing syndrome The clinical features of hypercortisolism, regardless of cause.

Cutaneous membrane Thin, flat organ, also known as *skin.* It is composed of two main layers: an outer, thinner layer, termed the *epidermis;* and an inner, thicker layer, termed the *dermis.*

CVA tenderness The costovertebral angle (CVA) is one of two angles that outline a space over the kidneys. Pain in this area is a common finding in pyelonephritis and other infections of the kidneys.

Cyanosis A blue coloration of the skin as a result of poor saturation of hemoglobin with oxygen.

Cyanosis is usually not evident until saturation falls below 75%.

Cystectomy Surgical removal of all or part of the urinary bladder.

Cystic fibrosis An autosomal-recessive condition with abnormal chloride channel function, producing lung and pancreatic disease in children.

Cystic kidney disease Acquired cystic kidney disease and polycystic kidney disease, where cysts form in the kidneys.

Cystitis Inflammation of the urothelium (lining of the bladder) resulting from infection, irritation, presence of foreign body, or trauma.

Cystocele Protrusion of a portion of the urinary bladder into the anterior vagina at a weakened part of the vaginal musculature. Predisposing factors include obesity, aging, inherent weakness, history of heavy-object lifting, or injury during childbirth or surgery.

Cytokine A peptide factor released by cells to influence the behavior of target cells. Cytokines have signaling, inflammatory, growth, and inhibitory functions.

Cytopathic Pertaining to significant cellular injury or death.

Cytoskeleton System of protein filaments in the cytoplasm of a cell that give the cell its shape and the capacity for purposeful movement.

D

Decubitus ulcer Localized area of cellular necrosis resulting from prolonged pressure between a bony prominence and an external object such as a bed or wheelchair. The tissues are deprived of blood supply and eventually die. Also called *pressure sore.*

Deep partial-thickness burn Second-degree burn characterized by destruction of entire dermis, leaving only epidermal skin appendages. All physiologic functions of skin are absent.

Deep vein thrombosis (DVT) A disorder involving a thrombus in one of the deep veins of the body, most commonly in the lower extremities.

Degranulate The release of granules by mast cells and basophils; the granules contain proinflammatory chemicals.

Degranulation Exocytosis of stored molecules contained in cytoplasmic vesicles.

Dehydration Excessive loss of water from body tissues, accompanied by an increase in serum osmolality and an increase in serum sodium level (hypernatremia).

Dehydroepiandrosterone (DHEA) An androgenic steroid hormone secreted largely by the adrenal cortex and found in human urine.

Deletion The loss of a piece of a chromosome.

Delirium An acute organic mental disorder characterized by confusion, disorientation, restlessness, and incoherence.

Delusion A fixed, false belief that is held despite considerable contradictory evidence.

Delusional disorder A behavioral constellation dominated by a system of fixed, false beliefs that are tenacious and typically refractory to contrary evidence.

Dementia Syndrome characterized by a general loss of intellectual abilities caused by either reversible or progressive disorders, most typically Alzheimer disease or multiinfarct dementia.

Demyelination Destruction, removal, or loss of the myelin sheath of a nerve or nerve fibers.

Dendritic cells A cell that captures antigens and migrates to the lymph nodes and spleen, where it presents the processed antigen to T cells.

Deoxyribonucleic acid (DNA) The biomolecule that carries genetic information in the cell. DNA is composed of covalently linked nucleotides that form long polymers.

Depressed fracture A fracture in which the fragment is displaced below the level of the surface of the bone, usually in the skull.

Depressed mood A hallmark symptom of major depression. This change in mood is relatively constant and is recognized both by the depressed person and by others.

Depression An abnormal mood disturbance characterized by exaggerated feelings of sadness and melancholy. Also known as *clinical depression.*

Dermatitis Inflammation of the skin.

Dermatome An area of skin that is innervated by a specific spinal nerve.

Dermatomyositis A rare collagen disorder characterized by the acute or insidious onset of muscle pain, weakness, fever, arthralgia, and, in some cases, a puffy erythematous eruption that is usually confined to the face and the eyelids.

Dermatophyte A fungus that causes infection of the skin. The most common dermatophytes are *Microsporum, Trichophyton,* and *Epidermophyton.*

Dermatosis Any disorder of the skin, especially those not associated with inflammation.

Dermis Inner, thicker layer of the cutaneous membrane.

Dermoepidermal junction The specialized area where the cells of the epidermis meet the connective tissue cells of the dermis.

Desensitization The process of manipulating or "training" the hypothalamus to react less forcefully to a perceived threat or stressor. This technique works by changing the predominant brain waves of the individual from beta waves to alpha waves that are slower and more normal.

Desquamation The shedding of epithelial elements from the skin surface.

Detrusor muscle Smooth muscle of the bladder body.

Diabetes insipidus An endocrine deficiency of antidiuretic hormone manifesting as excretion of large quantities of very dilute urine and excessive thirst.

Diabetes mellitus An endocrine disorder characterized by impaired glucose entry into insulin-sensitive cells because of an absolute or relative deficiency of insulin.

Diabetic retinopathy A disorder of retinal blood vessels characterized by capillary microaneurysms. It occurs most frequently in patients with long-standing, poorly controlled diabetes mellitus.

Dialysate fluid Prepared solution with varying concentrations of glucose and electrolytes used to aid dialysis.

Dialysis An artificial process that replaces the renal functions of diffusion and filtration necessary to maintain homeostasis.

Diaper rash A skin irritation resulting from feces or the ammonia and alkali byproducts of urine breakdown.

Diarrhea An increase in the frequency and fluidity of bowel movements. It is usually a primary sign of gastrointestinal tract disorders.

Diarthrosis Also called *synovial joint;* a freely movable joint in which a contiguous bony surface is covered by articular cartilage and connected by a fibrous connective tissue capsule lined with a synovial membrane.

Diastole A phase of the cardiac cycle in which the ventricles are relaxing and filling with blood.

Diastolic blood pressure The lowest measured pressure in the arteries just before the next ventricular ejection.

Dicrotic notch A small, downward deflection observed on the downstroke of an arterial pressure waveform, representing closure of the aortic or pulmonic valves at the onset of ventricular diastole.

Diencephalon "Between" brain; part of the brain between the cerebral hemispheres and the midbrain.

Diffusion Passive movement of a gas or other substance from an area of high concentration to low concentration, or the process by which solutes move across a semipermeable membrane from an area of greater concentration to one of lesser concentration.

Diffusion coefficient A constant that depends on the properties of the tissue and the solute; the rate of movement of a solute is proportional to the diffusion coefficient.

Diploid Containing two sets of homologous chromosomes and therefore two copies of each gene—one from each parent.

Disconjugate An abnormal positioning of the eyes such that they deviate from one another in the direction of gaze.

Disease Sum of the deviations from normal structure or function of any part, organ, or system (or combination thereof) of the body manifested by a characteristic set of symptoms and/or signs and whose cause, pathogenesis, and prognosis may be known or unknown.

Dislocation Displacement of a bone from its normal position in a joint to the degree that the articulating surfaces lose contact.

Displaced fracture A fracture in which the ends of fragments are separated.

Disseminated intravascular coagulation (DIC) A grave coagulopathy resulting from the inappropriate stimulation of clotting and fibrinolytic processes within the vascular system; it is often precipitated by immune mechanisms.

Distributive shock State of insufficient perfusion of body tissues because of abnormal distribution of blood (for example, with anaphylaxis, sepsis, and spinal cord injury).

Disuse atrophy The tendency of cells and tissues to reduce size and function in response to lack of trophic stimuli.

Disuse osteoporosis Reduction in quantity of bone or atrophy of skeletal tissue in response to lack of weight-bearing activity. May occur with prolonged bed rest.

Diuresis Excretion of large amounts of urine usually as a result of the actions of a diuretic.

Diurnal variation The regular (24-hour) recurrence of certain biological phenomena under

conditions of illumination; recurring during the daytime, or period of light.

Divergence A separation or movement of objects or processes away from each other.

Diverticulitis Inflammation of one or more diverticula, or outpouchings, in the intestinal wall.

Diverticulosis The presence of diverticula, or outpouchings, in the wall of the colon.

Diverticulum Outpouching of one or more layers of the wall of a structure in the gastrointestinal tract, especially in the colon or esophagus.

DNA polymerase An enzyme complex that binds to DNA, using it as a template for synthesis of a complementary DNA strand.

Dominant Referring to the gene allele that is overtly expressed in the cell's phenotype. Opposite of recessive.

Dopamine hypothesis A hypothesis that postulates that schizophrenia is the result of neuronal overactivity dependent on dopamine dysregulation.

Down-regulation A decrease in the number of cell receptors for a specific hormone resulting from the cell's prolonged exposure to high concentrations of the hormone. Down-regulation results in a decrease in the target cell response to a hormone.

Drug-induced asthma Asthma related to an ingested drug. An attack may occur within minutes of ingestion or may be delayed up to 12 hours. Nonsteroidal antiinflammatory drugs including indomethacin (Indocin) and ibuprofen (Motrin, Advil) are common causes.

Duchenne muscular dystrophy The most common and most severe form of muscular dystrophy; it is inherited as an X-linked trait and therefore afflicts only males.

Ductus deferens Thick, muscular tube that is continuous with the epididymis. The ductus deferens travels along the pelvic wall and joins with the seminal vesicle duct at the prostate to form the ejaculatory duct. Also called *vas deferens.*

Dumping syndrome The rapid emptying or "dumping" of stomach contents into the proximal small intestine attributable to loss of pyloric regulation of gastric emptying. This loss of function may occur after a gastrectomy.

Dysconjugate See *disconjugate*.

Dysfunctional uterine bleeding Abnormal endometrial bleeding not associated with tumor, inflammation, pregnancy, or trauma. It is most common around the time of menarche and menopause.

Dyslipidemia Abnormality in the concentrations of lipids and lipoproteins in the blood, especially an elevated low-density lipoprotein (LDL) level and a reduced high-density lipoprotein (HDL) level.

Dysmenorrhea Pain associated with menstruation; usually classified as primary (unrelated to an identifiable disease) or secondary (related to the presence of an underlying disease).

Dyspareunia Pain during sexual intercourse because of vaginal muscle spasms.

Dysphagia Difficulty in swallowing as perceived by the individual. It may include the inability to initiate swallowing and/or the sensation of ingested substances sticking to the esophagus.

Dysphoria The constant experience of unpleasant emotions.

Dysplasia An alteration in cellular growth in which cell morphologic characteristics are variable and disorderly. Dysplastic cells may become cancerous and therefore are often termed *preneoplastic.*

Dyspnea Breathlessness or difficulty breathing.

Dysrhythmia An abnormality of heart rhythm, including altered rates or sites of impulse initiation and abnormal conduction pathways.

Dysthymia A state of chronic depression.

Dystrophic Abnormal tissue growth that impairs function. May result from disordered growth (trophic) signals.

E

Eating disorder A group of behaviors often fueled by unresolved emotional conflicts symptomized by altered food consumption. Disorders include anorexia nervosa, bulimia, and binge eating.

Eccentric contraction A lengthening contraction that occurs when the load is greater than the amount of tension that the muscle is able to generate, such as walking down stairs (eccentric contraction of the quadriceps muscles).

Ecchymosis Bluish discoloration of the skin (bruise) caused by escape of blood into the tissues.

Eccrine sweat gland A sweat gland that opens directly onto the skin surface.

Echocardiogram A graphic representation of heart structures and movement produced by ultrasonography.

Ectasia Dilation of a tubular structure, as in mammary duct ectasia (in which the collecting ducts beneath the nipple and areola become dilated, thinned, and filled with secretions).

Ectopic In an abnormal location.

Ectopic ureter A single ureter that implants during fetal growth in any position other than normal, or an additional ureter.

Ectopy (cardiac) A cardiac impulse initiated at a site other than the sinoatrial node.

Edema An excess of fluid in the interstitial compartment.

Efferent neuron A neuron that carries information away from the central nervous system to the muscle cells, glands, or postganglionic neurons.

Effusion Presence of fluid in a contained space, causing pressure on structures within the space.

Ejaculation Expulsion of the ejaculate from the posterior urethra through the urethral meatus.

Ejection fraction Stroke volume divided by end-diastolic volume; indicates pumping efficiency of the ventricle.

Elastin A protein found in tendons and ligaments that provides some elasticity or extensibility.

Elastosis Skin wrinkling due to changes in collagen, with fibers becoming cross-linked and rearranged in thicker bundles.

Electrocardiogram A graphic record produced by an electrocardiograph, which records electrical conduction through the heart.

Electrochemical gradient A difference in concentration of charged particles across a membrane. Driving force that moves charged particles across a membrane as a result of the combined influences of concentration gradient and electrical charge gradient.

Electroconvulsive therapy (ECT) The induction of a brief convulsion by passing an electric current through the brain for the treatment of affective disorders.

Electroencephalogram Graphic tracing of the brain's action potentials; used to evaluate nervous tissue function.

Electrolyte Substance that releases charged particles (ions) when dissolved.

Electromyography A technique for evaluating muscle contraction. Using electromyography, aspects of the contractile process, such as time relationships between the beginning of electrical activity and the actual contraction of the muscle, can be studied.

Electron transport chain A series of proteins on the inner mitochondrial membrane that move an electron from a higher to a lower energy level and create a proton gradient.

ELISA Abbreviation for enzyme-linked immunosorbent assay, a test used in screening for HIV antibodies and measuring the quantity of numerous substances in blood and urine.

Embolectomy A surgical incision into an artery for the removal of an embolus.

Embolus A collection of material (thrombus, air, fat, tumor cells, bacteria, amniotic fluid) propelled by blood flow to another site, where it lodges and causes obstruction of flow.

Embryoscopy A procedure in which a scope is passed through the mother's abdominal wall and into the uterus to visualize and sample embryonic tissues.

Emission One of the two phases of ejaculation. During emission, secretions from the periurethral glands, seminal vesicles, and prostate are deposited with sperm into the prostatic urethra.

Emphysema A chronic obstructive respiratory condition characterized by abnormal, permanent enlargement of air spaces distal to the terminal bronchiole with destruction of their walls and without obvious fibrosis.

Empyema Accumulation of pus in a body cavity, especially the pleural space.

Encapsulation Physiologic process of enclosure in a sheath composed of a substance not normal to the part. Prevents opsonization (recognition and binding) by antibodies and thus prevents the microorganism from being phagocytized.

Encephalitis An inflammatory condition of the brain.

Encopresis Fecal holding with constipation and fecal soiling.

Endemic disease A physical or mental disorder caused by health conditions constantly present within a community.

Endocardium A layer of endothelial cells that lines the chambers of the heart. The layer of heart muscle immediately under the endocardium is called the *subendocardium*.

Endocrine organ Any organ that manufactures and secretes hormones into the bloodstream.

Endocrine system The cells and organs that produce and secrete hormones into the bloodstream.

Endocytosis Cellular ingestion of extracellular molecules.

Endogenous depression Mental depression arising from characteristics within the person as opposed to depression resulting from external events.

Endometrioma A mass of endometrial tissue that grows outside the lining of the uterine cavity in the condition known as *endometriosis.*

Endometriosis Growth of endometrial tissue outside the lining of the uterine cavity; an abnormal condition with potentially destructive effects on the pelvic organs.

Endometrium The innermost lining of the uterus, consisting of two layers: a thin deep layer, called the *basilar layer;* and a thick superficial layer, referred to as the *functional layer.* During a woman's reproductive years, the endometrium displays a constant cyclic activity of alternate proliferation and sloughing of the functional layer in response to hormonal secretion.

Endomysium The connective tissue that surrounds the sarcolemma of an individual muscle fiber.

End-organ damage Target organs refer to major organs fed by the circulatory system, such as the heart, kidneys, brain, and eyes. Damage may be from uncontrolled hyperglycemia, hypertension, hypotension, or hypovolemia.

Endorphin One of a group of potent endogenous opioid peptides derived from cells in the hypothalamus; also found in the periaqueductal gray matter of the brain. β-Endorphin has been found to have analgesic properties.

Endoscopic retrograde cholangiopancreatography (ERCP) A procedure whereby an optical scope is passed through the mouth, esophagus, stomach, and duodenum and then guided in a retrograde fashion into the pancreaticobiliary system. Using this technique, physicians can complete a number of therapeutic procedures without performing a laparotomy.

Endoscopic sclerosis A procedure for the treatment of esophageal varices that is accomplished by passing a flexible needle through the gastroscope and injecting various sclerosant solutions into and around the bleeding varix.

Endospore A dormant, tough, and temporarily nonreproductive structure produced by a bacterium. It is not a true spore.

Endosteum The thin membrane that covers the medullary cavity in longer bones.

Endotoxin A heat-stable lipopolysaccharide derived from the cell wall of gram-negative bacteria that induces the release of pyrogens and inflammatory mediators from immune cells.

Energy The capacity to operate or work, measured in kilocalories (kcal); 1 kcal represents the amount of energy required to raise the temperature of 1 kg of water 1°C.

Enkephalins One of two types of pain-suppressing pentapeptides; they are produced in the body and located in the pituitary gland, brain, and gastrointestinal (GI) tract.

Enteropathic arthritis Refers to joint manifestations of inflammatory bowel diseases such as ulcerative colitis and Crohn disease.

Enuresis Involuntary voiding; the term is generally used when referring to inappropriate bed-wetting in children.

Eosinophil A leukocyte that is the same size as a neutrophil but contains a two-lobed nucleus and large, coarse, eosinophilic granules that fill the cell; eosinophils participate in antiparasitic, allergic, and inflammatory responses.

Epicardium A layer of epithelial cells that covers the outer surface of the heart and forms the inner (visceral) layer of the pericardial sac.

Epidemic An outbreak of a disease that occurs suddenly and affects numbers of people clearly in excess of normal expectancy.

Epidemiology The study of patterns of disease among human populations for the purpose of establishing programs to prevent and control their spread.

Epidermal proliferating unit Group of active basal cells, together with vertical columns of migrating keratinocytes, that are undergoing mitosis.

Epidermis Outer, thinner layer of the cutaneous membrane.

Epididymis Tightly coiled tube in which sperm mature and develop the ability to swim; lies along the top of and behind the testes.

Epididymitis Inflammation of the epididymis.

Epigenetics The transfer of heritable traits from parent cells to offspring that are not coded in the DNA.

Epiglottitis An inflammation of the epiglottis, characterized by fever and stridor. Can be a life-threatening condition.

Epileptogenic focus Cellular focus in the brain with the capacity to induce epilepsy.

Epimysium The connective tissue surrounding a muscle.

Epinephrine A neurotransmitter that produces some of the same effects as norepinephrine but has a greater influence on cardiac action. Epinephrine enhances myocardial contractility, increases heart rate, and increases venous return to the heart, thus increasing cardiac output and blood pressure.

Epiphyseal plate A segment of a long bone between the metaphysis and epiphysis developed from a center of ossification and distinct from the shaft. An area of growth in a bone.

Epispadias A congenital anomaly in which the urethra opens on the dorsal aspect of the penis at a point proximal to the glans.

Epistaxis Hemorrhage from the nose; nosebleed.

Epitope A site on the surface of an antigen that is specifically recognized by an immune cell, thus stimulating an immune response.

Equilibrium Sense of balance.

Erectile dysfunction Inability of the adult male to achieve or sustain a penile erection.

Erection A complicated interaction of vascular, neurologic, and hormonal factors that enables the penis to become rigid, achieve penetration, and deposit sperm.

Erythema Diffuse redness of skin.

Erythroblastosis Presence of erythroblasts in the blood due to premature release from the bone marrow.

Erythrocyte Mature biconcave red blood cell that has no internal organelles.

Erythromelalgia Painful erythema (redness of the skin) of the palms and soles due to congestion of the capillaries.

Erythron The blood as a single body system.

Erythropoiesis The process of red blood cell production.

Erythropoietin Hormone produced primarily by the kidneys that stimulates bone marrow to produce erythrocytes.

Eschar Burn tissue.

Escharotomy A surgical incision through eschar of a circumferential extremity burn for the purpose of restoring distal blood flow, or through eschar of the chest to restore respiratory expansion.

Esophageal atresia Congenital anomaly in which the esophagus is closed off in a blind pouch. It occurs in about 1 of every 4000 live births and requires immediate surgical correction.

Esophageal varix Abnormally dilated blood vessel lying immediately below the mucous membrane of the esophagus that connects the hypertensive portal system with the systemic circulation. Esophageal varices may rupture, causing massive hemorrhage.

Esophagitis Inflammation or infection of the esophagus.

Essential amino acids Amino acids that must be supplied in the diet because the body cannot manufacture them.

Estrogen One of a group of ovarian hormones that promote the development of female secondary sex characteristics. During the menstrual cycle, estrogen renders the female reproductive tract suitable for fertilization of the ovum, implantation of the zygote, and provision of nutrition for the early embryo.

Etiology Study of the assignment of causes or reasons for phenomena.

Euchromatin Chromatin that is less densely packed and potentially open to transcription, as opposed to heterochromatin that is condensed and not open to transcription. "Normal" chromatin.

Eukaryote A cell that has a true nucleus bounded by a nuclear membrane.

Euphoria Also called *expanded mood,* is a hallmark symptom of both mania and hypomania. Extreme cheerfulness, enthusiasm, and optimism are present, but the joyful, buoyant mood is disproportionate to events and surroundings.

Ewing sarcoma A malignant round cell tumor (marrow tumor) that is relatively uncommon but rapidly growing.

Exacerbation A relatively sudden increase in the severity of a disease or any of its signs and symptoms.

Exercise-induced asthma Asthma that manifests 5 to 10 minutes after the exercise period begins. The increased rate and depth of respiration during exercise, especially in cold air, leads to cooling and dehydration of the lower airways.

Exhaustion A stage in the stress response that occurs when the stressor is too great or prolonged, resulting in depletion of energy reserves.

Exocytosis The process of cellular secretion through the plasma membrane accomplished by opening vesicles into the extracellular space.

Exon The portion of an RNA transcript that remains after unwanted sections (introns) have been removed from the primary transcript. A linear section of DNA that serves as a template for synthesis of a particular RNA sequence.

Exophthalmos Protrusion of the eyeball.

Exotoxin Toxins, such as enzymes or pore-forming proteins, produced by bacteria that cause physiologic dysfunction in the host.

Extracellular fluid Body fluid that is not inside the cells; includes vascular, interstitial, and transcellular fluids.

Extraocular Outside the globe of the eyeball.

Extrapyramidal system Part of the brain that includes the corpus striatum, subthalamic nucleus, substantia nigra, red nucleus, and the interconnections with the reticular formation, cerebellum, and cerebrum.

Extrapyramidal tract Outside the pyramidal tract of the brain. Composed of the nuclei and fibers involved in motor activities, extrapyramidal tracts control and coordinate postural, static, support, and locomotor mechanisms. Do not cross over in the medullary pyramid.

Extrinsic Originating from sources outside of the individual.

Extrinsic asthma Also called *allergic asthma,* it commonly affects children and young adults. Attacks are related to specific antigens and are mediated by immunoglobulin E.

Extrinsic pathway of clotting The mechanism that produces fibrin after tissue injury, beginning with formation of an activated complex between tissue factor and activated factor VII and leading to activation of factor X, which induces the reactions of the common pathway of coagulation.

Exudate Fluid of high protein content that moves into tissues or cavities as part of a reaction to inflammation or injury.

F

Fascia A sheath of connective tissue that envelops muscles or other parts of the body.

Fascia of Buck Layer of deep fascia covering the penis.

Fasciculus Bundle of muscle fibers that compose individual muscles.

Facioscapulohumeral muscular dystrophy A rare, inherited, autosomal-dominant trait that affects the muscles of the shoulder girdle and the face and upper arms.

Fast twitch (type II, white) A muscle fiber that can develop high tension rapidly. It is usually innervated by a single α-motor neuron and has low fatigue resistance, low capillary density, low levels of aerobic enzymes, and low oxygen availability.

Fat The most concentrated dietary source of energy; derived from either animals or vegetables. Provides 9 kcal/gm of energy when metabolized.

Fat emboli syndrome A circulatory condition characterized by a plug of fat blocking an artery; it enters the circulatory system after the fracture of a long bone.

Fatigue A lack of physical or emotional energy or power.

Fatty acid An organic acid with a long, straight hydrocarbon chain that is a fundamental component of lipids. Some fatty acids are manufactured by the body; others are essential and must be supplied in the diet.

Fetotoxic Referring to a substance that is damaging to a developing fetus; similar to teratogenic.

Fibrillation Cardiac dysrhythmia characterized by rapid, random myocardial contractions and uncoordinated pumping action.

Fibrinolysis Dissolution or breakup of a fibrin clot.

Fibroblast Connective tissue cells that produce collagen fibers, which compose the bulk of the dermis.

Fibrocystic breast disease A benign condition in which the presence of palpable breast masses (cysts) corresponds to fluctuations in the menstrual cycle; the masses may be associated with pain and tenderness.

Fibromyalgia syndrome A painful, noninflammatory musculoskeletal disorder associated with fatigue and multiple somatic symptoms.

Fibrosis Condition of decreased elasticity because of excessive deposition of fibrin and collagen in the tissue (e.g., restrictive process characterized by thickening of the alveolar interstitium).

Fight-or-flight response The reaction of the body to stressors that is mediated by the sympathetic nervous system and produces elevated heart rate and release of glucocorticoids from the adrenal cortex.

Filaments or protein filaments Fine threadlike fibers found in most tissues and cells of the body.

Filtration Passive movement of fluid across capillary walls as a net result of opposing forces of hydrostatic and colloid osmotic pressures.

First-degree burn Superficial tissue destruction in the outermost layers of the epidermis. All physiologic functions of the skin remain intact.

Fistula An abnormal connecting passageway between two organs or between an internal organ and the body surface.

Flagellum Motile (whiplike) appendage that allows a cell to move or swim.

Flail chest A thorax in which there are two fractures on at least two adjacent ribs that cause instability in part of the chest wall and paradoxical breathing.

Flank pain Discomfort to the posterior portion of the body between the ribs and the iliac crest. Often associated with the ureters.

Flat affect Lack of appropriate emotional expression.

Focal segmental glomerulosclerosis A condition where only some of the glomeruli are involved, resulting in scarring of the glomerulus. It is a cause of nephrotic syndrome and renal failure.

Fomite An inanimate object that transmits a pathogen to a new host.

Foreskin Also called *prepuce;* penile skin that overlies the glans and is removed in circumcision.

Fossa navicularis Area of widening near the end of the penile urethra.

Fournier gangrene A rare condition involving a gangrenous necrosis of the scrotum, penis, or perineum.

Fourth-degree burn A full-thickness burn that penetrates the dermis to reach muscle or bone.

Fracture A break or disruption in the continuity of a bone, an epiphyseal plate, or cartilage.

Frank–Starling law of the heart Describes the relationship between diastolic stretch and subsequent increased strength of contraction. Also called the *length–tension relationship*.

Free-living bacteria Bacteria that can live outside the host cell.

Free radical An extremely reactive compound that avidly makes molecular bonds with other compounds.

Full-thickness burn Also called a *third-degree burn,* this type of burn is marked by destruction of epidermis, dermis, and underlying tissue. All physiologic functions of the skin are absent. This burn will not heal and requires autografting.

Full-thickness excision Removal by surgical knife of complete eschar to fascia. Full-thickness excision often leaves an uneven contour, which presents difficulty with grafting, resulting in poor cosmetic appearance.

Functional disorder of the endocrine system An endocrine disorder caused by a nonendocrine disease (e.g., chronic renal failure, liver disease, or heart failure).

Functional incontinence Loss of urine or feces as a result of factors external to the urinary or digestive tract, such as physical or cognitive impairment.

Functional syncytium A multinucleate mass of protoplasm that results from the merging of cells. It is characteristic of the gastrointestinal tract and heart, meaning that its separate cells have the ability to function in concert with one another in a unified manner.

Fungal infection Any inflammatory condition caused by a fungus.

Fungus A nonphotosynthetic, eukaryotic protist that is disseminated throughout the environment.

G

Gallbladder A distensible sac of about 30- to 50-mL capacity that connects the common hepatic duct to the common bile duct via the cystic duct.

Ganglion A group of neuronal cell bodies located outside of the central nervous system.

Gangrene Cellular death involving a large area of tissue; may be characterized as wet, dry, or gaseous.

Gap junction A cell-to-cell communication pore that allows small biomolecules to flow from the cytoplasm of one cell to the cytoplasm of an adjacent cell.

Gas exchange Diffusion of oxygen and carbon dioxide across the alveolar–capillary membranes of the lungs.

Gastrectomy Surgical removal of all or, more commonly, part of the stomach. This procedure may be used to remove a chronic peptic ulcer, to stop hemorrhage in a perforating ulcer, or to remove a malignancy.

Gastrin A stomach hormone that is released in response to certain types of food. Gastrin increases acid secretion by stomach parietal cells.

Gastritis Inflammation of the stomach lining. It may occur after the ingestion of irritating substances or in the presence of viral, bacterial, or chemical toxins.

Gastroenteritis Inflammation of the stomach and intestines; may occur on an acute or chronic basis and is commonly caused by viruses.

Gastroesophageal reflux disease (GERD) Backflow of gastric contents into the esophagus through the lower esophageal sphincter. GERD may or may not produce symptoms. The most

common manifestations of GERD are heartburn, regurgitation, chest pain, and dysphagia.

Gastroesophageal varices A complex of longitudinal tortuous veins at the lower end of the esophagus, enlarged and swollen as the result of portal hypertension.

Gene A unit of heredity consisting of a segment of DNA nucleotides that encodes a messenger RNA capable of being translated into a protein.

General adaptation syndrome (GAS) The total organism's nonspecific response to stress. Term was coined by Hans Selye.

Generalized anxiety disorder (GAD) Characterized by the continual presence of a moderate degree of anxiety without discrete periods of acute attacks. GAD symptoms include chronic anxiety and tension accompanied by headaches, abdominal problems, or sleep disturbances. Agoraphobia is rarely seen in GAD.

Generalized seizure A seizure that involves the whole brain surface and impairs consciousness.

Genital herpes A chronic herpes simplex viral infection usually transmitted by sexual contact.

Genital warts A condyloma of the genitals; caused by sexual transmission of human papilloma virus.

Genome The entire complement of genes and associated proteins located on chromosomes in the nucleus of a cell.

Genotype The genetic constitution of an individual; often described by listing the allele types at a certain gene locus.

Gerontology The study of all aspects of the aging process.

Gestational diabetes mellitus A condition of glucose intolerance first diagnosed in the mother during pregnancy. Usually disappears after delivery of the infant.

Ghon tubercle A nodule or swelling indicative of *Mycobacterium tuberculosis.*

Giant cell tumor Also called *osteoclastoma;* a benign but aggressive tumor with richly vascularized tissue consisting of plump spindle-shaped cells and numerous giant cells.

Gilbert syndrome A common, benign, autosomal-dominant condition that results in mild unconjugated (indirect) hyperbilirubinemia.

Ginglymus A type of joint that permits flexion and extension; examples include the interphalangeal joint of the finger, the elbow joint, or the knee joint. Also called a *hinge joint.*

Glasgow Coma Scale Scale developed by G. Teasdale and B. Jennett for the purpose of objectively assessing coma and impaired consciousness.

Glaucoma An abnormal condition of elevated pressure within the eye.

Glial Supporting or nonneuronal cells within the central nervous system.

Glomerulopathies Diseases of the renal glomeruli.

Glomerular filtration rate The rate of fluid filtration through all the glomeruli into Bowman capsules per minute; normally 125 mL/min.

Glomerulonephritis Inflammation of the glomerular capillary walls that causes impaired filtration and renal function.

Glucagon Hormone produced by the α cells of the pancreas that stimulates glycogenolysis and gluconeogenesis in the liver.

Glucocorticoid resistance model This model proposes a specific link between stress, immunity, and disease. Rather than viewing disease as a result of increased vulnerability due to stress, this model proposes that overwhelming stress reduces the sensitivity of the immune system to cortisol.

Glucocorticoids A class of steroid hormones secreted by the adrenal cortex; they are necessary for use of carbohydrates, fats, and proteins and for the body's normal response to stress.

Gluconeogenesis The production of glucose from amino acids and other substrates in the liver.

Glucose-6-phosphate dehydrogenase deficiency (G6PD) An inherited disorder characterized by red cells partially or completely deficient in G6PD, an enzyme critical in anaerobic glycolysis.

Glycogen A carbohydrate consisting of branched chains of glucose produced by muscles and the liver as a storage form of glucose.

Glycogenesis Production of glycogen from glucose in hepatic and muscle tissue.

Glycogenolysis Production of glucose from the breakdown of glycogen in hepatic and muscle tissue.

Glycolysis The anaerobic process of breaking down carbohydrates into simpler molecules, with the net production of two adenosine triphosphate and two pyruvate molecules per glucose molecule.

Glycosaminoglycan A protein polysaccharide contained in ground substances surrounded by collagen fibers in bone.

Glycosylated hemoglobin An index of glycemic control; the quantity of glucose attached to hemoglobin molecules (%), reflecting mean blood glucose values for a period of 120 days, usually reported as HbA_{1c}.

Goiter Enlargement of the thyroid gland.

Golgi apparatus A membrane-bound organelle in which the proteins and lipids that are synthesized in the endoplasmic reticulum are modified and sorted in preparation for transport to the lysosomes or plasma membrane.

Gomphosis joint An articulation created by the insertion of a conical process into a socket, such as the insertion of a root of a tooth into an alveolus of the mandible or the maxilla. Gomphosis is not a connection between true bones but is considered a type of fibrous joint.

Gonad An organ that produces sex cells. Derived from the urogenital ridge, the undifferentiated and primitive gonads become the testes in males and the ovaries in females.

Gonadotropins Hormones that stimulate the function of the testes and the ovaries (follicle-stimulating hormone [FSH] and luteinizing hormone [LH]).

Gonorrhea Common sexually transmitted infection involving the inflammation of epithelial tissue by the organism *Neisseria gonorrhoeae.* Characteristic symptoms include urethritis, dysuria, purulent urethral discharge, and redness and swelling at the site of the infection.

Goodpasture syndrome A chronic relapsing autoimmune disease usually associated with glomerulonephritis.

Gout A condition caused by lack of the enzyme uricase and by inability to oxidize uric acid into a soluble compound; characterized by recurrent attacks of articular and periarticular inflammation, accumulation of tophi (crystalline deposits) in bony and connective tissue, renal impairment, and formation of uric acid calculi.

Grading Assignment of degree of differentiation of tumor cells by histologic examination. The degree of anaplasia usually correlates with the degree of malignancy, with higher grades conferring greater malignant potential.

Gram stain A process by which it is determined whether bacteria can retain a basic dye after iodine fixation. This ability is the basis for classifying bacteria into gram-negative and gram-positive organisms.

Granulocyte A leukocyte with polymorphic nuclei and cytoplasmic granules. Neutrophils, basophils, and eosinophils are types of granulocytes.

Granulocytopenia An abnormal decrease in the total number of granulocytes in the blood.

Granuloma Tissue that forms into a nodular mass as a result of inflammation, infection, or injury.

Granuloma inguinale An ulcerative disease of the genital tract caused by the bacterium *Calymmatobacterium granulomatis.* The communicability of the disease is relatively low, and it is generally believed that repeated exposure is necessary for infection.

Granulomatous Relating to granulomas; chronic inflammatory lesions characterized by an accumulation of macrophages; epithelioid macrophages, with or without lymphocytes; and giant cells into a discrete granule.

Graves disease Hyperthyroid state characterized by exophthalmos and goiter from autoimmune stimulation of the thyroid.

Greenstick fracture An incomplete break in the bone with the intact side of the cortex flexed; this is usually seen in children.

Ground substance A material composed of a hydrated network of proteins, mainly glycoproteins and proteoglycans, that serves as the "cement" between layers of collagen fibers. Also known as *matrix.*

Growing pain A common soft tissue syndrome in children. The most common symptom is nocturnal pain that usually occurs in the calves, shins, and thighs. Also called *nonarticular rheumatism.*

Growth hormone A hormone secreted by the anterior pituitary gland with wide-ranging action, including effects on energy metabolism and increasing lean body mass.

Guillain–Barré syndrome Also called *acute idiopathic polyneuropathy* or *polyradiculoneuropathy;* Guillain–Barré syndrome is an inflammatory demyelinating disease of the peripheral nervous system.

Gustatory Pertaining to the sense of taste.

Gyrus A raised ridge or convolution on the surface of a structure (e.g., cerebral cortex).

H

H^+ Hydrogen ion; released by acids; determines pH; also called a *proton.*

HCO_3^- Bicarbonate ion; a base that binds and buffers H^+.

H_2CO_3 Carbonic acid; this acid is removed from the body in the form of carbon dioxide and water during exhalation.

H zone Corresponds to a region of the sarcomere occupied solely by myosin filaments with no actin filament overlap in cardiac and skeletal muscle.

Hair follicle Slender, cylindrical tube of epidermal cells in which hair grows.

Hallucination A perception for which there are no objective sensory data.

Haploid Containing only one set of chromosomes (as distinct from diploid), as in a sperm cell or egg cell.

Hapten Incomplete, lipid-soluble particle that is incapable of being an antigen by itself, but that becomes an antigen inside the body when it binds with a host protein called a *carrier*. When a hapten penetrates the epidermis and binds to a carrier, it can cause contact hypersensitivity.

Haustral churning The mixing movement of the haustra (the outpouchings in the colon wall) when material is in the proximal end of the colon.

Haversian system The basic unit of bone; also called *osteon.*

HBV infection Hepatitis B virus infection. Vertical transmission from an HBsAg-positive mother to the infant is a common mechanism of spread.

HCV infection Hepatitis C virus infection. A type of viral hepatitis that is usually chronic and transmitted most commonly by blood transfusion or percutaneous inoculation.

HDV infection Hepatitis D virus infection. A form of hepatitis that occurs only in patients coinfected with hepatitis B. HDV relies on HBV replication and cannot replicate independently. The disease usually progresses to a chronic state.

Heimlich maneuver An emergency procedure for dislodging an obstruction from the trachea. It consists of grasping the choking person from behind, placing the hands around the victim's waist just below the sternum (in a fist with the thumb toward the body), and pulling inward and upward with force to dislodge the obstruction. Now called *abdominal thrusts.*

Helicobacter pylori An infectious gastrointestinal tract bacterium first identified in 1982. Since then, *H. pylori* has generated worldwide attention for its role in the promotion of chronic gastritis, peptic ulcer disease, and gastric carcinoma. The mode of transmission of *H. pylori* is still unclear, although person-to-person, fecal-oral spread is suspected because of the tendency of *H. pylori* infections to cluster in families.

Hemangiomas Benign tumors consisting of a mass of blood vessels.

Hemarthrosis Blood in a joint cavity.

Hematemesis Blood in vomitus.

Hematochezia Feces containing bright red blood.

Hematoma A mass caused by extravasation of blood into a tissue or cavity (raised bruise).

Hematopoiesis Production of cells in the bone marrow, including red blood cells, white blood cells, and platelets.

Hematuria Blood in the urine.

Hemianopsia Partial loss of the same field of vision in both eyes.

Hemiparesis Motor weakness affecting one side of the body, usually occurring with lateral cerebral injuries such as strokes.

Hemiplegia Paralysis of one side of the body.

Hemochromatosis A disorder (usually genetic) of iron metabolism characterized by excess absorption of iron, elevation of ferritin levels, and deposition of iron in organs such as the liver.

Hemodynamics Physical laws governing blood flow.

Hemoglobin Oxygen-carrying protein in the red blood cells. Formed from four polypeptide globulins, each containing a porphyrin ring with a central iron molecule.

Hemolysis Separation of hemoglobin from red blood cells and its appearance in the fluid in which the corpuscles are suspended; red blood cell lysis.

Hemophilia A group of hereditary disorders characterized by a deficiency of one of the factors necessary for coagulation of the blood, usually factor VIII or IX.

Hemoptysis Expectoration of blood, the origin of which is the lungs or bronchial tubes.

Hemostasis Arrest of bleeding; prevention of blood loss.

Hemothorax Accumulation of blood and fluid in the pleural space.

Hepatic encephalopathy A neuropsychiatric manifestation of extensive liver damage. Also known as *hepatic coma*.

Hepatitis An inflammatory condition of the liver. Potential causes include viral, bacterial, fungal, and protozoal infections; drugs and toxins; autoimmune disorders; and metabolic disorders.

Hepatocellular carcinoma A common form of primary hepatic malignancy. Signs and symptoms include hepatomegaly, abdominal pain, weight loss, nausea, and, in advanced cases, jaundice and ascites. Also called *hepatoma.*

Hepatocellular failure Acute or chronic loss of essential liver function resulting in encephalopathy and a variety of other problems, including coagulopathy, infection, hypoglycemia, and low albumin.

Hepatoma A primary liver cancer arising from cells normally found in the liver, not to be confused with cancer metastatic to the liver from a distant site.

Hepatorenal syndrome A type of kidney failure characterized by gradual loss of function without signs of tissue damage in patients with chronic liver disease.

Hereditary hemochromatosis An autosomal-recessive disorder caused by the activity of a mutant gene called *HFE,* which allows excessive and uncontrolled iron absorption by the gastrointestinal (GI) tract.

Herpes simplex virus A virus causing infection of the skin and nervous system; often associated with painful fluid-filled vesicles.

Herpes zoster (shingles) An acute localized inflammatory disease of a dermatomal segment of the skin caused by the same herpesvirus that causes chickenpox.

Herpesviruses Important group of viral agents producing infections in humans. Two types of herpes simplex viruses, referred to as types 1 and 2, may be sexually transmitted.

Heterochromatin A type of chromatin (DNA) that is tightly compacted and genetically inactive.

Heterozygous Having two different alleles for a specific gene product.

HEV infection Hepatitis E virus infection. A self-limited type of hepatitis acquired by ingestion of fecally contaminated water or food.

Hiatal hernia A defect in the diaphragm that allows a portion of the stomach to protrude through the diaphragmatic opening into the thorax.

High blood pressure A persistent elevation of blood pressure >140 mm Hg systolic or >90 mm Hg diastolic. High blood pressure is an elevated blood pressure reading. If it is consistently >120/80 mm Hg but <140/90 mm Hg, then it is called *prehypertension*.

Hilum Concave portion of the kidney that faces the vertebral column through which nerves, blood vessels, and ureters enter and exit the kidney.

Hinge joint A joint that permits flexion and extension, such as the interphalangeal joint of the finger, the elbow joint, or the knee joint. Also called a *ginglymus joint.*

Hirschsprung disease A congenital disorder of the large intestine in which the autonomic nervous system ganglia in the smooth muscle are absent or markedly reduced in number, which causes poor or absent peristalsis.

Hirsutism Excessive growth of hair or the presence of hair in unusual places.

Histamine A compound found in cells, especially mast cells; it is produced by the breakdown of histidine (an amino acid). It is released in allergic inflammatory reactions.

Histiocyte A type of cell normally present in small numbers around blood vessels, but in pathologic conditions it can migrate in the dermis as a tissue monocyte. It can also form abundant reticulum fibers. When it phagocytizes bacteria and particulate matter, it is referred to as a *macrophage.*

Histone A protein around which linear DNA is wrapped.

Histrionic Theatrical, dramatic.

Hodgkin disease A progressive malignancy of the lymph node characterized by the presence of Reed–Sternberg cells and slow, predictable dissemination through the lymphatic vessels.

Homeostasis A dynamic steady state, representing the net effect of a number of physiologic actions and reactions.

Homologous Corresponding in structure. For example, the labia majora are *homologous* with the scrotum of the male.

Homologous chromosomes A pair of chromosomes in a diploid cell that contain similar gene loci, each being derived from a different parent.

Homozygous Having two identical alleles for a specific gene product.

Hormone A bloodborne chemical messenger that affects target cells anatomically distant from the secreting cells.

Hormone agonist A chemical that binds to a hormone receptor and initiates intracellular activities identical to those caused by hormones. Some medications exert their therapeutic effects through this process.

Hormone antagonist A chemical that competes with hormones for cell receptors. Antagonists bind to cell receptors and prevent the occurrence of intracellular activities associated with hormone-receptor binding. Some medications produce their therapeutic effects through this process.

Hormone receptor A protein on or within a target cell that binds to circulating hormones and allows the cellular response to a specific hormone.

Hormone-receptor binding is the first step in the cellular response to a particular hormone.

Host–parasite relationship The interaction between the host and the microorganisms that reside on or in it.

Human immunodeficiency virus (HIV) Retroviruses (HIV-1, HIV-2) that infect CD_4+ cells, causing a defect in cell-mediated immunity and failure of the immune system to function properly.

Human leukocyte antigen (HLA) complex The major histocompatibility complex (MHC) in human leukocytes that present protein antigens on the surface of the cell.

Human papilloma virus (HPV) infection A virus that is the cause of common warts of the hands and feet, as well as lesions of the oral, anal, and genital cavities. More than 50 types of HPV have been identified.

Humoral immunity A form of immunity mediated by circulating antibodies that coat the antigens and target them for destruction.

Hyaline membranes Membranes in alveolar tissue that look like glass. The alveoli are filled with proteinaceous fluid and epithelial cells.

Hydrocele Accumulation of fluid in the tunica vaginalis testis; one of the most common causes of scrotal swelling.

Hydrocephalus Increase in the amount of cerebrospinal fluid attributable to blocked circulation or absorption and the consequent enlargement of the brain ventricles.

Hydronephrosis Distention of the pelvis and calyces of the kidney by urine that cannot flow past an obstruction in the ureter.

Hydrophilic Soluble in water but not in lipid.

Hydrophobic Insoluble in water but soluble in lipid.

Hydropic swelling An increase in intracellular fluid volume and changes in intracellular organelles in association with cell injury. Also termed *oncosis.*

Hydrostatic pressure Pressure exerted by a liquid.

Hydroureter Distention of a ureter with urine, usually resulting from an obstruction process.

Hyperacusis Exceptionally acute hearing, the hearing threshold being unusually low. It may or may not be accompanied by pain.

Hyperaldosteronism A condition characterized by hypersecretion of aldosterone, occurring as a disease of the adrenal cortex or as a response to adrenal disease.

Hyperalgesia An increased sensitivity to painful stimuli characterized by a lower-than-normal pain threshold.

Hypercalcemia Serum calcium concentration greater than normal.

Hypercapnia An abnormally high amount of carbon dioxide in the blood.

Hypercortisolism Elevated serum level of cortisol.

Hyperemesis gravidarum A Latin term meaning "excess vomiting in pregnant women." Unlike the transient nausea and vomiting that occurs in about half of women in the first trimester of pregnancy, hyperemesis gravidarum continues throughout the entire pregnancy.

Hyperemia Localized redness produced by increased blood flow.

Hyperkalemia Serum potassium concentration higher than normal levels.

Hyperkeratosis Horny overgrowth of epidermis, such as callus formation.

Hypermagnesemia Serum magnesium concentration greater than normal levels.

Hypermetabolic state A condition of abnormally high basal metabolic rate. May be indicated by an increase in skin temperature, such as that occurring in hyperthyroidism and after sun exposure or sunburn.

Hypernatremia Serum sodium concentration greater than normal levels; results from a gain of salt relative to water or a loss of water relative to salt; water deficit.

Hyperopia Farsightedness, or the inability of the eye to focus on nearby objects.

Hyperparathyroidism An abnormal endocrine condition characterized by hyperactivity of the parathyroid glands.

Hyperphosphatemia Serum phosphate concentration greater than normal levels.

Hyperplasia Abnormal multiplication or increase in the number of normal cells in normal arrangement in a tissue.

Hypersensitivity An abnormal excessive response to a sensitizing antigen.

Hypertension (see *high blood pressure*) A persistent elevation of blood pressure >140 mm Hg systolic or >90 mm Hg diastolic.

Hypertensive crisis A sudden, severe increase in blood pressure that could be life threatening.

Hyperthyroidism Overactivity of the thyroid gland.

Hypertonic fluid Fluid that has a higher particle concentration (osmolality) than normal body fluid; causes a net flow of water across cell membranes out of cells.

Hypertrichosis lanuginosa Excessive hair growth over the entire body.

Hypertrophy An increase in cell or tissue size and function.

Hypocalcemia Serum calcium concentration below normal values.

Hypochromia An abnormal decrease in the hemoglobin content of the erythrocytes resulting in pale appearance.

Hypodermis Loose subcutaneous layer rich in fat and areolar tissue lying beneath the dermis. Also known as *superficial fascia.*

Hypoglycemic sweating Sweating caused by low blood glucose concentration. Usually distinguishable from other causes of sweating attributable to the additional symptoms of weakness, tachycardia, hunger, headache, and "inward nervousness" manifested as mental irritability and confusion.

Hypogonadism A deficiency in the size or function of the ovary or testis.

Hypokalemia Serum potassium concentration less than normal levels.

Hypomagnesemia Serum magnesium concentration less than normal levels.

Hypomania Also called *partial mania,* can include any symptoms of mania but without the loss of reality testing, without psychosis (e.g., hallucinations, delusions), and without significantly impaired functioning.

Hypomenorrhea A deficient amount of menstrual flow, usually the result of an endocrine or systemic disorder that interferes with hormonal function. It may also result from partial obstruction of the menstrual flow by the hymen or a narrowing of the cervical os.

Hyponatremia Serum sodium concentration less than normal levels; results from gain of water relative to salt or loss of salt relative to water; water intoxication; water excess.

Hypophosphatemia Serum phosphate concentration less than normal levels.

Hypophysis The pituitary gland, which consists of anterior and posterior lobes.

Hyposensitization Reduction in sensitivity to an allergen, accomplished by administering low doses of the allergen, which binds with immunoglobulin G.

Hypospadias A congenital anomaly in which the urethral meatus is located on the undersurface of the penis or on the perineum.

Hypothalamic–pituitary–adrenal (HPA) axis Refers to a hierarchy of control mechanisms whereby the hypothalamus regulates the anterior pituitary and the pituitary regulates the secretion of hormones from the adrenal cortex.

Hypothalamus A group of nuclei at the base of the brain concerned with regulation of body processes: temperature, thirst, hunger, satiety, and adaptive sexual behaviors.

Hypothyroidism Underactivity of the thyroid gland.

Hypotonic fluid Fluid that has a lower particle concentration (osmolality) than normal body fluid; causes a net flow of water across cell membranes into cells.

Hypoventilation Decreased exchange of air in the alveoli in relation to oxygen consumption; influenced by a decreased rate and depth of respiration and evidenced by elevated $Paco_2$ values.

Hypovolemic shock A state of physical collapse and prostration usually caused by massive blood loss.

Hypoxemia An abnormally low amount of oxygen in the blood.

Hypoxia A reduction in oxygen at the tissue level that may lead to failure of aerobic production of adenosine triphosphate.

I

I bands "I" indicates "isotropic"; these bands are light in color and correspond to the position of thin actin filaments extending in both directions from the Z line in sarcomere of striated muscle.

Iatrogenic Resulting from the activity of a health care provider.

Icterus Also called *jaundice.* A yellow discoloration of the skin, mucous membranes, and sclerae of the eyes that is caused by greater-than-normal amounts of bilirubin in the blood.

Idiopathic Without known cause.

Idiopathic hypertension High blood pressure of unknown cause; also called *primary hypertension.*

Ileocecal valve Sphincter between the small and large intestines that is normally closed so that the contents of the large intestine cannot move in a retrograde fashion back into the small intestine. It opens in response to peristaltic contractions in the small intestine, bringing intestinal contents toward it.

Illusion The misperception of a real sensory stimulus.

Immobilization A mechanical action of limiting or preventing movement of the body or a body part. Prolonged immobilization may cause a shortening of connective tissue, a breakdown of cartilage, a weakening of ligaments, and a decrease in the muscles' ability to contract, as well as increased bone resorption.

Immunity A state of active resistance to a particular pathogen, which requires functional T- and B-cell memory cells.

Immunization Exposure of a susceptible host to an altered pathogen that does not cause disease but instigates the host to create antibodies to that pathogen.

Immunodeficiency Failure of immune system mechanisms to defend against pathogens. There are two broad categories of immunodeficiencies, based on the mechanism of lymphocyte dysfunction: primary and acquired.

Immunogen Foreign substance, cell, toxin, or protein that causes the components of the immune system to react and respond, inducing the formation of antibodies. Also known as an *antigen.*

Immunogenicity The ability to stimulate an immune response.

Immunoglobulin Any of five structurally distinct classes of proteins that function as antibodies in the serum and secretions of the body.

Immunosuppression The inability to produce an immune response to an antigen, resulting in reduced resistance to infection.

Immunotherapy The use of immune products such as monoclonal antibodies to treat specific diseases.

Impacted fracture A fracture caused by excessive force that telescopes or drives one fragment into another.

Impaired fasting glucose A disorder of glucose tolerance, not diagnostic of diabetes, that is characterized by a fasting blood glucose value between 100 and 125 mg/dL.

Impaired glucose tolerance A disorder of glucose tolerance, not diagnostic of diabetes, that is characterized by a 2-hour postprandial blood glucose value of between 140 and 200 mg/dL.

Impetigo An acute, contagious skin infection characterized by the formation of vesicles, pustules, and yellowish crusts.

Impotence Failure to achieve and maintain an erection of the penis; erectile dysfunction.

Impulsivity Spontaneous acting out of impulses accompanied by failure to plan ahead, predict consequences, or consider other possibilities.

Incomplete fracture A fracture in which the cortex of the bone buckles or cracks without disrupting bone continuity.

Incontinence The inability to adequately control urination or defecation.

Induction The process of stimulating and determining morphogenetic differentiation in a developing embryo. Also the initial phase of cancer chemotherapy.

Induration Hardness, such as that resulting from multiple intramuscular or subcutaneous injections of medication.

Inert tissue Soft tissue that possesses no ability to contract or relax; this includes the joint capsule, ligament, bursa, fascia, dura mater, and nerve root.

Infectious disease A pathologic process caused by a microorganism that is transmissible from one host to another.

Infiltrate Fluid or material that has moved into tissues.

Inflammation The body's protective response at the site of injury or tissue destruction. It is important to recognize that although infectious agents can produce inflammation, infection is not synonymous with inflammation.

Inflammatory bowel disease A general term for inflammatory diseases of the bowel, such as ulcerative colitis and Crohn disease.

Inhalation injury Cellular injury to lung tissue as a result of inhalation of a toxic substance such as smoke. Smoke inhalation significantly increases the morbidity and mortality from burn injury.

Innate immune response Part of the host defense system that is composed of mechanical and biochemical barriers, phagocytes, and chemical mediators. Does not include specific B and T cells.

Inotropy The force or rate of cardiac contraction; similar to contractility.

Insulin Hormone produced by the β cells of the pancreas; has wide-ranging effects on energy metabolism and protein synthesis.

Insulin resistance The condition of requiring an increased amount of insulin for the same level of tissue glucose utilization.

Integrins A large family of transmembrane proteins that mediate adhesion of cells to the extracellular matrix.

Integument Covering; refers to the skin.

Integumentary system The skin and its appendages, including the hair and nails.

Intercurrent Occurring during the course of an already existing disease.

Interferons Natural glycoproteins formed by cells exposed to a virus.

Interleukins One of a large group of proteins produced by immune and other cells. Includes growth factors, inflammatory mediators, and other regulatory signaling peptides.

Intermembranous ossification The process in which osteoblasts in the inner layer of the periosteum are responsible for the increase in width of bones.

Interstitial cystitis Inflammation of the bladder that is believed to be associated with an autoimmune or allergic response.

Interstitial fluid Fluid that lies between the cells; a component of extracellular fluid.

Interstitial space The space between cells.

Intertrigo An erythematous irritation of opposing skin surfaces caused by friction.

Intestinal obstruction Failure of the contents of the intestines to progress through the lumen of the bowel because of mechanical blockage.

Intestinal villi Fingerlike projections that line the small intestine and serve to increase the surface area of the intestine for digestion and absorption of nutrients.

Intraaortic Within the aorta. For example, in intraaortic balloon counterpulsation, a catheter with a balloon at the distal segment is inserted through the femoral artery and positioned in the aorta just distal to the left subclavian artery.

Intracellular fluid Fluid that is inside the cells.

Intracellular obligate parasites Bacteria that must live inside a living cell.

Intraocular Inside the globe of the eyeball.

Intrarenal Occurring within the kidney.

Intravenous pyelography A diagnostic procedure in which an iodine-based contrast material is injected into the vascular system to allow visualization of the kidneys and urinary tract.

Intrinsic Originating from within the individual.

Intrinsic asthma Asthma caused by pathophysiologic disturbances that do not involve IgE-mediated mechanisms. This type of asthma frequently develops in middle age. Psychological stress factors, pulmonary irritants, and exercise may precipitate an asthma attack.

Intrinsic (enteric) nervous system Neural structures belonging entirely to the gastrointestinal (GI) system that control most GI functions and are responsible for many reflexes occurring locally in the GI tract. The intrinsic nervous system is composed of two layers: the myenteric plexus and the submucosal plexus.

Intrinsic pathway of coagulation A sequence of reactions leading to fibrin formation, beginning with the contact activation of factor XII, followed by the sequential activation of factors XI and IX, and resulting in the activation of factor X, which in activated form initiates the common pathway of coagulation.

Intron The portion of a primary RNA transcript that is removed before translation of the RNA message occurs.

Intussusception A telescoping of a portion of the bowel into an adjacent distal portion. It is most common in infants and occurs three times more often in males than in females.

Invasion The process by which malignant cells move through the basement membrane and gain access to blood vessels and lymphatic channels.

Inversion An abnormal condition in which a section of chromosome is reversed and reinserted into DNA.

Involucrum A sheath or coating of new bone growth outside of existing bone seen in pyogenic osteomyelitis. It results from the stripping of the periosteum by the accumulation of pus within the bone and new bone growing from the periosteum. It can be seen with x-rays, but is extremely rare in developed countries because osteomyelitis is rarely left untreated.

Ipsilateral Referring to the same side of the body.

Iron-deficiency anemia A microcytic hypochromic anemia caused by inadequate supplies of iron needed to synthesize hemoglobin.

Irritable bowel syndrome The presence of alternating diarrhea and constipation accompanied by abdominal cramping in the absence of any identifiable pathologic process in the gastrointestinal tract.

Ischemia Inadequate blood flow through the arterial system, producing tissue hypoxia.

Ischemic hepatitis A lack of blood or oxygen supply to the liver that causes injury to liver cells.

Isoforms Isomeric forms of the same protein, with slightly different amino acid sequences but with the same function.

Isoimmunity The condition occurring when an individual's immune system reacts against

antigens on tissues from other members of the same species, such as a blood transfusion reaction in which a person with type A blood reacts against a transfusion with type B blood.

Isolated systolic hypertension An elevation in systolic blood pressure greater than 140 mm Hg without an increase in diastolic blood pressure. Most commonly occurs in the elderly.

Isolated urinary tract infection A first infection or an infection that occurs more than 6 months after a previous infection.

Isometric contraction A contraction in which no movement takes place and the muscle maintains its specific length. For example, holding a weight in the hand with the elbow flexed produces an isometric contraction.

Isotonic fluid Fluid that has the same particle concentration (osmolality) as normal body fluid.

Isovolumic contraction The early phase of systole, in which the myocardial muscle fibers have begun to shorten and all valves in the heart are closed.

J

Jaundice Yellowness of skin associated with elevated serum bilirubin levels.

Joint A point of contact between bones. Also called *articulation.*

Joint capsule A dense layer of connective tissue surrounding a synovial joint. The capsule is solidly attached to the periosteum of the adjacent bony components. The joint capsule provides strength to the joint and, through its neural receptors, detects motion, compression, tension, vibration, and pain.

Juxtaglomerular apparatus The collection of macula densa cells in the distal convoluted tubule and the afferent and efferent arterioles and also the juxtaglomerular cells located around the arterioles; they work together to control glomerular filtration rate.

Juxtamedullary nephron A nephron with long loops of Henle that extend deep into the medulla and create a concentrated interstitium via the countercurrent mechanism.

K

Keratin Tough, water-repellent protein produced by keratinocytes and found in hair, nails, and horny tissue.

Keratinocyte One of several types of epithelial cells. Keratinocytes are able to synthesize DNA and produce keratin.

Keratosis Any skin lesion in which there is overgrowth and thickening of the cornified epithelium.

Kernicterus An abnormal toxic accumulation of bilirubin in central nervous system tissues caused by hyperbilirubinemia.

Ketoacidosis Acidosis accompanied by an accumulation of ketones; it is usually associated with poorly controlled type 1 diabetes mellitus.

Ketone bodies The result of fatty acids in the liver that are transformed to acetyl coenzyme A, which is then processed into one of three compounds known as *ketone bodies.*

Kilocalorie The unit of measure for the energy value of foods.

Kinin A vasoactive peptide produced during inflammation and injury.

Koilonychia Dystrophy of the fingernails in which the nails are concave; also known as *spoon nail.*

Korotkoff sounds Sounds heard during auscultation of arterial blood pressure. As the pressure in the blood pressure cuff is released, blood begins to flow turbulently through the artery, producing Korotkoff sounds.

Kyphoscoliosis An abnormal condition characterized by an anteroposterior and lateral curvature of the spine.

L

Lactation Formation and secretion of milk from the breasts for the nourishment of the infant.

Lactic acidosis An increase in the anaerobic production of lactate, which, when released into the bloodstream, creates a condition of metabolic acidosis.

Laminar flow Flow of air or fluid in which there is no turbulence and the direction of flow is linear and parabolic.

Langerhans cell One of several types of epithelial cells. Langerhans cells are thought to have a role in immunologic reactions that affect the skin and may serve as a defense mechanism for the body.

Laparoscopic cholecystectomy Surgical removal of the gallbladder using an optical scope and instruments inserted through four small abdominal incisions.

Laparoscopy Examination of the abdominal cavity via a small incision to permit the insertion of a variety of optical scopes for diagnosis or therapy.

Latency period A period when there is no apparent change in status even though a process has begun to occur.

Leiomyoma Benign neoplasm of the smooth muscle of the uterus that is characteristically firm, well circumscribed, and round. Uterine leiomyomas usually appear and exhibit growth activity during the reproductive years. Fibroid tumors.

Leprosy A chronic infectious disease of the skin caused by the intracellular bacillus *Mycobacterium leprae.*

Leptomeninges The combined structures of the pia mater and arachnoid mater.

Lesion A general term for a demonstrable structural change produced in the course of a disease. Lesions may be evident at a gross or microscopic level.

Leukemia A malignant disease of bone marrow stem cells with accumulation of immature blasts in the marrow and peripheral blood.

Leukocyte A cell that mediates immune function. Leukocytes protect the body by phagocytosis of microorganisms and production of antibodies and memory cells. Also called *white blood cell.*

Leukoderma Patch of depigmentation, also called *vitiligo.*

Leukopenia A deficiency of white blood cells in the peripheral circulation, which is usually indicative of bone marrow failure.

Leukotrienes A class of biologically active compounds produced by leukocytes that trigger allergic and inflammatory reactions similar to those of histamine.

Leydig cell An interstitial cell in the testes that produces and secretes testosterone.

Libido Sexual drive; feeling of sexual desire.

Lichen planus A relatively common, chronic, pruritic disease involving inflammation and papular eruption of the skin and mucous membranes.

Ligament A dense connective tissue with parallel-fibered collagenous tissues designed to connect bone to bone.

Ligands A general term for a molecule, such as a hormone, neurotransmitter, or drug, that attaches to a receptor protein; usually involved in cell-to-cell communication.

Limbic system A group of structures surrounding the corpus callosum that process various emotional feelings.

Lipid bilayer A double layer of lipid molecules that forms cellular membranes, including the plasma membrane, organelle membranes, and vesicles.

Lipodystrophy A group of conditions caused by defective metabolism of fat, resulting in atrophy of subcutaneous fat.

Lipolysis Production of free fatty acids resulting from the breakdown of fat in adipose tissue.

Lipoproteins A group of biomolecules composed of differing amounts of cholesterol, triglyceride, and protein, such as high-density lipoprotein (HDL) and low-density lipoprotein (LDL).

Lithotripsy Mechanical or chemical fragmentation of a calculus (stone).

Longitudinal fracture Fracture in which a bone is split along its length.

Loose association Flow of thought in which ideas shift from one subject to another in a completely unrelated or noncohesive way. When severe, speech may be unintelligible.

Lower esophageal sphincter The circular band of muscular tissue at the lower end of the esophagus. It serves to prevent the highly acidic gastric contents from moving in a retrograde motion back into the esophagus.

Lumpectomy Surgical removal of only the malignant tumor, or "lump," from an affected breast.

Lung compliance A measure of the ease of expansion of the lungs and thorax determined by pulmonary elasticity (volume divided by pressure).

Lunula "Little moon"; the crescent-shaped white area nearest the root of the nail body.

Lupus erythematosus A chronic inflammatory disease affecting many systems of the body. It is an example of an autoimmune disease.

Lusitropy Rate of relaxation of cardiac muscle and chambers.

Lyme disease An infectious, immune-mediated multisystem disease caused by a tickborne spirochete. It is characterized by an erythema migrans rash in which the area of redness begins at the site of the tick bite.

Lymphadenopathy A pathologic lymph node enlargement, which is usually painless and may be associated with malignancy, especially lymphoma. Lymphadenopathy must be distinguished from normal reactive lymph node enlargement in response to infection. Reactive nodes are usually tender and are situated "downstream" from a site of infection.

Lymphedema Swelling produced by an obstruction of lymphatic flow.

Lymphoblast A large, immature cell that develops into a lymphocyte. When found in the circulation lymphoblasts are indicative of leukemia.

Lymphocyte A white blood cell derived from the lymphoid stem cell that is not affected by diseases of the myeloid stem cell. Lymphocytes are of three basic types: T, B, and natural killer cells.

Lymphogranuloma venereum Highly contagious systemic infection caused by a number of strains of *Chlamydia.* It has progressive stages of development in which an initial lesion forms, and systemic disease occurs after dissemination via the lymphatic system.

Lymphoid group Dermal cells consisting of lymphocytes commonly found in inflammatory lesions of the skin.

Lymphopoiesis Formation of lymphocytes in the bone marrow.

Lysis Destruction or breakage of a cell membrane or molecule.

Lysosome An organelle containing hydrolytic enzymes that function to digest intracellular materials.

Lysozyme An enzyme secreted by macrophages and neutrophils to control foreign particle activity.

M

M line Marks the center of the sarcomere A band and the midpoint of myosin filaments in striated muscle.

Macrocyte An abnormally large erythrocyte.

Macromolecule A molecule of colloidal size, such as a protein.

Macrophage A mature monocyte that migrates from the blood vessels to sites in the tissues. Macrophages are powerful phagocytes and secrete a number of cytokines that stimulate inflammation.

Major depressive disorder A mood disorder characterized by persistent dysphoria, anxiety, irritability, and fear.

Major histocompatibility complex (MHC) The regions on chromosome 6 that contain the genes for MHC proteins. Class I proteins are present on virtually all nucleated cells. Class II proteins are found mainly on professional antigen-presenting cells: B cells, macrophages, and dendritic cells.

Malabsorption Failure of the gastrointestinal tract to absorb or normally digest one or more dietary constituents.

Maladaptation Ineffective, inadequate, or inappropriate change in response to new or altered circumstances.

Malignant tumor A type of tumor that has a tendency to invade local tissues and spread to distant sites (metastasis). Malignant tumors are generally poorly differentiated and are associated with a poor prognosis if not promptly managed.

Mallory–Weiss syndrome Mild to massive bleeding due to a tear in the mucosa or submucosa of the cardia or lower esophagus. The tear is usually longitudinal and is caused by forceful or prolonged vomiting during which the upper esophageal sphincter fails to relax.

Malnutrition A disorder of nutrition that may result from an unbalanced, insufficient, or excessive diet or from impaired absorption.

Malunion An imperfect union of previously fragmented bone or other tissue.

Mania Also called *full mania,* it is characterized by an overwhelming increase in energy and drive manifested as nonstop activity, grandiose thinking, impulsivity, euphoria, impaired judgment, acting-out behaviors, and hypersexuality.

Mass lesion Any lesion in the cranium that behaves like a space-occupying mass. Mass lesions tend to progress and cause signs and symptoms of increased intracranial pressure.

Mast cell Also called *histiocytic cell.* Mast cells have intracytoplasmic basophilic metachromatic granules containing heparin and histamine. The normal skin contains relatively few mast cells, but their number is increased in many different skin conditions, particularly the itching dermatoses.

Mastectomy Surgical removal of a breast. See *modified radical mastectomy* and *radical mastectomy*.

McBurney point Located in the lower-right quadrant of the abdomen; situated in the normal area of the appendix midway between the umbilicus and the anterior iliac crest.

Mean arterial pressure The average pressure in the arterial system through the cardiac cycle. It is calculated by adding the systolic pressure to two times the diastolic reading and dividing the sum by 3.

Measles Known as *hard measles, 7-day measles,* or *rubeola*. This is a communicable viral disease caused by paramyxovirus of the genus *Morbillivirus* with a characteristic macular and blotchy rash; sometimes the macules become confluent.

Mediastinum The area of the chest between the sternum and vertebral column and between the lungs. Mediastinal structures include the trachea, esophagus, aorta, heart, and lymph nodes.

Medullary cavity Central cavity in longer bones.

Megacolon Abnormal massive dilation of the colon that may be congenital, toxic, or acquired. There is a risk of rupture and peritonitis.

Megakaryocyte A large bone marrow cell that sheds platelets into the circulation from its cytoplasm.

Megaloblastic dysplasia Abnormal development of large red blood cells and nonlymphocytic bone marrow cells.

Megaloblasts Large, abnormal, hematopoietic bone marrow cells.

Meiosis A type of cell division that results in daughter cells with one-half the normal number of chromosomes. Meiosis occurs in gonadal germ cells: egg and sperm.

Melanin Dark pigment found in melanocytes that gives color to hair and skin.

Melanocyte One of several types of epithelial cells. Melanocytes contribute color to the skin and serve to filter ultraviolet light.

Melena Tarry, black feces due to the action of gastrointestinal secretions on blood in the intestine.

Membranous urethra The urethral segment that passes through the muscular layers of the urogenital diaphragm.

Memory cells T and B lymphocytes that mediate immunologic memory through development of clones of long-lived cells.

Menarche The first menstrual period at the time of puberty, usually occurring around age 12 years in North America.

Meniere disease A chronic disease of the inner ear characterized by recurrent episodes of vertigo and hearing loss.

Meninges Membranes surrounding the brain and spinal cord, which include the dura mater, arachnoid, and pia mater.

Meningitis Any infection or inflammation of the membranes covering the brain and spinal cord.

Meningocele Hernial protrusion of meninges through a neural tube defect in the skull or vertebral column.

Meniscus Curved, fibrous cartilage located in the knee and other joints. Menisci facilitate rotation at the knee by allowing better contact of the tibial surfaces with the femoral condyles.

Menopause A process by which the supply of ovarian follicles and estrogen hormones declines, usually beginning between the ages of 45 and 55 years.

Menorrhagia An increase in the amount or duration of menstrual bleeding, usually resulting from a lesion of the female reproductive organs and hormones.

Menstrual cycle The rhythmic pattern of changes in hormonal secretions and in sexual organs occurring approximately every 28 days during a female's reproductive years. The cycle culminates in the production of an ovum and the preparation of the uterus for implantation of a fertilized ovum.

Merkel cells One of several types of epithelial cells. Merkel cells consist of free nerve endings attached to modified epidermal cells. It is generally agreed that Merkel cells function as touch receptors.

Mesoblastic nephroma Benign congenital renal tumor.

Mesonephros One of three distinct stages in the development of the renal system; it begins at about the fourth to fifth week of gestation.

Metabolic acidosis Any of the types of acidosis resulting from accumulation in the blood of noncarbonic, nonvolatile acids; characterized by a low HCO_3^- concentration.

Metabolic alkalosis A disturbance in which the acid–base balance shifts to alkaline because of uncompensated loss of acids, ingestion or retention of excess base, or depletion of potassium.

Metabolic syndrome A disorder of metabolism including at least three of the following: abdominal obesity, hypertriglyceridemia, low level of high-density lipoproteins, hypertension, and high fasting plasma glucose level. It is associated with an increased risk for development of diabetes mellitus and cardiovascular disease.

Metabolism Synthesis and breakdown of molecules in a living organism. Metabolism involves both the use (anabolism) and the release (catabolism) of energy.

Metanephros The final stage of development of the renal system; begins the fourth week of gestation when the ureteral bud grows out of the mesonephric duct.

Metaplasia Transformation of one kind of tissue to another fully differentiated tissue.

Metastasis Dissemination of cancer cells from the location of origin to other distant areas in the body.

Methemoglobin A transformation product of oxyhemoglobin that is formed when the iron of the hemoglobin molecule is oxidized to the ferric state (Fe^{3+}).

Metrorrhagia Bleeding between menstrual periods, usually the result of slight physiologic bleeding from the endometrium during ovulation. It may also result from other causes such as uterine malignancy, cervical erosions, endometrial polyps, or estrogen therapy.

Microbial adherence The ability of the microorganism to latch onto and gain entrance into the host.

Microcytic Referring to an abnormally small erythrocyte.

Micropenis A small, normally formed penis with an engorged length more than two standard deviations below the mean, or an engorged length less than 2.5 cm.

Micturition Urination.

Mineralocorticoids A class of steroid hormones secreted by the adrenal cortex that regulate the mineral salts (electrolytes) and water balance in the body.

Minimal change disease A glomerular kidney disorder characterized by subtle changes in glomerular structure.

Mitochondrion A membrane-bounded organelle that carries out oxidative phosphorylation to synthesize most of the adenosine triphosphate in a eukaryotic cell.

Mitosis A type of cell division that results in daughter cells with chromosomes that are identical to the parent cell. Mitosis occurs in somatic cells.

Mitral valve The cardiac valve that lies between the left atrium and left ventricle. The valve is normally closed during ventricular systole and open during ventricular diastole. Mitral valve closure contributes to heart sound S_1. Also called the *bicuspid atrioventricular valve*.

Mixed acid–base imbalance A combined disturbance of acid–base balance in which a primary respiratory disorder and a primary metabolic disorder coexist.

Mixed incontinence Loss of bladder control with symptoms of both stress and urge incontinence.

Mobility The ability to achieve purposeful movement.

Modified radical mastectomy Surgical removal of the breast accompanied by dissection of a portion of the axillary lymphatic system.

Molluscum contagiosum A viral skin disease with two forms. One form affects children and is spread through indirect contact; the other is sexually transmitted and occurs in young adults. It is characterized by pink and white lesions on the genitalia with an exudative core.

Mongolian spots Caused by selective pigmentation. They usually occur on the buttocks or sacral area and are commonly seen in Asian Americans and African Americans.

Monoclonal antibodies Identical antibodies produced in the body or in a laboratory from a single clone of B lymphocytes.

Monocyte An immature circulating macrophage.

Monosaccharide A simple sugar (e.g., glucose, fructose).

Monosomy Having only one member of a homologous chromosome pair, as in monosomy X, also called *Turner syndrome.*

Monounsaturated fatty acid A fatty acid with one double bond.

Mood Sustained expression of an emotion that affects one's outlook.

Mood disorder A disturbance of mood that may be caused by either organic damage to the brain or chemical alterations in neurotransmission. Mood disorders may also have no known biological basis.

Mood swings Oscillation between periods of euphoria (elevated mood) and depression or anxiety.

Morphogen A substance that triggers growth, proliferation, and differentiation of cells in a concentration-dependent manner.

Morphogenesis Arrangement of cells in a particular order during the development of complex organisms.

Morphologic changes Structural and associated functional alterations in cells or tissues that are either characteristic of the disease or diagnostic of the etiologic process.

Motor end-plate The points of contact between an α-motor neuron and the skeletal muscle cells it innervates.

Mucocutaneous candidiasis Candidal infection of the mucous membrane and the skin.

Mucosal edema Swelling of the membranes of the respiratory, gastrointestinal, or urogenital systems; usually a consequence of inflammation.

Müllerian ducts Genital structures, also called *paramesonephric ducts* because they develop alongside the mesonephric ducts. A pair of embryonic ducts that become the fallopian tubes, uterus, and vagina in females.

Multifactorial Pertaining to or characteristic of any condition or disease resulting from the interaction of many factors.

Multipennate muscle A muscle with several central tendons toward which the muscle fibers converge like the barbs of feathers.

Multiple myeloma (plasma cell myeloma) A malignant disorder of antibody-secreting plasma cells that produce large quantities of monoclonal antibodies and have a predilection to settle in the skeleton, where osteoclastic bone lesions are produced.

Multiple sclerosis A chronic demyelinating disease of the central nervous system that causes significant disability in young adults. It is thought to be an autoimmune disorder that results in inflammation and scarring (sclerosis) of the myelin sheaths covering nerves.

Muscular dystrophy Term referring to a group of genetically determined myopathies characterized by progressive degeneration of muscle fibers.

Mutagen A physical or chemical agent capable of causing alterations in an organism's DNA by inducing mutations.

Mutation A heritable change in the nucleotide sequence of a chromosome; it is passed on to daughter cells when the cell divides.

Myasthenia gravis A chronic autoimmune disease affecting the neuromuscular function of voluntary muscles and characterized by profound muscle weakness and fatigability.

Mycosis Infection caused by a fungus.

Myeloid group Cells of the dermis consisting of polymorphonuclear leukocytes and eosinophilic leukocytes. These cells occur commonly with allergic dermatoses.

Myelomeningocele or meningomyelocele Hernial protrusion of meninges, spinal fluid, and a portion of the spinal cord with its nerves through a defect in the vertebral column.

Myocardial infarction Localized area of cardiac necrosis most often associated with coronary heart disease and sudden acute occlusion of a coronary artery by a thrombus.

Myocardium The middle layer of the heart; composed of cardiac muscle tissue.

Myopia A condition of nearsightedness caused by elongation of the eyeball or by an error in refraction so that parallel rays are focused in front of the retina.

Myosin A cytoskeletal protein that comprises the thick filament of the muscle sarcomere in skeletal and cardiac muscle. It is also present as a contractile filament in other types of cells. Myosin binds with actin during muscular contraction.

Myositis ossificans An abnormal calcification within a muscle.

Myxedema Nonpitting edema caused by advanced hypothyroidism in adulthood.

N

Narcissism Self-absorption; excessive self-love.

Natural killer cells A lymphocyte that is capable of binding to and killing virus-infected cells and some tumor cells by releasing cytotoxins; is innate and does not have T-cell or B-cell receptors.

Necrosis Death and degradation of body cells or tissues in response to irreversible injurious events.

Necrotizing enterocolitis A disorder occurring most often in premature infants (less than 34 weeks' gestation) and infants with low birth weight (less than 5 lb or 2.25 kg). This disorder is characterized by diffuse or patchy intestinal necrosis accompanied by sepsis.

Necrotizing inflammation Also called *vasculitis.* This response can occur when antigen and antibody react in blood vessels in the skin. Necrotizing inflammation can be caused by drug allergies; disorders such as systemic lupus erythematosus, rheumatoid arthritis, and glomerulonephritis; and certain infectious diseases such as hepatitis B.

Negative feedback A term used to explain homeostatic mechanisms. Negative feedback causes the controller to respond in a manner that opposes or negates deviation from the normal level (set point). Many body systems operate on the principle of negative feedback.

Negative nitrogen balance The condition of protein catabolism (breakdown) exceeding daily protein intake and synthesis.

Negative symptoms (schizophrenia) Symptoms of schizophrenia that are thought to be mediated by dopamine D_1 receptors in the brain. Drugs that block D_1 receptors may alleviate some of the negative symptoms, which include social withdrawal, flat affect, poverty of speech, ritualistic posturing, and autism.

Neoantigen formation A new specific antigen that develops in a tumor cell.

Neologism A new word, often created by combining syllables of other words; or a word given special or private significance.

Neoplasia New growth. The term implies an abnormality of cellular growth and may be used interchangeably with the term *tumor*.

Neoplasm A new and abnormal proliferation of cells. If malignant, the growth infiltrates tissue, metastasizes, and often recurs, even after attempts at surgical removal.

Nephralgia Renal pain.

Nephrectomy The surgical removal of a kidney.

Nephritic syndrome A group of signs and symptoms of a urinary tract disorder, including hematuria, hypertension, and renal failure.

Nephroblastoma (Wilms tumor) The most common childhood malignant kidney tumor, resulting from a defect on chromosome 13.

Nephrogenic rests Remnants of embryonic tissue found in or around the kidney retained after the period of embryonic development. Nephrogenic rests are sometimes precursors (forerunners) of Wilms tumor.

Nephrolithiasis The presence of a stone or calculus anywhere in the urinary tract.

Nephroma Tumor of the kidney or area of the kidney.

Nephron Functional unit of the kidney composed of epithelial cells forming the glomerulus, proximal convoluted tubule, loop of Henle, distal convoluted tubule, and collecting duct.

Nephropathy A pathologic process in the kidney, including inflammatory, degenerative, and sclerotic conditions. Many disorders can lead to nephropathy, as in diabetic nephropathy, toxic nephropathy, ischemic nephropathy, and obstructive nephropathy.

Nephrotic syndrome A common set of symptoms (hypoproteinemia, generalized edema) caused by damage to the glomeruli in which proteins cross the glomerulus and are lost in the urine at a rate of >3.5 g/day.

Nephrotoxic Poisonous to the kidney.

Neural thread protein Microscopic protein thread that binds and helps stabilize microtubules (the cell's internal support structure or skeleton). In Alzheimer disease, the threads become chemically altered and twist into paired helical filaments known as *neurofibrillary tangles.* See *tau protein*.

Neurocrine Secretion of hormone-signaling molecules into the bloodstream from neurons.

Neurofibrillary tangle Abnormal bundle of twisted threads inside a nerve cell that is the collapsed remains of the neuron's microtubules, which normally provide intracellular transport.

Neurogenic bladder Bladder dysfunction caused by a lesion at any level in the nervous system.

Neurogenic shock Often called "fainting," neurogenic shock may be caused by severe pain, fear, an unpleasant sight, or other strong stimuli that overwhelm the usual regulatory capacity of the nervous system. Blood pools in the extremities, resulting in a drop in cardiac output.

Neuroglia, glia A group of cell types, including astrocytes, microglia, ependymal cells, and oligodendrocytes, that support nerve cells and do not themselves conduct action potentials.

Neurohormones A hormone secreted by a specialized neuron into the bloodstream, the cerebrospinal fluid, or the intercellular spaces of the nervous system.

Neuropathic osteoarthropathy A neurologic disease that leads to bone abnormalities and joint involvement. The mechanics of disease development are probably a combination of neurovascular and neurotraumatic processes.

Neurovascular injury An injury that affects the nerves that control the caliber of blood vessels.

Neutropenia A type of leukopenia in which the absolute neutrophil count is below 500 cells/μl. Neutropenia is associated with a high risk of bacterial sepsis.

Neutrophilia A high blood neutrophil count.

Neutrophil/neutrophilic granulocyte A cell that contains small lysosomal granules and a segmented nucleus with two to five lobes. These cells compose 60% to 70% of circulating leukocytes.

Nevus Congenital discoloration of a circumscribed area of the skin; commonly called *mole* or *birthmark*.

Nociception Activation of nociceptors by potentially tissue-damaging stimuli resulting in the perception of pain by the central nervous system. Nociception includes the processes of receptor transduction, signal transmission, perception, and signal modulation.

Nociceptor Pain receptor.

Nocturia Excessive urination at night.

Nonarticular rheumatism A common soft tissue syndrome in children. The most common symptom is nocturnal pain that usually occurs in the calves, shins, and thighs. Also called *growing pain*.

Noncontractile tissue Soft tissue that possesses no ability to contract or relax; this includes the joint capsule, ligament, bursa, fascia, dura mater, and nerve root. Also called *inert tissue.*

Nondisjunction The failure of homologous chromosomes to separate normally during meiosis or mitosis, resulting in unequal distribution of chromosomes to daughter cells.

Nondisplaced fracture A fracture in which the fragments remain in alignment and position.

Non-Hodgkin lymphoma A varied group of malignant disorders of lymph node cells involving B cells, T cells, and natural killer cells. In comparison with Hodgkin disease, these lymphomas tend to spread unpredictably and metastasize early, and thus carry an overall worse prognosis.

Non–Q-wave infarct A subendocardial infarct affecting only the inner third to half of the ventricular wall and generally associated with less severe symptoms. No pathologic Q-waves are observed on the electrocardiogram.

Non-STEMI A myocardial infarction that does not have ST-segment elevation observed on the electrocardiogram.

Nonunion Pertaining to a fractured bone that fails to heal properly.

Norepinephrine A major monoamine neurotransmitter of the sympathetic nervous system that is involved in cardiopulmonary, vascular, and gastrointestinal regulation and has been implicated in the development of mood disorders.

Normochromic Pertaining to a blood cell having normal color resulting from the presence of an adequate amount of hemoglobin, measured by the mean corpuscular hemoglobin concentration (MCHC).

Normocytic Description of a typical adult red blood cell of average size, measured by the mean corpuscular volume (MCV).

Nuclei A cluster of neuronal cell bodies located in the central nervous system. Also refers to the nucleus within individual cells.

Nucleocapsid The core of the human immunodeficiency virus (HIV), which contains two strands or chains of RNA, protein, and enzymes.

Nucleotide A biomolecule composed of a purine or pyrimidine base linked to a ribose or deoxyribose sugar, with one or more phosphate groups attached to the sugar. DNA and RNA are polymers of nucleotides.

Nucleus A cellular organelle that contains chromosomal DNA.

Nutritional screening A method for quickly determining the nutritional status of an individual from a selected group of anthropometric and biochemical tests.

Nutritional status The state of an individual's nutrition, resulting from the consumption and utilization of nutrients to meet metabolic needs.

Nystagmus Involuntary, rapid, rhythmic movements of the eyeball; these occur commonly in a horizontal direction, but can also occur in a vertical or a rotational direction. May indicate inner ear or cranial nerve disorders.

O

Oblique fracture Fracture resulting from a rotational force; however, unlike a spiral fracture, the break is along an oblique course (45-degree angle) and does not rotate around the entire bone.

Obsession A powerful, persistent, intrusive thought, impulse, or image that dominates the mental life of the individual to the extent of seriously interfering with normal living.

Obsessive-compulsive disorder (OCD) An anxiety disorder characterized by recurrent and persistent thoughts, ideas, and feelings or repetitive acts.

Obstruction Refers to an anomaly that compromises or prevents flow because of abnormal narrowings. Stenosis or atresia (failure to develop) of valves and coarctation of the aorta are examples of the most common cardiac obstructive defects.

Obstructive sleep apnea A form of sleep apnea involving a physical obstruction in the upper airways.

Occupational asthma Allergic asthma from exposure to protein allergens in the work environment.

Odynophagia A severe sensation of burning, squeezing pain with swallowing; may accompany dysphagia, or difficulty with swallowing.

Ogilvie syndrome The idiopathic syndrome of intestinal pseudoobstruction; may result in megacolon.

Olfactory Pertaining to the sense of smell.

Oligomenorrhea Infrequent menstruation, usually the result of failure to ovulate as a result of dysregulated hormonal function.

Oligosaccharide A compound consisting of 2 to 10 joined monosaccharide (sugar) units.

Oliguria Urine output of less than 500 mL/day.

Omphalocele Congenital anomaly in which a herniation of viscera at the base of the umbilical cord is present; requires surgical correction.

Oncocytoma Benign renal tumor consisting of large eosinophilic cells that have granular cytoplasm and round, uniform nuclei.

Oncogene A gene associated with the initiation of cancerous behavior in a cell. Oncogenes typically start as normal growth-promoting genes called *proto-oncogenes* that become overactive as a result of mutations.

Onycholysis Separation of the nail from its nail bed.

Oogonia The cells present in the female ovaries during prenatal development that ultimately develop into ova. The entire lifetime supply of ova is established prenatally; no new oogonia arise after birth.

Open fracture Fracture occurring when bone is broken and an external wound leads to the fracture site.

Ophthalmia neonatorum Purulent gonococcal conjunctivitis and keratitis in the newborn resulting from exposure of the infant's eyes to infected maternal secretions during the passage through the vagina at birth.

Ophthalmoscopic examination Examination of the structures of the eye, both extraocular and intraocular, using an ophthalmoscope.

Opioid Any of a group of drugs with an affinity for opioid receptors in the central nervous system. Morphine is the standard opioid with which others are compared for characteristics and potency. Opiate; synthetically produced drugs with actions similar to opioids.

Opportunistic infection An infection caused by organisms that are usually nonpathogenic but that become pathogenic because of decreased function of the immune system.

Opsonization The process of proteins, usually antibodies or complement fragments, binding to an antigen, making the antigen easier for phagocytic cells to locate. Phagocytic cells have receptors for opsonins.

Organelles The membrane-bound structures in the cell cytoplasm, including nucleus, mitochondria, endoplasmic reticulum, and Golgi apparatus.

Orthopnea Difficulty with breathing that is instigated or exacerbated by lying supine.

Orthostatic (postural) hypotension A form of low blood pressure that occurs after positional change from supine to standing. Diagnosed by an increase in heart rate of more than 15% and/or a decrease in either systolic blood pressure by more than 15 mm Hg or diastolic blood pressure by more than 10 mm Hg.

Osmolality A measure of degree of concentration; number of particles per kilogram of solvent.

Osmosis Movement of water across a semipermeable membrane to equalize the particle concentration of the fluid on both sides of the membrane.

Ossicle One of several small bones in the middle ear responsible for transmitting sound waves to the inner ear.

Osteoarthritis A common degenerative joint disease characterized by progressive loss of articular cartilage and by formation of new bone from subchondral bone at joint margins.

Osteoblast A bone-forming cell that is derived from the embryonic mesenchyme and, during the early development of the skeleton, differentiates from a fibroblast to function in the formation of bone tissue.

Osteochondroma A benign bone tumor consisting of bone and cartilage.

Osteoclast A cell responsible for bone resorption.

Osteocyte A mature bone cell.

Osteonecrosis The destruction and death of bone tissue, such as from ischemia, infection, malignant neoplastic disease, or trauma.

Osteoid osteoma A painful but benign bone-forming tumor that is often found in the cortex of the tibia and femur.

Osteomalacia An abnormal condition of lamellar bone, characterized by a loss of calcification of the matrix and consequent softening of the bone. The condition is the result of an inadequate amount of phosphorus and calcium available in the blood for mineralization of the bones.

Osteomyelitis A severe pyogenic infection of bone and local tissue that requires intensive management.

Osteon The basic unit of bone; also called the *haversian system.*

Osteoporosis A common metabolic bone disease in which reduction in bone mass results from bone resorption proceeding at a rate faster than that of new bone formation. Usually defined as a bone mass 2.5 standard deviations below the mean.

Osteoprogenitor A type of bone stem cell that lines bone surfaces.

Osteosarcoma A malignant bone-forming tumor and the most common primary malignant bone tumor that develops in the metaphyseal region of long bones.

Otosclerosis A hereditary condition of unknown cause in which irregular ossification occurs in the ossicle bones of the middle ear, causing hearing loss.

Ototoxic Damaging to the structures of the inner ear, usually referring to medications and toxins.

Ovarian cyst A sac on an ovary that contains fluid or semisolid material. It may develop at any time between puberty and menopause; the cause is presently unknown.

Overflow incontinence Loss of bladder control associated with urinary retention and bladder distention due to obstruction, detrusor underactivity or inactivity, or sphincteric malfunction.

Oviduct Another term for *fallopian tube.* Each oviduct runs laterally from the uterus to the uterine end of the ovary. The free end of the oviduct adjacent to the ovary is called the *infundibulum.*

Oxidative phosphorylation An adenosine triphosphate (ATP)-generating process in which oxygen serves as the final electron acceptor. The process occurs in mitochondria and is the major source of ATP generation in aerobic organisms.

Oxygen consumption The amount of oxygen used by the tissues in 1 minute, usually expressed as $\dot{V}o_2$.

Oxygen delivery The amount of oxygen delivered to the tissues each minute. Oxygen delivery ($\dot{D}o_2$) is calculated by multiplying cardiac output and arterial oxygen content.

Oxytocin Hormone secreted by the posterior pituitary. It causes uterine contraction and is thought to have a major role in promoting increased uterine contractility during parturition. After birth, it is secreted in response to suckling by the infant and stimulates the release of milk, called the *milk ejection reflex.*

P

$Paco_2$ Partial pressure of dissolved carbon dioxide in arterial blood; an indicator of the effectiveness of respiratory excretion of carbonic acid.

Paget disease Also called *osteitis deformans;* a slowly progressive metabolic bone disease characterized by an initial phase of excessive bone resorption followed by a reactive phase of abnormal excessive bone formation.

Pallor Paleness of skin, nail beds, lips, and conjunctivae. A possible symptom of anemia.

Pancreas Gland located in the abdomen; has both endocrine and exocrine functions. The endocrine pancreas produces insulin, glucagon, and somatostatin.

Pancreatitis Inflammation of the pancreas; may be acute or chronic.

Pancytopenia Decreased production of red blood cells, white blood cells, and platelets.

Pandemic An epidemic that affects large geographic regions, possibly spreading worldwide.

Panhypopituitarism A condition of deficiency in most or all pituitary hormones.

Panic disorder Psychiatric disorder characterized by recurrent, unexpected episodes of acute anxiety, fear, and panic; often accompanied by the subject's belief that he or she is having a heart attack, is unable to breathe, is losing control, or is dying.

Panmyelosis A pathologic condition characterized by excessive proliferation of bone marrow cells of all types.

Papillary layer One of two layers of the dermis. The papillary layer consists of bumps (papillae) that project into the epidermis.

Papillary muscles Muscles within the ventricles that connect the chordae tendineae to the ventricular wall. Papillary muscles contract during ventricular systole to place tension on the valves and prevent backflow through them.

Paracrine Referring to hormonelike chemicals, the target cell of which is located near to the cell secreting the chemical.

Parametritis An infection of the connective tissue between the broad ligaments underlying the female reproductive organs.

Paraneoplastic syndrome A cluster of systemic conditions associated with cancer, such as hypercalcemia, hyponatremia, or Cushing syndrome.

Paraphimosis Painful constriction of the glans penis by the foreskin, which has been retracted behind the corona.

Parasite One of a variety of protozoa (single-celled animals), nemathelminths (roundworms), platyhelminths (flatworms), and arthropods (invertebrate animals with jointed appendages). Parasites depend on another organism for survival.

Parkinsonism Parkinson disease symptoms; a neurologic disorder characterized by tremor; muscle rigidity; hypokinesia; a slow, shuffling gait; and difficulty in chewing, swallowing,

and speaking caused by various lesions in the extrapyramidal motor system.

Paroxysm A sudden outburst or change from the norm, as in a sudden burst of electrical activity seen on electroencephalography, as occurs with seizure activity; or a sudden fit of coughing or burst of fast heart rate.

Paroxysmal nocturnal dyspnea A sudden severe feeling of suffocation; usually occurs at night and awakens the person from sleep.

Partial seizure A seizure in which part of the brain surface is involved in the seizure. The person retains consciousness although it may be impaired.

Parturition The process by which an infant is born.

Passive immunity A form of acquired immunity resulting from antibodies that are transmitted naturally through the placenta to a fetus or through the colostrum to an infant, or artificially by injection.

Pathogen An agent that causes disease.

Pathogenesis Development or evolution of disease. A description of the pathogenesis includes the processes that occur in the body from the initial stimulus to the ultimate expression of manifestations of the disease.

Pathologic Pertaining to a condition that is caused by or involves a disease process.

Pathology Study of the causes, characteristics, and effects of disease.

Pathophysiology The study of the biological and physical manifestations of disease, including etiology, pathogenesis, clinical manifestations, and treatment implications.

Pauciarticular onset Affecting four or fewer joints; used in association with juvenile rheumatoid arthritis.

Pedigree Genetic lineage or family history of traits; used to trace the pattern of inheritance.

Pelvic inflammatory disease Any acute or subacute recurrent or chronic infection of the oviducts and ovaries with involvement of the adjacent reproductive organs.

Pemphigus A group of disorders including vulgaris, vegetans, foliaceus, and erythematosus. The pemphigus group disorders are characterized by bullous eruptions (blisters) thought to be caused by autoimmune reactions.

Pemphigus vulgaris Included in the pemphigus group of disorders, pemphigus vulgaris has the worst prognosis. Bullae can erupt on the skin and mucous membranes (e.g., esophagus), and toxemia and infection can cause death if proper treatment (cortisone) is not administered.

Penile urethra The longest segment of the male urethra, extending about 15 cm in length from the membranous urethra to the external meatus.

Penis Male organ of copulation and urinary excretion.

Penumbra The margin or fringe surrounding a central part. In the case of stroke, a penumbra of viable tissue surrounds the necrotic core that can survive if optimal conditions exist and if it is not subject to further insults.

Peptic ulcer disease Disorder of the upper gastrointestinal tract caused by the action of acid and pepsin. This disorder may include injury to the mucosa of the esophagus, stomach, or duodenum and may range from a slight mucosal injury to severe ulceration.

Percutaneous Referring to a procedural approach that traverses the skin and is less traumatic than open surgical methods.

Perfusion The delivery of blood flow to a specific organ or an area of the body.

Pericardial effusion The escape of blood or other fluid into the pericardial sac.

Pericarditis Inflammation of the pericardium associated with trauma, malignant neoplastic disease, or infection.

Pericardium A protective covering of the heart that is made of two layers separated by a fluid-filled space. The inner (visceral) layer is attached to the heart itself, whereas the outer (parietal) layer forms a sac around the heart.

Perimysium The connective tissue surrounding the fasciculi.

Peripheral vascular disease (peripheral arterial disease) Decreased localized blood flow, often to the feet, resulting in decreased arterial pressure and chronic ischemia; often the result of atherosclerosis in the peripheral artery.

Peristalsis The basic propulsive movement of the gastrointestinal (GI) tract. During normal functioning, this coordinated, rhythmic, serial contraction of smooth muscle propels the contents of the GI tract in a downward direction.

Peritoneal dialysis A procedure performed to correct an imbalance of fluid or electrolytes in the blood or to remove toxins by intermittent infusion and removal of dialysis fluid through a catheter in the peritoneal cavity.

Permissive hypothesis A hypothesis positing that poor dampening by serotonin of other neurotransmitter systems (e.g., norepinephrine and dopamine) allows wide variations in mood.

Permissiveness The process in which one hormone increases the number of cellular receptors for a second hormone, thus increasing the cellular response to the second hormone.

Peroxisome (microbody) Small, membrane-bound, intracellular organelle that uses molecular oxygen to degrade organic molecules.

Perseveration Persisting response to a previous stimulus after a new stimulus has been presented. Often applied to an excessively focused thought process.

Personality disorder A disorder that represents immature, inflexible, and persistently maladaptive ways of dealing with the intrapersonal and interpersonal aspects of life.

Pessary A device inserted in the vagina to treat uterine prolapse.

Petechiae Nonblanching, pinpoint red or purple spots caused by capillary hemorrhages.

Peyronie disease Formation of palpable, fibrous plaques on the surface of the corpora cavernosa of the penis.

pH The negative logarithm of the hydrogen ion concentration; a measure of the acidity or alkalinity of a solution.

Phagocytosis Ingestion of pathogens by leukocytes using the process of receptor-mediated endocytosis.

Phagosome A cellular lysosome containing substances obtained by phagocytosis.

Phenotype The physical, biochemical, and biological composition of an individual; expressed as recognizable traits.

Pheochromocytoma Tumor of the adrenal gland that secretes catecholamines, resulting in elevated blood pressure. An example of a condition that causes secondary high blood pressure.

Phimosis A condition in which the penile foreskin fits so tightly over the glans that it cannot be retracted.

Phlebitis Inflammation of a vein.

Photoreceptor A receptor found in the eye that responds to light.

Photosensitivity An abnormal response to exposure to light. Certain medications are photosensitive and can cause a skin reaction if the person is exposed to excessive sunlight.

Physiologic jaundice of the newborn A harmless, short-term condition caused by immature bilirubin conjugation and transport mechanisms; characterized by yellowish staining of the skin and sclera.

Physiology The study of the specific characteristics and functions of a living organism and its parts.

Pigmentary disturbance Interruption of any organic coloring material produced in the body, such as melanin.

Pinocytosis A process of ingesting fluids and small particles that is common to most cell types. Also called "cellular drinking."

Pituitary gland A gland located at the base of the hypothalamus. It consists of anterior and posterior lobes. Also called *hypophysis*.

Pityriasis rosea A rash of unknown origin that primarily affects young adults. The characteristic lesion of a macule or papule with surrounding erythema is thought to be viral in origin.

Pivot joint A synovial joint that allows rotation as its single axis movement. Examples include the superior radioulnar joint of the elbow and the union between the first and second vertebrae. Also called *trochoid joint*.

Placenta A highly vascularized organ through which the fetus receives nutrients and by which wastes are removed. It also is an endocrine organ, producing several hormones, most notably human chorionic gonadotropin.

Placenta previa Condition of pregnancy in which the placenta is implanted abnormally over the internal cervical os. It occurs in varying degrees of severity and may result in sudden massive hemorrhage after dilatation of the internal os.

Plaque A flat patch on the skin or a patch of atherosclerosis.

Plasma A complex, aqueous liquid in blood and lymph containing a number of organic and inorganic substances from which blood cells have been removed.

Plasma cell 1. An antibody-secreting B lymphocyte. 2. A dermal cell rarely seen in normal skin secretions, occurring in small numbers in most chronic inflammatory diseases of the skin and in larger numbers in granulomas.

Plasma membrane The lipid bilayer that surrounds a living cell.

Plasmalemma Plasma membrane. Cell membrane.

Plasmapheresis Removal of plasma from withdrawn blood, with retransfusion of the formed elements into the donor.

Platelet A circulating cytoplasmic fragment of megakaryocytes that is essential in the formation of blood clots and in the control of bleeding.

Pleural effusion A collection of fluid in the pleural cavity resulting from a disease process.

Pleurodesis Instillation of a chemically irritating drug (e.g., tetracycline, sterile talc, bleomycin, doxycycline) into the pleural space to stimulate inflammation and adhesion.

Pneumonia An acute inflammation of lung tissue caused by an infectious agent or by aspiration of chemically irritating fluid.

Pneumothorax Accumulation of air in the pleural space.

Poliomyelitis An infectious disease of the neuromuscular system caused by polioviruses.

Polyarteritis nodosa A form of systemic vasculitis that can cause inflamed arteries in visceral organs, brain, and skin.

Polyarticular onset (oligoarticular) Affecting five or more joints; used in association with juvenile rheumatoid arthritis.

Polycystic kidney disease A progressive genetic disease characterized by multiple dilations of the collecting ducts of the kidneys, which appear as if they are fluid-filled cysts, as a result of renal pathologic processes.

Polycythemia An excess of circulating red blood cells.

Polydipsia Excessive thirst.

Polygenic Referring to a trait determined by multiple genes at different loci, all having additive effects.

Polymenorrhea An increased frequency of menstruation, which may be associated with ovulation due to endocrine or systemic factors.

Polymorphism Inherited structural differences in proteins as a result of many alleles for a particular gene locus.

Polymorphonuclear neutrophil (PMN) Also called *segmented neutrophil.* A white blood cell that contains small lysosomal granules and a segmented nucleus with two to five lobes. PMNs compose 60% to 70% of leukocytes.

Polymyositis Inflammation of many muscles, usually accompanied by deformity, edema, insomnia, pain, sweating, and tension.

Polyp A general descriptive term used for any mass of protruding tissue. Polyps may be either benign or malignant, although the term usually refers to the benign form.

Polyploidy Having more than two sets of homologous chromosomes.

Polysaccharide A saccharide containing 10 to 10,000 monosaccharide units.

Polysomy Having more than the usual number of chromosomes.

Polyunsaturated fatty acid A fatty acid with several double bonds.

Polyuria Excretion of large amounts of urine.

Portal hypertension Abnormally high blood pressure in the blood vessels draining the intraabdominal alimentary tract, pancreas, gallbladder, and spleen. It may be due to increased resistance to blood flow, as in cirrhosis, or, rarely, to abnormally increased blood flow, as in arteriovenous communications.

Portal systemic encephalopathy A neuropsychiatric syndrome caused by liver dysfunction and resulting in mental status changes ranging from mild cerebral dysfunction to deep coma (hepatic coma) and death.

Positive end-expiratory pressure A method in which a ventilator is used to maintain positive airway pressure at the end of expiration, resulting in increased functional residual capacity and decreased shunt.

Positive feedback A term used to explain homeostatic mechanisms. Positive feedback increases deviation from the set point. Although most systems of the body operate on the principle of negative feedback, sneezing and childbirth are two examples of positive feedback.

Positive nitrogen balance The condition of dietary intake of proteins exceeding output.

Positive symptoms (schizophrenia) Symptoms of schizophrenia that are thought to be due to excessive dopamine D_2 receptor activation in the brain. Disorganized thinking (inability to connect thoughts logically), disorganized speech (rambling, tangentiality), delusions (fixed system of false beliefs), and hallucinations (sensory perception when no apparent stimulus exists) are typical positive symptoms.

Positron emission tomography (PET) A technique of brain imaging. PET studies measure changes in brain utilization of glucose.

Postictal phase The phase after a seizure during which the person is sleepy and confused.

Postobstructive diuresis Increased urinary output after resolution of partial or total obstruction of the urinary tract.

Postrenal A term referring to structures distal to the kidney, including the ureters and urethra, that may become obstructed and lead to kidney failure.

Posttraumatic stress disorder (PTSD) Psychiatric disorder characterized by an acute emotional response to a previous traumatic event.

Potter syndrome Congenital condition often associated with renal agenesis, but always manifesting with the following anomalies: wide-spaced eyes with epicanthal folds, low-set ears, broad and flat nose, hypoplastic lungs, and limb deformities.

Poverty of speech Speech that gives little information due to vagueness, empty repetitions, or obscure phrases.

Predictive value A measure used by clinicians to interpret diagnostic test results, as in positive predictive value and negative predictive value

Preeclampsia-eclampsia Elevated blood pressure during pregnancy associated with edema and proteinuria. Blood pressure returns to normal after delivery. Eclampsia is present when preeclampsia progresses to seizures. Also known as *pregnancy-induced hypertension.*

Pregnancy-induced hypertension The rapid rise of arterial blood pressure associated with a loss of large amounts of protein in the urine occurring during pregnancy. Women at risk for pregnancy-induced hypertension include teenagers and women in their late 30s and early 40s (also known as *toxemia of pregnancy* and *preeclampsia-eclampsia*).

Preload The volume of blood in the cardiac chamber just before systole (end-diastolic volume).

Prepuce Also called *foreskin;* penile skin that overlies the glans and is removed with circumcision.

Prerenal Pertaining to the area proximal to the kidney, generally referring to blood flow to the kidney, which if disrupted can result in prerenal renal failure.

Presbycusis Hearing loss associated with aging.

Presbyesophagus Presence of slow or disorganized esophageal motility in the older adult.

Presbyopia A refractive condition in which the accommodative ability of the elderly eye cannot meet the accommodative demand for near vision.

Pressure sores Localized areas of cellular necrosis resulting from prolonged pressure between any bony prominence and an external object such as a bed or a wheelchair. The tissues are deprived of blood supply and eventually die. Also called *decubitus ulcers.*

Priapism Painful, persistent erection.

Prickly heat A rash caused by midepidermal obstruction and rupture of the sweat glands from prolonged exposure to a warm and humid environment.

Primary biliary cirrhosis A slowly progressive disease that destroys small to medium-sized bile ducts and results in cirrhosis and liver failure.

Primary dysthymia A long-term state of chronic depression not associated with any other disorder. It is neither a prelude to major depression nor a state existing between episodes of a cyclic form of mood disorder.

Primary endocrine disorder Direct malfunction of a hormone-producing gland not induced by the pituitary.

Primary glomerulopathy Disease states resulting from alterations in the structure and function of the glomerular capillary circulation, in which the kidney is the only or primary organ involved.

Primary (essential, idiopathic) hypertension High blood pressure of unidentified cause. Accounts for 90% of cases of high blood pressure.

Primary lesion Injury that originates in the skin and has not been altered by scratching or by treatment.

Primary prevention The first level of health promotion, designed to prevent disease.

Primary sclerosing cholangitis A progressive chronic fibrosing inflammation of the bile ducts of unknown cause, occurring most commonly in young men and frequently associated with chronic ulcerative colitis.

Prodromal period The period preceding the onset of a disorder. Symptoms indicate an impending seizure, migraine, or other problem.

Progesterone A hormone produced by the adrenal cortex and the corpus luteum during the luteal phase of the menstrual cycle; it promotes uterine changes essential for the implantation and growth of the fertilized ovum.

Prognosis A forecast about the probable outcome of a disease; the prospect of recovery from a disease indicated by the nature, signs, and/or symptoms of the case.

Progression (cancer) A phase of carcinogenesis when clones of cells that have undergone

mutations begin to develop new properties that allow them to become increasingly malignant.

Progressive familial intrahepatic cholestasis A rare autosomal-recessive disorder comprising severe jaundice, pruritus, and malabsorption attributable to a defect in bile salt excretion.

Proinsulin Precursor to insulin produced by the β cells of the pancreas.

Prokaryote A cell that does not have a membrane-bound nucleus or other membrane-bound organelles (e.g., bacteria).

Prolactin Hormone secreted by the anterior pituitary. After birth of an infant, prolactin stimulates milk production.

Proliferation The reproduction or multiplication of similar forms. The term is usually applied to an increased number of cells as a result of mitosis.

Pronephros One of three distinct phases in the development of the renal system. The pronephros is the earliest state in humans, corresponding to the mature structure in primitive vertebrates.

Prostate Gland located below the bladder; its secretions help activate sperm and maintain their motility.

Prostatic urethra The widest and most distensible part of the male urethra.

Prostatitis Inflammation of the prostate.

Prostatodynia Pain in the prostate.

Protein A molecule composed of nitrogen, carbon, hydrogen, oxygen, and occasionally sulfur; when hydrolyzed, proteins yield amino acids.

Proteoglycan Any of a group of polysaccharide–protein conjugates occurring primarily in the matrix of connective tissue and cartilage; composed mainly of polysaccharide chains, particularly glycosaminoglycans, as well as minor protein components.

Proto-oncogene A normal cellular gene that is growth promoting and usually inhibited in nonproliferating cells. When erroneously activated, it becomes an oncogene and promotes tumor formation.

Provirus The viral DNA that is spliced into the host cell's DNA.

Pruritus Itching of the skin.

Pseudocyst A collection of fluid within or adjacent to the pancreas that often has a direct communication to the pancreatic duct. It is the most common localized complication of acute pancreatitis.

Pseudoglandular period The first stage in fetal lung development when the bronchial divisions are differentiated and the major elements of lung tissue are present except for those involved in gas exchange: the respiratory bronchioles and alveoli.

Pseudomembranous enterocolitis An acute inflammation and necrosis of the small and large intestines caused by *Clostridium difficile,* usually affecting the mucosa but sometimes extending to other layers.

Psoriasis A common chronic skin disease characterized by papules and plaques with an overlying silvery scale. Lesions can appear on any area of the body but especially the knees, elbows, lower part of the back, scalp, and nails.

Psoriatic arthritis An inflammatory arthritis associated with psoriasis occurring in approximately 0.1% of the population in the United States. Peak age of onset is 30 to 55 years of age, and the arthritis can occur in patients who have had psoriasis for many years.

Psychogenic Produced or caused by emotional or psychological factors rather than organic factors.

Psychosis The most serious and debilitating of mental disorders. The hallmarks of psychosis are delusions and hallucinations, thought disorders, and inappropriate emotional responses or social behavior.

Psychosomatic medicine The discipline involving the physiologic impact of psychic stress on the emergence of disease.

Pulmonary embolism (PE) The blockage of a pulmonary artery by fat, air, tumor tissue, or thrombus.

Pulmonary function testing (PFT) A procedure for determining the volumes and capacities of the lungs.

Pulmonary hypertension Abnormally high blood pressure within the pulmonary circulation.

Pulmonary tuberculosis Infection of the lungs by *Mycobacterium tuberculosis.*

Pulmonic valve The cardiac valve that lies between the right ventricle and the pulmonary artery. It is open during ventricular systole and closed during ventricular diastole. Pulmonic valve closure contributes to heart sound S_2.

Pulse pressure The difference between the systolic and diastolic blood pressures.

Purpura Hemorrhagic lesions 2 to 4 mm in diameter; petechiae that occur in groups or patches caused by a vascular or bleeding disorder.

Putamen nuclei The larger, darker, and more lateral part of the lentiform nucleus in the brain.

Pyelonephritis An infection of the kidney medulla or cortex.

Pylorus Muscular sphincter between the stomach and the duodenum that controls gastric emptying and limits the reflux of bile from the small intestine.

Pyogenic Creating the formation of pus; typically at the site of an inflammation caused by bacterial infection.

Pyrogenic Producing or produced by fever.

Pyrosis A substernal burning sensation that may radiate to the neck or throat. It is caused by the reflux of gastric contents into the esophagus (also called *heartburn*).

Pyuria The presence of an excessive number of white blood cells in the urine. It is generally a sign of urinary tract infection.

R

Radiation Emissions of radioactive energy, rays, or waves. Can cause radiation sickness, but often used in the treatment of cancer.

Radical mastectomy Surgical removal of the entire breast, lymphatic drainage structures, and underlying pectoral muscles.

Radiculopathy A disease involving compression and dysfunction of a spinal nerve root.

Radiolysis Lysis or splitting of water molecules into H^+ and OH^- ions by the action of radioactive particles.

Rapidly progressing glomerulonephritis A syndrome that combines abrupt hematuria and proteinuria followed by a swift decline in renal function.

Rarefaction Decrease in density and weight of bone, but not in volume.

Raynaud phenomenon Blanching or cyanosis of fingers or hands on exposure to cold or emotional stress. The phenomenon is attributed to vasospasm and structural disease of blood vessels. Puffiness and swelling of the hands and fingers are noted clinically.

Reality testing The act of evaluating and considering the differences between internal experiences and external events.

Receptor activation Binding of a ligand to a cellular receptor, resulting in a change in intracellular cell signaling or function.

Receptor specificity The principle of allowing intracellular processes to be activated only by certain hormones. If a cell does not have the specific receptors for a hormone, it will not respond to the hormone.

Recessive Referring to a gene allele that fails to be expressed in the phenotype when a dominant allele is present. The trait carried in a recessive allele is apparent only when two identical copies are present.

Recruitment The process of calling in additional motor units in response to an increase in stimulation of motor nerves, or calling in additional gas exchange units in the lung when pulmonary arterial pressure increases.

Rectocele Protrusion of the anterior rectal wall into the posterior vagina at a weakened part of the vaginal musculature. It usually results from an injury during either childbirth or surgery, and it may also be the result of the aging process or an inherent weakness in the vaginal wall.

Recurrent urinary tract infection Repeated infections within a short period after verified resolution of an earlier infection.

Red blood cell (erythrocyte) Hematologic cell that has no organelles and is responsible for transporting oxygen to the tissues, removing carbon dioxide from the tissues, and buffering blood pH.

Reed–Sternberg cell A malignant cell type thought to be derived from B cells and found in affected lymph nodes of patients with Hodgkin disease. The presence of Reed–Sternberg cells differentiates Hodgkin disease from all other forms of malignant lymphoma.

Reentry The proposed mechanism for many dysrhythmias, including premature complexes and fibrillation. Reentry occurs when an impulse is able to activate the cardiac muscle more than once because of abnormalities in conduction through a portion of the heart.

Referred pain Pain felt at a site different from that of an injured or diseased organ or body part.

Reflex incontinence Urine loss that occurs without sensory warning or awareness.

Reflux (urinary) Retrograde flow of urine from the bladder to the kidney.

Refractory period The time during which a nerve or muscle membrane is unable to respond to a stimulus by generating an action potential.

Regurgitation (valvular) Retrograde blood flow through a cardiac valve when the valve is supposed to be closed.

Reiter syndrome Seronegative arthritis that appears 2 to 6 weeks after onset of an infection; characterized clinically by diffuse swelling of fingers and toes, swelling in the Achilles tendon or plantar fascia, and low back pain.

Relative anemia Anemia characterized by normal total red blood cell mass with disturbances causing excessive plasma fluid volume resulting in low hematocrit value.

Relative polycythemia Polycythemia characterized by normal total red blood cell mass with reduced plasma volume resulting in elevated hematocrit value.

Relaxation time The period between peak tension and zero tension.

Releasing hormone A hormone secreted by the hypothalamus that stimulates the anterior pituitary gland to secrete other hormones.

Reliability The extent to which a test measurement produces the same results with different investigators or with repeated measures over time.

Remission Disappearance of clinical manifestations of disease. In leukemia, complete remission is determined by the absence of leukemic blasts in the bone marrow aspirate and peripheral blood. Undetected malignant stem cells still may be present, and remission does not imply cure.

Renal adenoma A tumor smaller than 3 cm with a cellular makeup similar to that of renal cell carcinoma.

Renal agenesis Failure of one or both kidneys to develop.

Renal angiomyolipoma (hamartoma) Benign renal tumor composed of abdominal blood vessels, clusters of fat cells, and sheets of smooth muscle.

Renal calculus Concretion of crystals of material (e.g., uric acid, calcium phosphate, struvite) that initially form in the calices or pelvis of the kidney. Calculi may migrate down the urinary tract and cause pain, obstruction, and infection.

Renal cell carcinoma Most common malignant tumor of the kidney.

Renal hypoplasia An abnormally small kidney that is morphologically normal but has either a reduced number of nephrons or smaller nephrons.

Renal or ureteral colic Intermittent flank or abdominal pain caused by spasms in the kidneys and/or ureters.

Renin An enzyme stored and released by the juxtaglomerular cells; converts angiotensinogen to angiotensin I.

Renin–angiotensin–aldosterone system (RAAS) The regulation of sodium balance, fluid volume, and blood pressure by the cascade of reactions beginning with release of renin, an enzyme that cleaves angiotensinogen to angiotensin I, followed by conversion of angiotensin I to angiotensin II by angiotensin-converting enzyme (ACE). Angiotensin II stimulates the release of aldosterone from the adrenal cortex.

Reperfusion injury Damage to tissues when blood flow is restored after a period of ischemia.

Resident flora Microorganisms that usually reside in a certain environment on or in the host without causing disease. Also called *commensal organisms*.

Residual urine Urine that remains in the bladder after urination.

Resistance In the stress response, resistance occurs when sympathetic activity declines while secretion of adrenocortical hormones is high. Resistance in a vessel or bronchiole represents the impedance to flow and is correlated with lumen radius.

Respiratory acidosis A pulmonary hypoventilation condition that tends to cause an excess of carbonic acid and results in acid–base imbalance with a high $PaCO_2$.

Respiratory alkalosis A pulmonary hyperventilation condition that tends to cause a deficit of carbonic acid and results in acid–base imbalance with a low $PaCO_2$.

Respiratory quotient The ratio of the volume of carbon dioxide produced to the volume of oxygen used in tissue metabolism. Varies depending on whether the source of fuel is carbohydrates or fats.

Respiratory syncytial virus A member of a subgroup of myxoviruses that in tissue culture cause formation of giant cells or syncytia.

Resting membrane potential The transmembrane voltage that exists in nonexcitable cells and in neurons and muscle cells when not producing an action potential.

Reticular formation Fibers in the brainstem that arouse the cerebrum.

Reticular layer One of two layers of the dermis; consists of a denser reticulum (network) of fibers than the papillary layer above it. The reticular layer is made of collagen and elastin and contains skeletal (voluntary) and smooth (involuntary) muscle fibers.

Reticulocytosis Increase in the number of circulating reticulocytes (immature red blood cells).

Reticulohistiocytic group Cells of the dermis consisting of fibroblasts, histiocytes, and mast cells. Immature cells of the reticulohistiocytic group are called reticulum cells.

Retinal detachment A separation of the retina in the back of the eye.

Retinopathy A group of noninflammatory eye disorders. Major contributing conditions include diabetes, hypertension, and atherosclerotic vascular disease.

Retrodisplacement Posterior displacement, such as an alteration in the position of the uterus in which the body of the uterus is displaced from its normal location overlying the bladder to a position in the posterior pelvis.

Retroperitoneal Referring to the anatomic space in the abdomen behind the peritoneal cavity, where the kidneys reside.

Retrovirus An RNA virus capable of transcribing its own RNA into DNA, which can be integrated into the host genome through the actions of a viral polymerase, also called *reverse transcriptase.*

Reverse transcriptase An enzyme that allows certain viruses to convert RNA to DNA and incorporate their genome into the DNA of the host cell.

Reye syndrome Primarily a children's disease, it is characterized by fatty infiltration of the liver with severe hepatic dysfunction.

Rheumatoid arthritis A systemic, inflammatory, autoimmune connective tissue disease. Enzymes are released into the joint fluid, causing inflammation, proliferation of synovium, and tissue damage. The hands, wrists, knees, and feet are most commonly involved. Joint involvement is symmetric.

Rhinorrhea Drainage of a watery fluid from the nasal mucosa.

Rhonchus Coarse, bubbling sound usually heard on expiration, but which also can be heard on inspiration; it is caused by secretions in the airways.

Rhus dermatitis An inflammatory reaction to poison ivy, poison oak, and poison sumac. Clinically, rhus dermatitis begins within 48 hours of contact. The first symptom is pruritus, followed by erythema and vesicle formation, sometimes in linear fashion.

Rhythmicity The ability to beat regularly without external stimuli. Also called *automaticity*.

Ribonucleic acid (RNA) A nucleic acid found in both the nucleus and the cytoplasm of cells that has several roles in the translation of the genetic code and the assembly of proteins.

Ribosome A cellular structure containing ribosomal RNA and proteins that bind to mRNA and perform the task of translating the RNA message into a protein.

Rickets A condition caused by the deficiency of vitamin D; seen primarily in infancy and childhood and characterized by abnormal bone formation.

Rigor mortis The stiffening of muscles throughout the body after death resulting from the formation of persistent actin–myosin cross-bridges.

Risk factor Characteristic related to the probability of a certain outcome; a risk factor may be shown to cause an outcome or may be correlated with an outcome.

RNA polymerase An enzyme complex that binds to a gene segment of DNA, using it as a template for the synthesis of an RNA strand.

Rocky Mountain spotted fever An infection caused by a tick that carries *Rickettsia rickettsii*. The characteristic rash is a macular or maculopapular one that spreads to the rest of the body. Other symptoms include generalized edema, conjunctivitis, petechial lesions, photophobia, lethargy, confusion, and cranial nerve deficits.

Roseola infantum A contagious viral disease that generally affects children younger than 4 years and usually those about 1 year of age. It produces a characteristic maculopapular rash covering the trunk and spreading to the appendages.

Rostral A positional term referring to the head end.

Rough endoplasmic reticulum A portion of the endoplasmic reticulum that is studded by ribosomes. Ribosomes attached to the endoplasmic reticulum synthesize proteins that are destined for the plasma membrane or lysosomes.

Rubella Also known as *3-day measles* or *German measles*. Rubella is a childhood disease caused by the rubella virus. It is characterized by a diffuse punctate, macular rash that begins on the trunk and spreads to the arms and legs.

Rugae Folds in the lining of the body of the bladder (and stomach) that allow it to distend to accommodate volume at a low pressure.

S

Saddle joint A joint in which the surfaces are convex in one plane and concave in the other,

permitting both flexion-extension and adduction-abduction movements; the surfaces of a saddle joint fit together as a saddle fits a horse. The carpometacarpal joint of the thumb is a saddle joint. Also called a *sellar joint*.

Saline deficit Extracellular fluid volume deficit.

Saline excess Extracellular fluid volume excess.

Salivary amylase Digestive enzyme (also called *ptyalin*) contained in saliva; initiates carbohydrate digestion in the mouth.

Sarcoidosis A chronic disorder of unknown origin characterized by the formation of tubercles of nonnecrotizing epithelioid tissue.

Sarcolemma The plasma membrane that encloses a muscle cell.

Sarcomere The unit of muscle contraction in striated muscle. A sarcomere extends from one Z disk to another and consists of overlapping actin and myosin filaments.

Sarcoplasm The cytoplasm of the muscle fiber.

Sarcoplasmic reticulum A calcium-storing structure in muscle cells analogous to the endoplasmic reticulum; it fills the space between myofibrils and forms sacs.

Saturated fatty acid Having the maximal number of hydrogen atoms present so that only single bonds exist in the carbon chain, as in saturated fatty acids.

Saturation The condition of being saturated. The degree of hydrogen saturation refers to the number of double bonds between the carbon atoms in the hydrocarbon chain. Hemoglobin saturation occurs when all four ions in the hemoglobin molecule are bound to oxygen.

Scabies Infestation with the mite *Sarcoptes scabiei* in humans. Scabies begins with eggs laid in the stratum corneum. These eggs hatch into larvae within 3 to 4 days and reach adulthood within 2 months.

Scarlet fever A systemic reaction to the toxins produced by group A β-hemolytic streptococci. It occurs when the person is sensitized to the toxin-producing variation of streptococci. Scarlet fever frequently occurs in association with streptococcal sore throat (strep throat), but it may also be associated with a wound, skin infection, or puerperal infection.

Schizoaffective disorder A psychiatric disorder in which either a major depressive or manic episode develops concurrently with symptoms of schizophrenia.

Schizoid Indifferent to social interaction and possessing a limited range of emotional experience and expression.

Schizophrenia A syndrome or combination of mental disorders characterized by paranoia, delusions, and hallucinations associated with impaired interpretation of reality.

Scleroderma A disorder characterized by massive collagen deposition with fibrosis accompanied by inflammatory reactions and vascular changes in the capillary network.

Scoliosis A lateral deviation of the spine resulting in an S- or C-shaped spinal column. The disorder, most common in adolescent girls, can be a consequence of congenital, connective tissue, or neuromuscular disorders.

Scrotum Pouchlike sac containing the testes, epididymis, and spermatic cord.

Seasonal affective disorder A condition in which lethargy results from seasonal changes of decreased periods of daylight and longer nights.

Sebaceous gland Oil- or sebum-producing gland that anoints hair and skin.

Seborrheic dermatitis A common chronic inflammatory skin disease characterized by greasy scales and yellowish crusts.

Seborrheic keratosis Benign skin tumor common in the elderly; composed of immature epithelial cells.

Second messenger An intracellular signal that is produced in response to an extracellular signal binding to receptors on the cell membrane.

Secondary dysthymia A long-term state of chronic depression that is associated with some other nonmood disorder, which may be a classic mental disorder (e.g., anorexia nervosa) or a physical illness.

Secondary endocrine disorder A malfunction of the hypothalamus/pituitary cells that control the hormone-producing gland.

Secondary glomerulopathy Alterations in the structure and function of the glomerular capillary circulation resulting from drug exposure, infections, or glomerular injury in the setting of multisystem or vascular abnormalities.

Secondary hypertension High blood pressure in which the cause can be identified.

Secondary lesion Injury modified by normal progress over time or by such external actions as scratching.

Secondary prevention The second level of health promotion, based on early detection and screening.

Second-degree burn A burn that affects the epidermis and the dermis; classified as superficial or deep according to the depth of injury.

Secretin A digestive hormone that is produced by the S cells lining the duodenum and jejunum when protein of partially digested food enters the intestine from the stomach; it stimulates the pancreas.

Segmentation contractions A set of movements that occur in the small intestine. The primary effect of these contractions is progressive mixing of solid chyme particles with secretions of the small intestine.

Seizure A transient neurologic event of paroxysmal abnormal or excessive cortical electrical discharges that is manifested by disturbances of skeletal motor function, sensation, autonomic visceral function, behavior, or consciousness.

Self-identity disturbance In schizophrenic patients, self-identity disturbances can be profound, both in terms of the ability to differentiate one's physical self from the physical environment and in terms of the psychological discernment of self as distinct from others.

Semen Male reproductive fluid that contains spermatozoa; released with ejaculation.

Seminal vesicles Two glands that contribute rich nutrients to the seminal fluid.

Senile purpura A skin condition affecting older people and characterized by fragile blood vessels that rupture with minimal trauma.

Sensitivity Susceptibility to a substance, such as a medication or antigen.

Septic shock A form of distributive shock that occurs in septicemia when endotoxins or exotoxins are released from certain bacteria in the bloodstream causing inappropriate vasodilation.

Sequela, sequelae A condition or conditions caused by and following a disease.

Sequestra Fragment of dead bone that is partially or entirely detached from the surrounding or adjacent healthy bone.

Serotonin A major monoamine neurotransmitter that is an etiologic factor in mood disorders. Also called *5-hydroxytryptamine.*

Sertoli cell An elongated cell that supports and provides nutrition to attached spermatids until they mature into spermatozoa.

Severe acute respiratory syndrome (SARS) An infectious respiratory illness first reported in Asia.

Sex chromosome A chromosome that confers gender to the individual. In humans, females are designated as 46XX, whereas males are 46XY. The Y chromosome confers male gender.

Sexually transmitted infection One of many infections that can be transmitted by sexual contact, regardless of whether the disease has manifestations in the genital organs (previously referred to as *venereal disease* or *sexually transmitted disease*).

Sharpey fibers Fibers that attach tendons to bones; they are continuous with the perimysium.

Shock A condition of severe hemodynamic and metabolic disturbance resulting in an imbalance between oxygen supply and oxygen demand at the cellular level. The common types of shock are cardiogenic, hypovolemic, obstructive, and distributive (septic, anaphylactic).

Short-bowel syndrome Severe diarrhea and significant malabsorption that develop after the surgical removal of large portions of the small intestine. The severity of the manifestations depends on the amount and location of the bowel resected.

Shunt (right-to-left) An abnormal route of blood flow through the heart or lungs that allows movement of blood into the arterial system without passing through oxygenated areas of the lung.

Sickle cell A red blood cell containing abnormal hemoglobin that causes the cell to assume a sickle shape under decreased oxygen tension.

Sign Objectively identifiable manifestation of the disease. Fever, reddening of the skin, and a palpable mass are signs of disease; as opposed to symptoms that are not observable.

Skeletal muscle Striated muscle that is attached to bone. Constituting 40% of total body weight, skeletal muscle enables bones to move at the joint and provides strength and protection to the skeleton by distributing and absorbing shock.

Skeletal system Rigid system of bony structures designed to protect internal organs and provide bony attachments for muscles and ligaments; presents rigid levers to allow for functional movement of the body and its separate parts.

Skin A relatively flat membrane composed of an outer, thinner layer (epidermis) and an inner, thicker layer (dermis).

Skin cancer A cutaneous neoplasm caused by ionizing radiation, certain genetic defects, or chemical carcinogens, including arsenics, petroleum, tar products, and fumes from some molten metals, or by overexposure to the sun or other sources of ultraviolet light.

Sleep study Recording of electroencephalograph (EEG) motor activity and respirations during sleep to determine duration and type of sleep and number of awakenings.

Sliding filament theory This theory of muscle contraction is suggested by the anatomic configuration of the sarcomere. Muscle shortening is accomplished by increasing the amount of overlap between actin and myosin filaments. Also called *cross-bridge theory.*

Slow twitch (type I, red) A muscle fiber that develops tension more slowly than a fast-twitch fiber. This fiber is usually fatigue resistant and relies on oxidative phosphorylation for energy.

Slow-wave electrical activity One of the basic types of electrical activity in the gut. Slow waves represent an ongoing basic oscillation in membrane potential occurring in the smooth muscle of the gastrointestinal tract between 3 and 12 times per minute.

Smooth endoplasmic reticulum A portion of the endoplasmic reticulum that has no ribosomes. Smooth endoplasmic reticulum is a site of lipid synthesis.

Soft tissue injury Any trauma to soft tissue with disruption of circulatory and lymphatic systems.

Solar elastosis Wrinkled, weather-beaten appearance of skin caused by overexposure to sunlight.

Somatosensory receptors Specialized nerve endings located in the dermis of all skin areas. Receptors permit the skin to serve as a sensory organ, transmitting sensations of pain, pressure, touch, and temperature.

Somatotopic Referring to the sequential arrangement of neurons related to sensory or motor function in specific anatomic regions.

Specific immune response Creation of clones of reactive lymphocytes that recognize a specific antigen. Frequently involves production of specific antibodies that leads to the destruction of the antigen.

Specificity The quality of being distinctive and the probability that the test will be negative among patients who do not have the disease.

Spermatid An immature sperm cell.

Spermatocele A painless, cystic mass containing sperm.

Spermatogenesis Production of sperm cells.

Spermatozoon A mature sperm cell.

Spherocyte An abnormal spherical (round) erythrocyte that is less biconcave than a normal erythrocyte.

Spherocytosis The presence of spherocytes in the blood; usually from a genetic disorder.

Spike potential Sudden increase in membrane potential in the smooth muscle of the gastrointestinal tract that appears on the peaks of slow waves in response to certain conditions, including stimulation by stretching or the effects of acetylcholine or parasympathetic excitation.

Spina bifida A developmental anomaly characterized by defective closure of the bony encasement of the spinal cord (neural tube) through which the spinal cord and meninges may or may not protrude.

Spinal shock A temporary physiologic suspension of spinal cord function and reflexes below the level of cord injury.

Spiral (bacteria) Referring to any bacterium of the genus *Spirochaeta* that is motile and spiral-shaped with flexible filaments. See *spirochete*.

Spiral fracture A fracture resulting from rotational forces and causing bone to separate in the form of an S around the bone.

Spirochete Spiral-shaped bacterium.

Splinter hemorrhage A linear hemorrhage, appearing as a red or brown streak, running parallel to the finger in the nail bed; may be linked to bacterial endocarditis and trichinosis.

Spontaneous abortion Expulsion of the products of conception from the uterus before the period of fetal viability. It is usually called "miscarriage" by laypersons, and is differentiated from an elective abortion.

Squamous cell carcinoma A type of cancer that often occurs in areas of skin excessively exposed to sunlight and arising from keratinocytes.

Staging The process of determining the extent and location of cancer in an individual.

Standard deviations A mathematical statement of the dispersion of a set of values or scores from the mean.

Stasis A "staying" of a substance in an anatomic location. A reduction in the normal rate of flow. A *stasis* of bile in the gallbladder promotes an increase in gallstone formation.

Status asthmaticus Severe, prolonged asthma attack that does not respond to routine therapy.

Status epilepticus Rapid succession of seizures without intervals of consciousness. Brain damage may result.

Steatorrhea Passage of high fat content in the feces; seen in malabsorption diseases where there is lack of pancreatic enzymes, such as cystic fibrosis.

STEMI A myocardial infarction with the finding of ST elevation on the 12-lead electrocardiogram.

Stenosis (valvular) Obstruction to blood flow through cardiac valves that open incompletely.

Steroid A hormone produced from cholesterol and secreted by the adrenal cortex and other cells. Includes glucocorticoids, mineralocorticoids, and androgens.

Still disease Systemic-onset juvenile rheumatoid arthritis.

Stomatitis An inflammation of the tongue that may extend to the buccal mucosa, lips, and palate.

Strabismus An abnormal ocular condition in which the visual axes of the eyes are not directed at the same point.

Stratum Layer.

Stratum corneum Outermost layer of the epidermis; composed of flat, compact cells that have lost their nuclei.

Stratum germinativum Also known as the *basal cell layer,* the final (fifth) layer of the epidermis is a line of cuboidal cells that marks the lowest boundary of the epidermis and divides it from the dermis.

Stratum granulosum Third layer of the epidermis, comprising flatter cells that contain protein granules, called *keratohyalin granules.*

Stratum lucidum Second layer of the epidermis, appearing as a translucent line of flat cells. This layer of the skin is present only on the palms and the soles.

Stratum spinosum Fourth layer of the epidermis, composed of upwardly migrating and maturing keratinocytes. This layer forms the bulk of the epidermis over most of the body.

Strawberry hemangioma A soft vascular nevus, usually present on the face or neck, occurring at birth or shortly afterward.

Strawberry tongue Bright red papillated tongue, characteristic of scarlet fever.

Stress The sum of biological reactions produced when an organism's homeostasis is disrupted.

Stress fracture A fracture of one cortical surface of the bone; often caused by repetitive activity such as running.

Stress incontinence Loss of bladder control caused by increased intraabdominal pressure (e.g. from coughing) combined with pelvic muscle laxity.

Stressor An agent or condition capable of producing stress. The term denotes both physical (gravity, mechanical force, pathogen, injury) and psychological (fear, anxiety, crisis, joy) forces that an individual may experience.

Striation (muscle) The typical pattern of banding apparent upon microscopic inspection of a skeletal or cardiac muscle cell.

Stricture A narrowing or constriction of the lumen of a tube, duct, or hollow organ, such as the intestine, ureter, or urethra.

Stroke volume The volume of blood ejected from the ventricle in one contraction (end-diastolic volume minus end-systolic volume).

Structural scoliosis The most severe form of scoliosis; can be progressive in which the mechanics of the curve are such that rotation of the vertebrae occurs in combination with lateral curvature. This usually produces a protuberance of one side of the rib cage, seen best when a person bends forward.

Subarachnoid space The space between the arachnoid and the pia mater.

Subendocardium The part of the myocardium lying in proximity to the endocardial surface.

Subluxation Displacement of a bone from its normal position (articulating surface) in a joint; less severe than dislocation.

Substantia gelatinosa Another term for areas in laminae II and III that are important in the transmission of pain signals in the spinal cord.

Substantia nigra The layer of gray matter separating the tegmentum of the midbrain from the crus cerebri; part of the basal ganglia from which dopamine-secreting neurons project to other basal ganglia.

Sudden cardiac death (sudden cardiac arrest) Death due to cardiac causes (or successful resuscitation) within 1 hour of symptom onset.

Sulcus A deep furrow or groove on the surface of a structure (e.g., cerebral cortex).

Summation The additive response to repetitive or multiple stimuli.

Superficial fascia Loose subcutaneous layer rich in fat and areolar tissue, which lies beneath the dermis. Also known as *hypodermis*.

Superficial partial-thickness burn Marked by destruction of the epidermis and dermis, a superficial partial-thickness burn is also a second-degree burn. The water vapor barrier is absent, but tactile and pain sensors are intact.

Suppurative disease A disease that produces purulent material (pus).

Surface film Thin film of emulsified material spread over the surface of skin.

Surfactant A surface tension–reducing agent produced by type II pneumocytes in the lung important for preventing alveolar atelectasis.

Suture joints A joint that unites bone with a thin but dense layer of fibrous tissue. Found only in the skull.

Sweat gland Most numerous of the skin glands. Sweat glands are of two types: apocrine and eccrine.

Symphysis joint A joint that connects bony segments by a fibrocartilaginous plate or disk.

Symptom Subjective feeling that an affected individual can report to an observer. Nausea, malaise, and pain are *symptoms* of disease; as opposed to signs that are objectively observed.

Symptomatically Defined by symptoms.

Synarthrosis A fibrous or cartilaginous (nonsynovial) joint.

Synchondrosis A joint connecting cartilage to a bony component; allows bone growth while providing stability.

Syncytium A complex of fused cells that act in concert.

Syndesmosis A fibrous joint in which opposing surfaces that are relatively far apart are connected by ligaments.

Syndrome A collection of signs and symptoms that occur together.

Syndrome of inappropriate antidiuretic hormone (SIADH) secretion Excessive antidiuretic hormone secretion either from the posterior pituitary gland or from other tissues. SIADH results in retention of water, hemodilution, and hyponatremia.

Synostosis The bony union that results from the fusion of a suture joint.

Synovial fluid A clear, pale yellow, viscous fluid similar to blood plasma but containing hyaluronic acid and a glycoprotein called *lubricin*. Synovial fluid reduces friction between the capsule and joint surfaces, lubricates the surface of the cartilage, resists shear forces, and provides nourishment for cartilage.

Synovial joint Freely movable joint in which contiguous bony surfaces are covered by articular cartilage and connected by a fibrous connective tissue capsule lined with a synovial membrane.

Synovial sheath Also called *joint capsule*. A dense layer of connective tissue surrounding synovial joints. The capsule is solidly attached to the periosteum of the adjacent bony components. The synovial sheath provides strength to the joint and, through its neural receptors, detects motion, compression, tension, vibration, and pain.

Synovitis An inflammatory condition of the synovial membrane of a joint as the result of an aseptic wound or traumatic injury.

Syphilis Sexually transmitted disease caused by the spirochete *Treponema pallidum* and characterized by distinct stages of effects over a period of years.

Systemic lupus erythematosus A chronic inflammatory autoimmune disease resulting from a type III hypersensitivity reaction. It is a multisystem relapsing disease that can affect skin, mucosa, lung, heart, kidneys, central and peripheral nervous systems, and blood components. Arthralgias and synovitis are common features. Skin lesions are often present as a butterfly rash. Renal failure is the leading cause of death with this disease.

Systemic vascular resistance The impedance to blood flow exerted by the arterioles; determined primarily by vessel diameter.

Systole A phase of the cardiac cycle in which the ventricles are contracting to develop force and eject blood.

Systolic blood pressure The maximal pressure in the aorta and major arteries during ventricular ejection of blood.

T

T cell Lymphocyte that provides cellular immunity. T cells have T-cell receptors (TCRs) and mature in the thymus.

Tangential excision Also called *full-thickness excision*. Done to remove eschar in thin layers until viable tissue is visible.

Target cell or target organ The cell or organ that is stimulated by the effects of a hormone.

Tau protein Neuronal protein that organizes microtubules. Also known as *neural thread protein*, tau protein becomes chemically altered in Alzheimer disease and twists into abnormal bundles, known as *neurofibrillary tangles*. See *neural thread protein*.

Telangiectasia A lesion created by dilated blood vessels.

Telogen The resting phase of hair growth.

Telomerase An enzyme that permits addition of nucleotides to the tips of the chromosomes to prevent progressive shortening of the telomeres during cell division. Telomerase is produced by cancer cells, enabling them to become immortal.

Telomere The end cap of the chromosome; this section shortens with each cell division.

Tendinitis Inflammation of the tendon within the sheath.

Tension pneumothorax Presence of air in the pleural space that develops a positive pressure and compresses mediastinal structures.

Teratogen An agent or factor that causes damage or physical defects in a developing embryo.

Terminal hair Long, coarse, thick, visible strands of tightly fused keratinized epidermal cells.

Terminal sac period The third stage in fetal lung development when terminal sacs become thinner, preparing the lung tissue for gas exchange. Proliferation of pulmonary capillaries is also prominent during this period.

Tertiary prevention The third phase of health promotion, based on supporting independent function and preventing further disease-related deterioration.

Testis The male gonad or reproductive gland that produces spermatozoa and houses the Leydig cells.

Testosterone Male sex hormone produced by interstitial cells in the testes.

Thalamus Portion of diencephalon; mass of gray matter involved in relay of sensory information, emotion, arousal, and complex reflexes.

Thalassemia An inherited form of anemia characterized by microcytic red blood cells that lack either α- or β-hemoglobin chain genes.

Thoracentesis Surgical perforation of the chest wall and pleural space with a needle to aspirate fluid for diagnostic or therapeutic purposes or to remove a specimen for biopsy.

Thoracotomy Surgical opening of the chest wall.

Threshold The lowest level at which a stimulus can produce a response.

Thrill Vibration palpated over a blood vessel or heart chamber reflecting turbulent blood flow.

Thromboangiitis obliterans (Buerger disease) An occlusive vascular condition usually associated with smoking and affecting a leg or a foot in which the small- and medium-sized arteries become inflamed and thrombotic.

Thrombocyte Circulating cytoplasmic fragment of a megakaryocyte that is essential in the formation of blood clots and in the control of bleeding. Also called *platelet*.

Thrombocytopenia A deficiency of platelets (thrombocytes) in the peripheral blood. Any reduction in platelet count below normal is called *thrombocytopenia,* but significant risk of bleeding does not occur until the count drops below about 20,000/μl.

Thrombocytosis An abnormal increase in the number of platelets in the blood.

Thromboembolus An embolus that originated as a thrombus then broke free to travel with the blood flow; most commonly in the venous system of the lower extremities.

Thrombophlebitis Inflammation of a vein accompanied by the formation of a blood clot.

Thrombus Stationary blood clot formed within a vessel.

Thrush Candidiasis of the tissues of the mouth. The condition is characterized by the appearance of creamy white patches of exudates on an inflamed tongue or buccal mucosa.

Thyroid-stimulating hormone (TSH) A peptide secreted by the anterior lobe of the pituitary gland that controls the release of thyroid hormone and is necessary for the growth and function of the thyroid gland.

Thyroid storm Extreme thyrotoxicosis. Massively elevated levels of thyroid hormones cause an increased basal metabolic rate, tachycardia, hypertension, and fever, eventually leading to cardiovascular collapse.

Thyroxine (T_4) A hormone containing four iodine molecules secreted by the thyroid gland.

Tight junction Cell-to-cell junction that seals adjacent epithelial cells together and prevents the passage of most substances through the epithelial sheet.

Tinea The infection caused by fungal infections of the skin in any cutaneous area, including the hair and nails.

Tinnitus Ringing, buzzing, or roaring in the ears; commonly associated with exposure to loud noise or disorders such as Meniere disease.

Tonic Referring to continuous stimulation or muscular contraction.

TORCH complex An acronym used to describe the usual offending infectious agents that commonly cause congenital anomalies: ***t***oxoplasmosis, ***o***thers, ***r***ubella, ***c***ytomegalovirus, ***h***erpesvirus.

Torsion Act of twisting or condition of being twisted. Applies to the state of the testes when abnormally rotated.

Trabecular bone Cancellous bone cells arranged in response to mechanical stress placed on the bone. Configuration of bone cells increases strength of bone.

Tracheoesophageal fistula Congenital anomaly in which an abnormal opening between the trachea and esophagus exists. It requires immediate diagnosis and surgical correction.

Transcellular fluid Body fluid contained in special compartments, such as the synovial or cerebrospinal compartments; a component of extracellular fluid.

Transcription The process by which a segment of DNA is used as a template to produce a complementary sequence of messenger RNA.

Transient flora Microorganisms that temporarily reside in a certain environment on the host.

Transient ischemic attack A temporary episode of cerebrovascular insufficiency that is usually associated with partial occlusion of a cerebral artery.

Transitional flow Airflow occurring in the larger airways, especially at bifurcations. Also known as *mixed pattern of airflow.*

Translation Formation of a polypeptide chain in a sequence dictated by messenger RNA.

Translocation Shifting of a segment of one chromosome into another chromosome.

Transmission of infection The process by which a pathogenic organism is transferred from one host to another.

Transmural Denoting the entire thickness of a wall (e.g., the myocardial wall). A transmural myocardial infarction extends throughout the entire cardiac muscle layer.

Transportation The movement or transference of biochemical substances from one site to another.

Transudate Fluid of low protein content that passes through membranes because of a difference in hydrostatic pressure.

Transverse fracture Fracture that occurs in a straight line at approximately a 90-degree angle to the longitudinal axis of the bone.

Traumatic alopecia Hair loss as a result of tight plaiting of hair or use of hot oil and tension on the scalp. Gradual damage to hair follicles occurs, leading to hair thinning and loss.

Traumatic brain injury Nondegenerative, noncongenital insult to the brain from an external mechanical force.

Triad asthma A subcategory of drug-induced asthma representing a combination of intrinsic asthma, aspirin sensitivity, and nasal polyposis.

Triaxial joint A joint that permits movement around three axes so that motion can occur in three planes. Permits gliding movement between two bones as exemplified by the carpal joints of the hand.

Tricuspid valve The cardiac valve that lies between the right atrium and right ventricle. The valve is normally closed during ventricular systole and open during ventricular diastole. Tricuspid valve closure contributes to heart sound S_1.

Trigeminal neuralgia A neurologic condition of the trigeminal cranial nerve characterized by paroxysms of pain on one side of the face.

Triglyceride A simple fat compound consisting of three molecules of fatty acid and glycerol.

Trigone Triangular area, usually referring to the bladder muscle.

Triiodothyronine (T_3) An active hormone containing three iodine molecules secreted by the thyroid gland and produced in tissues by the enzymatic deiodination of T_4.

Trisomy Having three homologous chromosomes instead of the usual pair, as in trisomy 21, also called *Down syndrome.*

Tropic/trophic hormone A hormone that stimulates the growth and maintenance of a tissue or gland.

Tropical sprue Inflammation of the mucosa of the small intestine secondary to infection.

Tropomyosin A structural protein involved in regulation of actin–myosin cross-bridge formation. Associated with the thin filament of the sarcomere in skeletal and cardiac muscle, where it blocks cross-bridge formation when intracellular calcium levels are low.

Troponin A regulatory protein involved in regulation of actin–myosin cross-bridge formation. Associated with tropomyosin and actin in the thin filament of the sarcomere. Troponin binds calcium ions in the cell and regulates the position of tropomyosin, allowing muscle contraction when intracellular calcium levels rise.

Tuberculosis A chronic infection caused by the acid-fast bacillus *Mycobacterium tuberculosis*.

Tumor marker Biochemical substance, such as a specific enzyme, receptor, or surface protein, that helps identify the tumor cell.

Tumor suppressor gene A gene that regulates a group of growth-promoting genes and suppresses tumor formation. Mutation and underexpression of tumor suppressor genes are associated with the development of cancer.

Tunica albuginea Fascial layer covering the testes and erectile bodies of the penis.

Turbulent flow The friction and increased resistance caused by nonlinear flow in airways or blood vessels.

Twitch The mechanical response to a single stimulus of a motor unit.

Type 1 diabetes mellitus Insulin-dependent diabetes mellitus; characterized by an absolute deficiency of insulin; usually from autoimmune destruction of pancreatic beta cells.

Type 2 diabetes mellitus Non–insulin-dependent diabetes mellitus; characterized by tissue insulin resistance and insufficient insulin production by the pancreas.

U

Ulcerative colitis An inflammatory disease of the mucosa of the rectum and colon. Most commonly it affects the most distal portions of the colon, but eventually it may affect the entire colon. It is typically characterized by exacerbations and remissions. The clinical manifestations of ulcerative colitis are abdominal pain, diarrhea, and rectal bleeding.

Ultrafiltration Filtration through a filter capable of removing colloidal particles from a dispersion medium, as in the filtration of plasma at the capillary membrane.

Ultrasound An imaging modality that uses sound waves to assess the size, structure, and function of internal organs, tissue, or the fetus.

Uniaxial joint A joint that allows motion around a single axis.

Unipennate A muscle with a lateral tendon to which the fibers are attached obliquely, like one-half of a feather.

Unipolar depression Depression without periods of mania.

Unresolved urinary tract infection (UTI) A urinary tract infection in which bacteriuria remains after initial antibiotic treatment.

Unsaturated (fatty acid) Referring to an organic compound in which one or more pairs of carbon atoms are united by double or triple bonds, as in unsaturated fatty acids.

Up-regulation An increase in the number of cell receptors for a specific hormone resulting from chronically low concentrations of the hormone. Up-regulation helps maintain the target cell response to a hormone, even when circulating hormone levels are low.

Urea Substance produced in the liver from the breakdown of protein and excreted in the urine.

Uremia A clinical syndrome related to the severe loss of renal function and resulting in the accumulation of metabolic waste products in the blood.

Ureter One of a pair of fibromuscular, mucosa-lined narrow tubes that connect the kidneys to the bladder.

Ureterocele Congenital cystic dilatation of the distal ureter.

Ureterolithiasis The presence of calculi (stones) in the urinary system, particularly the ureters.

Ureteropelvic junction obstruction Disruption of urinary flow from the kidney(s) into one or both ureters.

Urethral stricture A fibrotic narrowing of the urethra, usually composed of scar tissue.

Urethral valve The most common cause of urinary obstruction in male newborns and infants. Posterior in location, occurring in the distal prostatic urethra, urethral valves are mucosal folds that resemble thin membranes and cause obstruction when the child attempts to void.

Urethritis Inflammation of the urethra.

Urethrorectal fistula A rare congenital anomaly almost always associated with an imperforate anus. The fistula results from failure of the urorectal septum to develop completely, leading to a persistent communication between the rectum posteriorly and the urogenital tract anteriorly.

Urge incontinence A strong and immediate urge to void instigated by involuntary detrusor overactivity.

Urinary bladder Muscular sac located in the anterior inferior pelvic cavity that holds urine until it is excreted through the urethra.

Urodynamic testing The study of the mechanics of urinary bladder filling, emptying, and voiding.

Urolithiasis The presence of calculi (stones) in the urinary system.

Urothelial tumor A malignant tumor of the lining of the renal pelvis, calyces, ureter, and bladder.

Urothelium Epithelial lining of the urinary tract from the renal pelvis to the bladder.

Urticaria A pruritic (itchy) skin eruption characterized by transient wheals of varying shapes and sizes with well-defined erythematous margins and pale centers.

Uterine prolapse A sinking of the uterus from its normal position. It usually occurs when supporting structures, such as the uterosacral ligaments and cardinal ligaments, relax, altering the relationship of the uterus to the vaginal axis.

V

Validity The extent to which a test measures what it is intended to measure.

Valvular incompetence An acquired or congenital disorder of a cardiac valve resulting in regurgitation of blood through the valve because it doesn't close properly; also called *valvular insufficiency*.

Varicose veins Incompetency of the superficial veins of the extremities, producing engorgement.

Vas deferens Thick, muscular tube that is continuous with the epididymis. It travels along the pelvic wall and joins with the seminal vesicle duct at the prostate to form the ejaculatory duct. Also called *ductus deferens*.

Vascular fluid Fluid that is in blood vessels; a component of extracellular fluid.

Vasculitis Inflammation of the lining (intima) of a blood vessel.

Vasoconstriction A decrease in the diameter of a blood vessel, usually referring to an arteriole, caused by contraction of vascular smooth muscle.

Vasodilation An increase in the diameter of a blood vessel, usually referring to an arteriole, caused by a relaxation of the smooth muscles in the vessel wall.

Vasospasm Sudden inappropriate constriction of a blood vessel, producing obstruction of flow.

Vellus Tiny hair strands that are almost unnoticeable. Vellus covers the whole body of a child, except the palms and soles. In contrast, terminal hair strands cover the arms and legs of adults.

Ventilation The process of moving air into the lungs and distributing air within the lungs to gas exchange units (alveoli) for maintenance of oxygenation and removal of CO_2.

Ventricle A fluid-filled compartment as in the brain or heart.

Ventricular ejection The forceful expulsion of blood from the ventricles into the aorta and the pulmonary arteries.

Ventriculoperitoneal shunt A shunt that extends all the way from the ventricular system of the brain to the peritoneal cavity of the body.

Verrucae Circumscribed elevations of the epidermis, commonly called *warts*.

Vertigo A sensation of loss of equilibrium and spinning that frequently occurs with head movement; more than just dizziness.

Verumontanum A small elevation that is marked by a midline opening from the prostatic utricle; a remnant of the müllerian duct system.

Vesicle A portion of lipid bilayer membrane that forms a sphere and surrounds substances to be transported to a destination within the cell or plasma membrane.

Vesicoureteral reflux Retrograde movement of urine from the bladder to the kidney as a result of a disruption in the normal valvular mechanism at the ureter–bladder junction.

Vibrissae Large hairs of the nasal cavity.

Virion A virus particle.

Virulence The capacity of a microorganism to successfully evade host defenses and cause disease.

Virulent Referring to a microorganism's ability to evade host defenses and cause disease.

Viruses Tiny genetic parasites that are dependent on the host cell for replication. They take over the host cell "machinery" for energy and replication.

Visual acuity The ability of the eyes to focus clearly on an image at a known distance; it is assessed with a Snellen chart and often expressed as the ability of a person to accurately discern characters of various sizes at a distance of 20 feet.

Vitamin An organic substance that is essential in the diet and is required for the body's utilization of energy-containing nutrients.

Vitiligo Patch of depigmentation; also called *leukoderma*.

Voiding dysfunction A failure in the normal process of bladder emptying that results in urinary retention and/or incontinence.

Volvulus A twisting of the bowel on itself, which results in blood vessel compression.

Vulvovaginitis An inflammation of the vulva and vagina.

W

Wallerian degeneration Complete disintegration of the distal portion of an axon that has been severed from the cell body. The axon, myelin sheath, and terminal arborization all disintegrate.

Wart Also called *verruca;* caused by a virus that provokes a benign proliferation of keratinocytes.

Water imbalance Imbalance of body fluid concentration; osmolality imbalance; hypernatremia and hyponatremia.

Western blot A laboratory test to detect the presence of antibodies to specific antigens. It is regarded as more precise than the enzyme-linked immunosorbent assay (ELISA) and is sometimes used to check the validity of ELISA tests.

White blood cell A cell that mediates immune function. White blood cells include granulocytes, monocytes, and lymphocytes. Also called *leukocyte*.

White-coat phenomenon Elevated blood pressure readings when measured in a clinic setting by a nurse or physician.

Wilson disease A rare autosomal-recessive disorder in which excessive amounts of copper accumulate in the liver or other organs; also called *hepatolenticular degeneration*.

Wolffian ducts Mesonephric ducts that develop as nephric ducts but mature to form male genital ducts.

Wolff law States that bone is established where needed and resorbed where not needed.

X

Xerostomia Dryness of the mouth caused by cessation of normal salivary secretion.

Z

Z line A dark band that defines a sarcomere; Z lines are perpendicular to actin and myosin filaments. A sarcomere extends from one Z line to the next.

Zeitgeber A cue given by the environment, such as a change in light or temperature, to reset the internal body clock.

Zona pellucida The thick covering of the ovum.

Zone of calcifying cartilage A very thin line of chondrocytes and the weakest segment of the epiphyseal plate.

Zone of maturing cartilage Contains the enlarged and mature cartilage cells as they migrate toward the metaphysis.

Zone of resting cartilage Maintains adherence of the plate to the epiphysis.

Zone of young proliferating cartilage Demonstrates the most active cartilage cell growth.

Zygote The developing ovum, from the time it is fertilized until it is implanted in the uterus.

INDEX

A

A band, in sarcomeres, 1015
ABC transporters, 42, 43*f*
Abdominal pain, 721–723
Abducens nerve (CN VI), 861*t*, 863, 864*f*
Ablation procedures, for dysrhythmias, 431
Abnormal uterine bleeding patterns, 673
Abortion, spontaneous, 681
Abruptio placentae, 681
Abscess(es)
  brain, 912–913, 913*b*, 913*f*
  breast, 682
  liver, 775–776
  pancreatitis and, 750
Absence seizures, 916
Absorption, of nutrients, 715
Acalculous cholecystitis, 747
Acarbose, for diabetes, 830
Accessory conduction pathways, 429
Acclimatization, 6
Accommodative capacity, loss of, in presbyopia, 944
Accumulations, intracellular, in cell injury, 60–62, 62*f*
Accuracy, of data, 5
ACE inhibitors, 569, 569*t*
Acetaminophen poisoning, 775, 775*f*
Acetylcholine (ACh)
  deficiency of, in schizophrenia, 978*t*
  in neuromuscular junctions, magnesium imbalances and, 534, 535*f*
  as neurotransmitter, 867, 871*f*, 875–877, 876*b*, 876*t*, 877*f*
Acid-base homeostasis, 541–543, 545*b*
  buffers in, 542
  disorders of, 545–549. *see also* Acid-base imbalances
  regulation of, renal tubules in, 565–566, 567*f*
  renal contribution to, 543, 543*t*
  respiratory contribution to, 542–543, 543*t*
Acid-base imbalances, 545–549, 549*b*
  geriatric considerations in, 550*b*
  laboratory measurements in, 546*f*
  metabolic acidosis as, 545–546, 545*b*
  metabolic alkalosis as, 547–548, 548*b*
  mixed, 549
  pediatric considerations in, 550*b*
  respiratory acidosis as, 546–547, 547*b*
  respiratory alkalosis as, 548–549, 548*b*
Acidemia, definition of, 545
Acidosis
  definition of, 545
  metabolic, 545–546, 545*b*
  respiratory, 546–547, 547*b*
Acne vulgaris, 1071–1072, 1072*f*
  in adolescence, 1090
Acquired immunodeficiency syndrome (AIDS), 233–258, 237*t*. *see also* AIDS (acquired immunodeficiency syndrome); HIV-1 and HIV-2; HIV/AIDS; HIV infection
Acrochordons, 1091*f*
Acromegaly, 802, 802*f*
Acrosome, of spermatozoa, 637, 638*f*
Acrosome reaction, 639–640
Actin filaments, 30, 32*f*
  of cardiac myocytes, 363, 365–366, 367*f*
  in skeletal muscle, 1015
Action potential(s), 45–46, 46*f*–48*f*
  cardiac, 369–370, 370*f*
  in cardiac cells, 46
  in nerve and muscle cells, 872
  in nerve cells, 45, 46*f*–47*f*
Activated partial thromboplastin time
  alterations in, in hemostatic disorders, 305*t*
  normal value and significance of, 305*t*
Active immunity, 187, 188*f*
Active transport pumps, 41–42, 42*f*
Acute angle-closure glaucoma, 950, 950*f*
Acute bronchitis, 483–485
  clinical manifestations of, 485
  diagnosis of, 485
  etiology of, 483–485
  pathogenesis of, 485
  treatment of, 485
Acute cardiogenic pulmonary edema, in heart failure, 418
Acute cholecystitis, 747–748
Acute coronary syndrome (ACS). *see also* Myocardial infarction
  clinical features and management of, 388–393
  electrocardiographic changes in, 390–391, 391*f*
  etiology and pathogenesis of, 389, 389*f*–390*f*
Acute hyperglycemia, 825
Acute idiopathic polyneuropathy, 931
Acute infective endocarditis, 398
Acute kidney injury (AKI), 593–600, 600*b*–601*b*
  after burn injury, 1100
  clinical management of, 604, 607*b*
  clinical presentation of, 596–600, 596*f*, 597*t*–599*t*
  complicating shock, 449
  etiology and pathophysiology of, 594–596
  intrinsic/intrarenal, 595–596, 595*b*. *see also* Acute tubular necrosis
  oliguric phase of, 599–600
  postoliguric phase of, 600
  postrenal, 595, 595*b*
  prerenal, 594–595
  prodromal phase of, 599
  staging of, 606*t*
    Rifle classification for, 594*t*
  types of, 595*b*
Acute lymphoblastic leukemia, 223–224, 224*f*, 224*t*
Acute myeloid leukemia, 222, 222*f*
Acute pancreatitis, 748–750, 749*b*, 750*f*, 751*b*
Acute pericarditis, 402–403
Acute radiation sickness, signs and symptoms of, 74*f*
Acute renal failure. *see* Acute kidney injury
Acute respiratory distress syndrome (ARDS), 504–506, 507*b*
  clinical manifestations of, 505
  complicating shock, 449
  diagnosis of, 505–506, 507*f*
  etiology of, 504–505, 505*b*
  pathogenesis of, 505, 506*f*–507*f*
  treatment of, 506
Acute respiratory failure (ARF), 469–470, 471*b*
Acute rheumatic fever, 1052, 1053*b*
Acute tracheobronchial obstruction, 495–496
Acute tubular necrosis (ATN), 595
  pathogenesis of, 596, 596*f*
  phases of, 596*f*
Acute viral hepatitis, 767
  in elderly, diagnosis of, 781
Acyanotic congenital defects, 406–408
Adam's apple, 453–454
Adaptation, in general adaptation syndrome, 14–16, 14*t*
Adaptive immunity, 176–187
  cell-mediated, mechanisms of, 178–181. *see also* Cell-mediated immunity
  humoral, mechanisms of, 181–183. *see also* Humoral immunity
  major histocompatibility complex in, 176. *see also* Major histocompatibility complex
Adaptive response(s)
  allostasis in, 13, 20*b*, 23*b*
  to cell injury, 63, 63*f*. *see also* Cell adaptation
  coping in, 20–23
  endorphins and enkephalins in, 19
  general adaptation syndrome and allostasis in, 14–16, 15*f*
  growth hormone in, 19
  homeostasis in, 12–13
  illness and, 20–23
  immune cytokines in, 19
  neurohormonal mediators of, 17–19. *see also* Neurohormonal mediators, of stress and adaptation
  oxytocin in, 19
  prolactin in, 19
  sex hormones in, 19
  stress as concept and, 13–17
  stressors and risk factors in, 16–17
Addison disease, 806
Addisonian crisis, 806
Adenocarcinoma
  bladder, 617
  of prostate cancer, 654
  pulmonary, 475
Adenohypophysis, 788
  endocrine cell types of, 789*t*

Page numbers followed by *f* indicate figures; *t*, tables; *b*, boxes. **Bold entries** designate diseases or syndromes.

Adenomas
metanephric, 579*t*
renal cortical, 579*t*
Adenosine, as neurotransmitter, 876*b*, 876*t*, 879–880
Adenosine triphosphate (ATP)
in cellular metabolism, 34–39, 38*f*, 39*b*
citric acid cycle and, 36
glycolysis and, 36, 37*f*
oxidative phosphorylation and, 36–39, 39*f*
loss of, in acute brain injury, 892–893, 893*f*
mitochondrial production of, 34, 39*f*
in nutrient absorption, 715
Adherence, microbial, 144
Adhesive capsulitis, 1022–1023
Adolescents
blood pressure classification in, 338*t*
skin disorders in, 1090
Adrenal glands, 794–795, 795*f*
Adrenal medulla disorders, 810–811, 811*b*
Adrenocortical hormone disorders, 806–810, 806*b*, 807*f*–810*f*, 810*b*
adrenocortical insufficiency as, 806–808, 806*b*, 807*f*
congenital adrenal hyperplasia as, 808, 808*f*
hyperaldosteronism as, 809–810
hypercortisolism as, 806*b*, 808–809, 809*f*–810*f*
Adrenocortical hormones, 667
Adrenocortical insufficiency, 806–808, 806*b*, 807*f*
Adrenocortical steroids, in stress and adaptation, 18–19
Adrenocorticotropic hormone (ACTH), secretion of, 793
Adrenogenital syndrome, 808
Adult-onset Still disease, 1055–1056, 1056*b*
Aerosols, in skin care, 1086
Afferent nerve fibers, in gastrointestinal motility, 705
Afferent sensory pain fibers, 956, 956*t*, 957*f*
Affinity, receptor, 787
Affinity maturation, 185–186
Afterload, 337–338
in shock, 447
stroke volume and, 376
Agammaglobulinemia, Bruton X-linked, 212
Age
atherosclerosis risk and, 329
breast cancer risk and, 684
in epidemiology, 7
features contributing to restrictive lung disease related to, 501*t*
hematologic value changes related to, 262*t*
HIV/AIDS rates and, 234
hypertension and, 344
infection risk and, 144
maternal, Down syndrome frequency related to, 101, 102*t*
respiratory anatomic and physiologic variations by, 457, 459*t*
Age differences, in normality, 6
Age-related macular degeneration, 948, 949*f*
Aging
free radical theory of, 75
healthy, 9, 10*f*
physiologic changes in, 75, 75*b*, 75*t*
in blood vessels, 329
cellular, 74–75, 74*f*
in circulatory system, 315, 317*b*
in female reproductive system, 669*b*
in gastrointestinal system, 739
Aging *(Continued)*
in hearing, 939, 941*b*
in heart, 359*b*
in integumentary system, 1059–1061, 1060*b*, 1060*f*
in male reproductive system, 638*b*
metabolic, 846
in nervous system, 881, 881*b*–882*b*
in renal function, 570, 570*b*–571*b*
in respiratory system, 458*b*
in skeletal system, 1004–1006, 1005*b*–1006*b*
programmed senescence theory of, 74–75
Aging processes, 7
Agitation, psychomotor
in bipolar disorder, 983
in major depression, 980–981
Agnosia, in dementia, 918
Agonists, hormone, 788
AIDS (acquired immunodeficiency syndrome), 233–258. *see also* HIV-1 and HIV-2; HIV infection
conditions indicative of, 246*b*
epidemiology of, 234–238
incidence and deaths from, estimated, in U.S. 1985-2014, 236*f*
infections in patient with, agents of, 247*b*
progression of HIV infection to, 244–245, 244*f*
Air emboli, 325
Air pollutants
indoor, health effects of, 72*t*
outdoor, health effects of, 72*t*
Airbag burns, chemicals associated with, 1110
Airborne transmission of infection, 148
Airway lumen, obstructive conditions of, 491–496, 496*b*
acute tracheobronchial obstruction as, 495–496
bronchiectasis as, 491–493, 492*f*–494*f*
bronchiolitis as, 493–494
croup syndrome as, 496
cystic fibrosis as, 494–495
epiglottitis as, 496
Airway obstruction, in burn injury, 1101, 1101*f*
Airway resistance, ventilation and, 461–462, 462*f*
Alagille syndrome, 780
Alarm stage, of general adaptation syndrome, 14, 14*t*, 15*f*
Albinism, 105–106, 1081, 1081*b*
Albumin, serum, 261
Alcohol
cancer risk and, 120–123
teratogenicity of, 111
Alcoholic fatty liver, 773
Alcoholic hepatitis, 773–774
Alcoholic liver disease, 773–774
Aldosterone
excessive production of, 809
fluid volume and, 18–19
in renal tubular fluid reabsorption, 568–569, 568*t*
secretion of, 795, 796*f*
in stress and adaptation, 18–19
Aldosterone inhibitors, as potassium-wasting diuretics, 569–570, 569*t*
Alkalemia, definition of, 545
Alkalosis
definition of, 545
metabolic, 547–548, 548*b*
respiratory, 548–549, 548*b*
All-or-none response, of motor unit, 1016
Alleles, dominant and recessive, 95–96, 98*f*
Allergens, asthma triggered by, 479, 479*f*
Allergic asthma, 479–480, 479*f*
Allergic contact dermatitis, 207–209
Allergic purpura, 304
Allergic skin responses, 1072–1074
Allergy, 1072–1073
Allogeneic stem cell transplantation, 220*f*, 222
Alloimmunity, 201
Allostasis, 13
general adaptation syndrome and, 14–16
Allostatic overload. *see also* Stress
illness and, 21–23, 22*f*
Alopecia
androgenic, 1084, 1085*f*
traction, 1081, 1082*f*
Altitude, high, test data and, 6
Alveolar-capillary membrane, 466–467, 466*f*
Alveolar hypoxia, 466
Alveolar ventilation/oxygenation, 461
Alveoli, 454, 455*f*
Alzheimer disease, 918
Amblyopia, 945–947
Amenorrhea, 672–673, 672*f*
Amine neurotransmitters, 876*b*, 876*t*, 877, 878*f*
Amino acids, 714
chemical structures of, 85*f*
excitatory, in acute brain injury, 893–894, 894*f*
in glomerular filtration regulation, 563–564
as neurotransmitter, 876*b*, 877–879
RNA codons for, 80, 82*t*
Amniocentesis, in prenatal diagnosis, 10, 114
Amnion, 663
Amniotic cavity, 663
Amniotic fluid emboli, 325
Amphiphilic molecular structure, 27–28, 28*f*
Ampulla, of vas deferens, 637–639
Amputation, following electrical injury, 1108
Amylase, serum, activity of, increased, causes of, 751*b*
Amylins, for diabetes, 830
Amyloid plaques, 918, 918*f*
Amyotrophic lateral sclerosis (ALS), 928
clinical manifestations and treatment of, 928, 928*b*
etiology and pathogenesis of, 928
respiratory effects of, 511, 512*t*
Anabolism, 36, 839
Anal agenesis, 699
Analgesia, stress-induced, 959–960
Anaphylactic shock, 443–444
Anaphylaxis
triggers of, 443*b*
in type I hypersensitivity, 200, 201*b*
Anaplasia, 118, 118*f*
Androgenic alopecia, 1084, 1085*f*
Andropause, 648
Anemia, 272–274, 285*b*–286*b*
absolute, 272, 274*b*
laboratory findings in, 275*t*–276*t*
in acute blood loss, 285
in antibody-mediated drug reactions, 284–285, 284*t*
aplastic, 277. *see also* Aplastic anemia
in cancer, 137
of chronic renal failure, 277–278, 278*f*
laboratory findings in, 275*t*–276*t*
classification of, 274*b*
complicating chronic kidney disease, 603
management of, 605
definition of, 272

Anemia *(Continued)*
in extrinsic destruction of erythrocytes, 284–285
general effects of, 272–274
in glucose-6-phosphate dehydrogenase deficiency, 282–284
in hematologic neoplasms, 217
hemolytic, 281
in hemolytic disease of newborn, 284
in hereditary spherocytosis, 282
in inherited red blood cell disorders, 279–284
iron deficiency, 279
laboratory findings in, 275*t*–276*t*
in leukemia, management of, 219
other extrinsic abnormalities of, 285
pernicious, 278
relative, 272, 274*b*
laboratory findings in, 275*t*–276*t*
sickle cell, 281–282
in thalassemia, 279–281, 280*f*
transfusion therapy for, 286, 287*t*–289*t*
from vitamin B_{12} or folate deficiency, 278–279
Anemic hypoxia, 468
Anergy, in HIV infection, 243
Anergy test, in monitoring HIV disease status, 246
Aneuploidy, 99–100, 100*f*
Aneurysm(s), 325
aortic, dissecting, 331, 331*f*
cerebral, 909–910, 910*f*
classifications of, 330–331, 331*f*
clinical manifestations and diagnosis of, 331
treatment of, 331
Angelman syndrome, 110, 110*f*
Angina pectoris
classic, 386
Prinzmetal variant, 388
Anginal ischemia, 388
Angiogenesis, in tumor growth, 133
Angiography
in atherosclerosis diagnosis, 330
coronary, 379–380, 380*f*–381*f*
Angiomyolipoma, renal, 579*t*
Angiotensin II (AII)
in glomerular filtration regulation, 563
in myocardial remodeling and progression, 415
in renal tubular fluid reabsorption, 568–569, 568*t*
Angiotensinogen, 342–343
Anhydrous ammonia, burns from, 1109–1110
Ankylosing spondylitis, 1051–1052, 1052*f*
respiratory effects of, 512–513
Anomic aphasia, after stroke, 908
Antagonists, hormone, 788
Anterior hemiblock, 430
Anterolateral somatosensory tract, 883, 884*f*
Anterolateral tract, in pain transmission, 957, 958*f*
Anthropology, medical, 7
Antibiotic-associated colitis, 731
Antibiotic therapy, topical, for thermal burn wounds, 1103, 1104*t*
Antibody(ies)
drug reactions mediated by, anemia in, 275*t*–276*t*, 284–285, 284*t*
structure of, 183–187
Antibody-dependent cell-mediated cytotoxicity (ADCC), 166
Antibody titer, 187
Anticipatory anxiety, in panic disorder, 991
Anticoagulation therapy, for ischemic stroke, 907–908
Anticonvulsants, for bipolar disorder, 983–984
Antidepressants, for major depression, 981, 981*f*
Antidiabetic agents, oral, 829–830, 829*t*
Antidiuretic hormone (ADH)
action of, in renal tubules, 567–568, 568*t*, 569*f*
inappropriate secretion of, syndrome of, 813–814, 813*b*, 813*f*
secretion of, 789–790, 790*f*
Antidiuretic hormone disorders, 812, 813*b*–814*b*, 813*f*
Antigen(s)
integrated response to viral, 191–192
new, integrated immune response to, 189–191, 192*b*
sequestered, triggering autoimmunity, 195–196
Antigen recognition, by B cells, 175*b*–176*b*, 181–183, 182*f*
Antigenic mimicry theory, of autoimmunity, 195
Antihemophilic factor, action of, 300*t*
Antihistamines, for type I hypersensitivity, 200–201
Antimicrobial resistance, 145
Antioxidants, cancer risk and, 123
α_1-Antiprotease, in phagocytosis, 172
Antipsychotics
for bipolar disorder, 983
for schizophrenia, 976, 977*f*
Antiretroviral therapy (ART), 254–255
medications in, classes of, 255, 255*f*, 256*t*
α1–Atitrypsin deficiency, 779
Anxiety
anticipatory, in panic disorder, 991
avoidance, in panic disorder, 991
prevalence of, in women, 984, 985*t*
Anxiety disorders, 989–995
generalized, 991–992
obsessive-compulsive disorder as, 992–994, 993*f*
panic disorder as, 990–991
posttraumatic stress disorder as, 994–995, 995*b*
Aorta, coarctation of, 408, 408*f*
hypertension in children from, 348–349
Aortic aneurysm, dissecting, 331, 331*f*
Aortic events, in cardiac cycle, 358*f*, 360
Aortic regurgitation, 397–398, 398*f*
Aortic stenosis or atresia, 397, 397*f*, 408
Aortic stiffening, 360
Aortic valve, 355, 356*f*
Aortic valve disorders, 397–398
regurgitation as, 397–398, 398*f*
stenosis as, 395*f*, 397, 397*f*
Aphasia, after stroke, 908
Aplastic anemia, 277
drugs associated with, 277*t*
laboratory findings in, 275*t*–276*t*
Apoptosis, 65*f*, 66–68, 69*f*
in CD4 T-cell depletion in HIV infection, 243
induction of
chemotherapeutic agents in, 219
Fas ligand in, 67–68, 68*f*
intrinsic pathway of, 68, 69*f*
radiation-induced, 73, 73*f*–74*f*
"survival" signal withdrawal in, 67–68, 67*f*
mitochondria in, 34
radiation-induced, 73, 73*f*
Appendages, of integumentary system. *see also* Hairs; Nails; Sebaceous glands; Sweat glands
age-related changes in, 1061, 1061*b*
Appendicitis, 731
Appetite, regulation of, 841–844, 844*b*
genetics, epigenetics, and environment in, 841–842
hormonal, 842–843, 842*f*, 842*t*
Appetite disturbance, in depression, 980–981
Aquaporins, 41, 557–558, 567, 569*f*
activation of, 789–790, 790*f*
Arachnoid, 851, 852*f*
ART. *see* Antiretroviral therapy
Arterial blood flow alterations, 326–331, 331*b*
in acute arterial occlusion, 331
in aneurysms, 330–331, 331*f*
in atherosclerosis, 326–330
in Raynaud syndrome, 330
in thromboangiitis obliterans, 330
Arterial blood oxygen content (CaO_2), 269, 273*t*
monitoring of, in shock, 447–448
Arterial blood pressure, 337–340, 338*f*, 340*b*–341*b*
Arterial peripheral vascular disease, 326–327, 326*b*. *see also* Atherosclerosis
Arterial ulcer, 324, 327*f*
Arteriohepatic dysplasia, 780
Arteriolar sclerosis, 326
Arteriosclerosis, 325. *see also* Atherosclerosis
Arteriovenous fistulas (AVFs), 325
Arteriovenous malformation (AVM), 325, 911, 911*b*
Arteritis, blood flow alterations from, 325
Artery(ies). *see also* Blood vessels
anatomy of, 316, 316*f*
great, transposition of, 408–409, 409*f*
occlusion of, acute, 331
systemic, primary, 314*f*
thrombosis in, 323
Arthritis. *see also* Osteoarthritis
acute gouty, 1054–1055
enteropathic, 1053–1054
infectious, 1045
juvenile idiopathic, 1056–1057
psoriatic, 1053
reactive, 1052
rheumatoid, 1046–1050. *see also* Rheumatoid arthritis
Arthropathy, hemophilic, 1054
Arthus reaction. *see* Type II hypersensitivity
Articular cartilage, structure and function of, 1011–1012, 1012*b*
Articulations, 1006. *see also* Joint(s)
Ascites
in advanced liver disease, 764–765, 765*f*
pancreatic, 750
Asperger syndrome, 997
Aspergillus, in etiology of pneumonia, 514
Asphalt burns, 1110
Asthma, 479–482, 482*b*
acute episode of, laboratory values in, 547*t*
allergic, 479–480, 479*f*
clinical manifestations of, 481
diagnosis of, 481, 482*f*–484*f*
drug-induced, 479–480
etiology of, 479–480, 479*f*
exercise-induced, 479
food additive-induced, 480
nonallergic (intrinsic), 479
occupational, 479
pathogenesis of, 480–481, 480*f*
severity of
classifying, 481, 482*f*–483*f*
treatment based on, 484*f*
treatment of, 481–482, 485*f*
severity-based, 484*f*

Asthmatic bronchitis, 483–485
Astigmatism, 944–945
Astrocytes, 853, 856*f*, 868, 873*f*
Asystole, electrical, 424, 425*f*
Ataxia, in cerebral palsy, 922
Atazanavir (Reyataz), 256*t*
Atelectasis, 461
Atelectatic pulmonary disorders, 504–508
Atherosclerosis, 325
  clinical manifestations and diagnosis of, 330
  coronary
    coronary heart disease and, 383
    risk factors and mechanisms of, 383–385, 384*f*, 385*b*, 385*t*, 386*f*
  etiology and pathogenesis of, 326–327, 326*f*, 328*f*
  hypertension and, 346
  pathogenesis of, 384, 387*f*
  risk factors for, 327–330, 327*b*
    modifiable, 327–329
    nonmodifiable, 329–330
  treatment of, 330
Atherosclerosis obliterans, 326–327
Athlete's foot, 152
Atmospheric pressure changes, cell injury from, 72, 72*t*
Atonic seizures, 916
Atopic dermatitis, 1072–1073, 1073*f*
Atopic hypersensitivity, 199–201. *see also* Type I hypersensitivity
Atopy, 1072–1073
ATP. *see* Adenosine triphosphate
Atrial dysrhythmias, 425–426
Atrial events, in cardiac cycle, 358*f*, 360
Atrial fibrillation, 426, 426*f*
Atrial flutter, 426, 426*f*
Atrial natriuretic peptides (ANPs)
  renal tubular fluid reabsorption and, 568–569, 568*t*
  synthesis of, by cardiac myocytes, 376
Atrial septal defect, 406–407, 407*f*
Atrial tachycardia, paroxysmal focal, 425–426, 426*f*
Atrioventricular conduction disturbances, 428–429, 428*f*–429*f*
Atrioventricular node, 371
Atrioventricular valve, 355, 356*f*
Atrium(ia), 355–356
Atrophy, 63, 63*f*, 1063*f*–1064*f*
Attention-deficit/hyperactivity disorder (ADHD), 996–997
Atypical absence seizures, 916
Aura, in seizure disorders, 917
Auscultatory gap, in blood pressure measurement, 340, 340*f*
Autism spectrum disorder (ASD), 997–998, 998*b*
Autoantibody induction, in drug-induced immune hemolysis, 285
Autocrine signaling, 49, 50*f*, 124, 784, 784*f*
Autoimmune hepatitis, 772
  diagnosis of, 772
  management of, 772
Autoimmunity (autoimmune disorders), 195–198, 198*b*
  environmental triggers for, 196–197
  genetic factors in, 196, 197*t*
  pharmacotherapies for, 197–198
Autologous stem cell transplantation, 220*f*, 226
Autolysis, postmortem, 75
Automaticity, in dysrhythmias, 421
Automobile airbag burns, chemicals associated with, 1110
Autonomic dysreflexia, in spinal cord injury, 931, 932*t*
Autonomic nervous system (ANS), 867, 867*b*–868*b*, 869*f*–871*f*
  effects on organ system function, 872*t*
Autoregulation
  abnormal, in acute brain injury, 895
  of blood flow, 322
    tissue pressure hypothesis of, 323
    vascular endothelium in, 323
  of coronary circulation, 361
  myogenic, GFR and, 562
Autosomal chromosome disorders, 100–101
Autosomal-dominant disorders, 103–104, 105*t*
Autosomal dominant polycystic kidney disease, 578–579, 578*t*
Autosomal-recessive disorders, 104–106, 106*t*
Autosomal recessive polycystic kidney disease, 578
Autosomes, 104–105
Avascular necrosis, in fracture healing, 1029
Avoidance anxiety, in panic disorder, 991
Avulsion fracture, 1026, 1026*f*
Axoneme, 637
Axons, 91, 92*f*

B

B-cell lymphoma, 229, 230*t*, 231*b*
B lymphocytes (B cells), 159, 167, 167*f*
  activated, class switching in, 184*f*, 185–186, 186*f*
  antigen recognition by, 181–183, 182*f*
  immunodeficiency disorders involving, 210–212
  signaling pathways in, 183*f*–184*f*
  theories of autoimmunity involving, 196
B-type natriuretic peptide (BNP), secretion of, by cardiac myocytes, 376
Bacilli, 149, 151*f*
Bacteremia, 444
  definition of, 445*t*
Bacteria
  antiphagocytic factors in, 145
  cell injury from, 70
  culture and sensitivity testing of, 147*b*
  endospore formation by, 145
  endotoxin produced by, 144–145
  enzymes produced by, 145
  exotoxins produced by, 145
  gram staining of, 149
  morphology of, 149, 151*f*
  pathogenic, 146*t*–147*t*, 150, 150*f*, 153*f*
Bacterial emboli, 325
Bacterial infections
  autoimmunity triggered by, 196
  cutaneous, 1067–1069
Bacterial meningitis, 911, 912*t*
Bacterial peritonitis, spontaneous, 765–766, 766*f*
Bacterial pili, 145*f*
Bacterial pneumonias, 514, 515*t*, 516*f*
Bacterial prostatitis, 652
Balance, 937–942
Ball-and-socket joint, 1010, 1011*f*
Balloon tamponade, of gastroesophageal varices, 762, 763*f*
Band cells, in infection, 163, 163*f*
Barbiturate-hypnotic combinations, for migraine, 964*t*
Baroreceptors
  altered sensitivity to, in orthostatic hypotension, 350
  in blood pressure regulation, 341
  in control of respiration, 464*f*
  in response to heart failure, 413, 414*f*
  in response to shock, 437, 438*f*
Barrett esophagus, 725
Bartholin cyst, 677–678
Bartholin glands, 658
Bartholinitis, 677–678
Basal cell carcinoma, 1079–1080, 1079*f*
Basal ganglia, 855, 859*f*
  injury to, in cerebral palsy, 922–923
  in obsessive-compulsive disorder, 992
Basal metabolic rate (BMR), 839
  factors affecting, 839*t*
  in physiologic stress, 845
  in starvation, 845
Base pairs, in DNA, 78, 79*f*
Basophils, 164
  characteristics of, 263*t*
  functions of, 163*t*
  in type I hypersensitivity, 199
Beau lines, in nail in systemic diseases, 1084, 1085*f*
Becker muscular dystrophy, 1039
Beck's triad, 402
Bedbugs, 1075
Bell palsy, 932, 932*b*, 933*f*
Bence Jones protein, in plasma cell myeloma, 225
Benign neoplasms, specific, 683–684
Benign prostatic hyperplasia (BPH), 651–652, 651*f*–653*f*
Benign tumors
  characteristics of, 118, 118*t*
  nomenclature for, 119*t*
Benzodiazepines, long-acting, for panic disorder, 991
Berger disease, 588
Bernard-Soulier syndrome, 307
Berry aneurysm, 330–331
Biaxial joint, range of movement of, 1010, 1011*f*
Bicarbonate, reabsorption of, in acid-base balance regulation, 565, 567*f*
Bicarbonate buffer system, 542
Bifascicular block, 430–431
Biguanide, for diabetes, 829–830
Bilateral renal agenesis, 578
Bile, physiology of, 743, 744*b*, 745*f*
Bile salts, 743
Biliary atresia, 780
Biliary cirrhosis, 773
Biliary malignancy, 748
Bilirubin
  in erythrocyte destruction, 268
  intracellular accumulations of, 60–61
  metabolism of, 755, 759*f*
    abnormal, in neonatal period, 778
    disorders of, 779–780
Biliverdin accumulation, in systemic diseases, 1083
Biofilm, bacterial, 145
Biopsy, renal, 573
Bipennate muscle, 1014
Bipolar disorder (BD), 982–984
  clinical manifestations of, 983
  etiology and neurobiology of, 982–983
  forms of, 982
  risk factors for, 979*t*

Bipolar disorder (BD) *(Continued)*
treatment of
nonpharmacologic, 984
pharmacologic, 983–984
Bipolar neurons, 868, 871*f*
Bisphosphonates, for osteoporosis, 1035
Biventricular heart failure, 420
Bladder
capacity of, voiding and, 611
innervation of, 610*f*
male, 627–628, 628*f*
neurogenic, 614, 614*b*
structure of, 610*f*
Bladder cancer, 617–618, 617*b*
Bladder disorders, cystocele as, 675
Bladder pain syndrome, 613–614
Bladder tumors, TNM staging of, 619*f*
Blastocyst, 663, 664*f*
Bleeding time
alterations in, in hemostatic disorders, 305*t*
normal value and significance of, 305*t*
Blocking agent, 788
Blood
components of
cellular, 261–262
characteristics of, 263*t*
differing amount of, with age, 262*t*
maturation of, 160*f*, 264*f*
therapeutic indications and actions of, 287*t*–289*t*
organic and inorganic, 261, 261*t*
composition of, 260–262, 260*f*, 262*b*
in gas transport and acid-base balance, 268–270
loss of, acute, anemia in, 285
transfusion of. *see* Transfusion therapy
viscosity of, resistance and, 319
Blood-brain barrier (BBB), 317, 853, 856*f*
ischemia compromising, 895
Blood coagulation factors, 299, 301*f*. *see also* Clotting factors
Blood flow
autoregulation of, 322. *see also* Autoregulation, of blood flow
control of, 322–323, 323*b*
extrinsic mechanisms in, 322
intrinsic mechanisms in, 322–323
coronary, 357, 360–362, 362*b*–363*b*
regulation of, 361–362
vessels, anatomy of, 360, 361*f*–362*f*, 362*t*
distribution of, in shock, 448
laminar, 319–320, 320*f*
in microcirculation, 321–322
pressure and resistance affecting, 318–319, 319*f*
principles of, 317–322, 322*b*
pulmonary, 465–466, 466*b*
turbulent, 319–320, 320*f*
velocity of, 319–320, 320*f*
Blood flow alterations, 313–336. *see also* Arterial blood flow alterations; Lymphatic flow alterations; Venous blood flow alterations
general mechanisms causing, 323–326, 326*b*
from obstructions, 323–325
from structural alterations, 325
Blood flukes, 1076
Blood gases. *see also* Carbon dioxide *entries;* Oxygen *entries*
arterial, normal values for, 467*t*
Blood groups, major, 203*t*
Blood pH, normal range for, 541, 542*t*
Blood pressure
in adults, classification of, 339*t*
alterations in, 337–353
arterial, 337–340, 338*f*, 340*b*–341*b*
in children and adolescents, classification of, 338*t*
diastolic, 338
low, 350–351, 351*b*
measurement of, 338–340
accuracy of, 339–340, 340*t*
auscultated, recommended approach to, 339–340
components of, 338–339
direct, 339
indirect, 339–340
self-monitoring, 340
white coat effect on, 339–340
regulation of, 341–343, 341*f*, 343*b*
long-term, 341–343, 342*f*
short-term, 341
systemic
determinants of, 337–338
normal fluctuations in, 343
systolic, 338
Blood supply
to brain, 853, 855*f*
renal, 553–554, 553*f*–555*f*
Blood urea nitrogen (BUN), in renal function evaluation, 572
Blood vessels
complications of diabetes mellitus involving, 826
endothelium of, 323
geriatric considerations on, 329
hepatic, 755, 757*f*
length of, resistance to flow and, 318, 319*f*
mechanical compression of, blood flow alterations from, 325
permeability of, increased, in inflammation, 171, 172*f*, 172*t*
pressure and resistance of, blood flow and, 318, 319*f*
pulmonary, 465
radius of, resistance to flow and, 318, 319*f*
structure of, 315–317, 316*f*
alterations in, in blood flow and, 325
wall tension and compliance of, 320–321, 320*f*
Blood volume
maternal, in pregnancy, 667–668, 667*f*
regulation of, kidneys in, 567–570, 569*f*
total, 260–261, 260*f*
Blunt trauma, soft tissue, 1024
Body fluid
definition of, 522
extracellular, 522, 522*f*
homeostasis of, 522–524, 523*f*, 524*b*–525*b*
abnormal loss of fluids in, 524
disorders of. *see* Fluid imbalance(s)
fluid distribution in, 522*f*, 523–524
fluid excretion in, 524
fluid intake and absorption in, 522–523
geriatric considerations in, 537*b*
pediatric considerations in, 537*b*
interstitial, 522, 522*f*
intracellular, 522, 522*f*
transcellular, 522
vascular, 522, 522*f*
Body mass index (BMI), 844
Bohr effect, 269
Bone, 1006*b*
age-related changes in, 1004–1006, 1005*b*–1006*b*
bacteria targeting, 153*f*
calcium homeostasis in, 1004
composition of, 1002–1003, 1002*f*–1003*f*
disorders of, complicating chronic kidney disease, 603
management of, 605
functional properties of, 1003–1004
growth and ossification of, 1003, 1004*f*
injury to. *see* Bone fracture(s)
remodelling, 1004, 1005*f*
in fracture healing, 1006
stress on, response to, 1004–1006
structure and function of, 1002–1006
types of, 1003, 1026
Bone and joint tuberculosis, 1032
Bone fracture(s)
compression, 1027, 1027*f*
diagnosis of, 1027–1028
extent of, 1027
healing of, 1006, 1007*f*
healing process in, 1028–1029, 1029*f*
complications of, 1029–1030
treatment of, 1028
types of, 1026–1027, 1026*f*
Bone marrow, in hematopoiesis, 160–161, 160*f*
Bone marrow transplantation (BMT)
for aplastic anemia, 277
for leukemias, 219
Bone mineral density, 1033
Bone remodeling, in fracture healing, 1028, 1029*f*
Bone structure and mass alterations, 1032–1038, 1036*b*
metabolic, 1033–1036
osteomalacia as, 1035
osteoporosis as, 1033–1035, 1034*b*, 1034*f*
Paget disease as, 1035–1036, 1036*f*
rickets as, 1035
scoliosis as, 1032–1033, 1033*f*
Bone tumors, 1036–1038, 1037*f*
malignant, 1037–1038
Bouchard nodes, in osteoarthritis, 1043
Boutonnière deformity, in rheumatoid arthritis, 1047, 1048*f*
Bowel. *see* Intestines
Bowel patterns, alterations in, 723–724
Bowman capsule, 554*f*–555*f*
hydrostatic pressure in, 561, 562*f*
oncotic pressure in, 561
Brachial plexus, 864–865, 865*f*–867*f*, 866*t*
Bradycardia, sinus, 424, 424*f*
Bradykinin, 169, 169*f*
Brain, 853–861, 853*t*, 856*f*
bacteria targeting, 153*f*
blood supply to, 853, 855*f*
brainstem of, 860–861, 861*t*
cerebellum of, 860, 860*f*
cerebrum of, 853–856, 854*t*, 857*f*–859*f*
compression of, from increased intracranial pressure, 897–899
diencephalon of, 856–860, 859*f*
herniation of, from increased intracranial pressure, 897–899, 898*f*–899*f*
hyperactive circuits in, in generalized anxiety, 991
specialization of, 854, 854*t*
stress response pathways of, overbuilt, in panic disorder, 990
support and protection of, 851–853, 852*f*–854*f*

Brain abscess, 912–913, 913*b*, 913*f*
Brain-derived neurotrophic factor (BDNF), reduced activity of, depression and, 973
Brain disorders, chronic, 916–925
cerebellar disorders as, 924–925, 925*b*
cerebral palsy as, 922–923
dementia in, 918–920
hydrocephalus as, 923–924
Parkinson disease as, 920–922, 921*f*
seizures as, 916–917, 916*b*–918*b*
Brain function
basal ganglia in, 855, 859*f*
consciousness and memory in, 886–888, 889*f*
disorders of
acute, 891–914. *see also* Brain injury, acute
chronic, 916–925. *see also* Brain disorders, chronic
mapping of, 854, 858*f*
motor, 883–886
neuronal communication in, 872–880. *see also* Neuronal communication
sensory, 882–883
Brain injury, acute, 891–914
in arteriovenous malformation, 909–911
from central nervous system infections, 911–913, 912*t*
in cerebral aneurysm, 909–911
in cerebrovascular disease and stroke, 905–909. *see also* Cerebrovascular disease and stroke
manifestations of, 900–902
cranial nerve reflexes as, 901–902, 902*f*
Glasgow Coma Scale and, 900–901, 900*b*
level of consciousness as, 900, 900*t*, 901*f*
mechanisms of, 892–899
abnormal autoregulation as, 895
cellular energy failure as, 892–893, 893*f*
excitatory amino acids as, 893–894, 894*f*
increased intracranial pressure as, 895–899, 896*b*, 896*f*
ischemia and hypoxia as, 892–895
primary, 892
reperfusion injury as, 894–895
secondary, 892
traumatic, 902–905. *see also* Traumatic brain injury
Brain waves, in sleep, 888–889, 889*f*
Brainstem, 860–861, 861*t*
BRCA1 and *BRCA2* genes, defects in, breast cancer and, 128–129
Breast(s), 662–663, 663*b*
development of, 662–663
fat necrosis in, 682–683
lactation in, 663
lymphatic drainage of, 316*f*, 686*f*
silicone implants in, reactions to, 683
structure of, 662, 662*f*
Breast abscess, 682
Breast cancer, 684–686
age and, 684
BRCA1 and *BRCA2* gene defects and, 128–129, 684
clinical manifestations of, 684–685, 685*f*–686*f*
continuum of care in, 686
dietary factors in, 684
etiology of, 684
family history and, 684
hormonal factors in, 684
reproductive factors in, 684
Breast cancer *(Continued)*
risk factors for, 684
screening guidelines for, 121*t*, 684
treatment of, 685–686
Breast disorders, 682–686, 686*b*
benign, 683–684
fat necrosis as, 682–683
fibrocystic breast disease as, 683–684, 683*f*
malignant, 684–686. *see also* Breast cancer
mammary duct ectasia as, 682–683, 682*f*
reactions to foreign material as, 683
reactive-inflammatory, 682–683
Breathing, mechanics of, 461, 462*f*
Broca aphasia, after stroke, 908
Brodmann areas, 854, 858*f*
Bronchial epithelium, 452, 454*f*
Bronchial provocation tests, 472
Bronchiectasis, 491–493, 492*f*–494*f*
Bronchioles, 454, 455*f*
Bronchiolitis, 493–494
Bronchitis, 487*b*–488*b*
acute, 483–485
asthmatic, 483–485
chronic, 485–487, 486*f*–487*f*, 488*t*
Bronchoalveolar carcinoma, 476
Bronchopulmonary segments, of lung, 454, 457*f*
Bruit, 320
Bruton X-linked agammaglobulinemia, 212
Buerger disease (thromboangiitis obliterans), 330
Buffers, in acid-base homeostasis, 542
Bulbourethral glands, 628–629, 629*f*
Bulla, 1063*f*–1064*f*
Bundle branch block, 429
Burn injury(ies), 1093–1113
chemical, 1109–1110, 1110*b*
electrical, 1107–1108, 1107*f*, 1108*b*–1109*b*
nutritional considerations in, 848
special populations in, 1110–1111, 1111*b*
thermal, 1094–1107
acute management of, 1097–1099
assessment of, 1099, 1099*b*
cardiovascular dysfunction in, 1101, 1101*f*
cellular changes in, 1102–1103
depth classification of, 1096–1097, 1096*f*, 1096*t*, 1097*b*, 1098*f*
emergent phase of care in, 1103
nutritional support in, 1105–1106
surgery in, 1103–1104
topical antibiotic therapy in, 1103, 1104*t*
wound management in, 1103, 1104*t*
etiology of, 1094
extent of injury from, 1097
immune response in, 1103
incidence of, 1094
integument effects of, 1094–1096
metabolic changes in, 1102
mortality from, 1094
organ dysfunction in, 1101–1102
physiologic changes in pediatric and elderly in, 1095*t*
rehabilitation after, 1106–1107, 1107*b*
renal dysfunction in, 1102
respiratory dysfunction in, 1101–1102, 1101*f*
risk factors for, 1094, 1094*f*, 1095*t*
severity classification of, 1097, 1099*t*
survival of, elements of, 1103
wound healing after, 1106–1107, 1106*f*
Burn shock, 1099–1101, 1100*f*
acute resuscitation for, 1099–1101, 1100*b*
immune response in, 1103
Burn unit referral criteria, 1099*b*
Burr cells, in acute kidney disease, 277, 278*f*
Bursae injuries, 1023
Byler syndrome, 779

C

Cachexia, cancer, 136, 136*f*
Calcitonin, in calcium homeostasis, 811
Calcium
in capacitation of sperm, 639
in clot formation, 300*t*
intracellular accumulations of, in cell injury, 61–62
in muscle contraction, 367, 368*f*, 1015, 1016*f*
plasma, imbalances in, 533–534, 534*b*
Calcium homeostasis, 1004
Calcium overload
in acute brain injury, 894
cellular injury from, 67*f*, 69
Calcium transporters, membrane, 42, 42*f*
Calculi
bladder (vesical), 622
lower urinary tract, 622
renal, 584–586. *see also* Nephrolithiasis
ureteral, 622
Callus formation, in fracture healing, 1006, 1028, 1029*f*
Calluses, epidermal, 1078
Cancellous bone, 1003, 1026
healing in, 1027–1028, 1029*f*
Cancer
biliary, 748
biology of, cell cycle and, 54–55
bladder, 617–618, 617*b*
cervical, 679–680
colon, 738–739, 739*f*, 739*t*
deaths from
1930-2011, 122*f*
2015, 120*f*
effects of, on body, 136–137, 137*b*
endometrial, 680
epidemiology of, 120–123
esophageal, 737
of female genital structures, 679–680, 680*b*–681*b*
genetic mechanisms of, 123–129. *see also* Carcinogenesis
metastasis of, 132–135, 135*b*. *see also* Metastasis
new case of, 2015, 120*f*
nutritional considerations in, 848
ovarian, 680
pain related to, 964
pancreatic, 752
prostate, 652–655, 654*b*, 654*f*
risk factors for, 120–123, 123*b*
nutritional, 120–123
tobacco use as, 120, 122*f*–123*f*
screening guidelines for, 121*t*
of skin, 1079–1080
therapy of, 137–140, 140*b*
chemotherapy in, 138
drug in, 138
gene therapy in, 138–139
immunotherapy in, 138, 139*f*
molecular, 138–139
radiation in, 138
radiation-induced cell death and, 73
stem cell transplantation in, 139–140
surgery in, 137–138

Cancer *(Continued)*
tumor markers in, 133, 135*t*
tumors in
angiogenesis in, 133
grading and staging of, 134–135, 135*t*, 136*f*
vaginal, 680
of vulva, 680
warning signs of, 136*b*
Candidiasis, mucocutaneous, 1065–1067, 1068*f*
Capacitation, of sperm, 639
Capillary(ies)
anatomy of, 316–317, 318*f*
dynamics of, 321–322
Capillary fragility test, normal value and significance of, 305*t*
Capillary hydrostatic pressure, 321
Capillary leak syndrome, in burn injury, 1099, 1100*f*
Capillary permeability, 317
Capillary pressure gradient, 321, 321*f*
Carbohydrates
absorption of, 715
digestion of, 713, 714*t*
metabolism of, 839–840, 843*t*
Carbon dioxide, partial pressure of (PCO_2), 270
arterial ($Paco_2$), in autoregulation, acute brain injury and, 895
Carbon dioxide transport, 270, 271*f*, 467–468
Carbon monoxide poisoning, in burn injury, 1102
Carbonic acid, excretion of, by lungs, 541, 542
Carcinogen(s), 123
chemical, 131*b*
definition of, 130
Carcinogenesis
initiation of, 129–130, 130*f*–131*f*
multistep nature of, 129–132, 132*b*
progression of, 130*f*, 131–132, 132*f*
promotion of, 130*f*, 131
Carcinoma in situ, 64
Cardiac action potential, 369–370, 370*f*
Cardiac catheterization, 379–380, 381*f*
Cardiac cycle, 358–360, 358*f*, 360*b*
aortic and pulmonary artery events in, 358*f*, 360
atrial events in, 358*f*, 360
isovolumic contraction in, 359–360
isovolumic relaxation in, 360
pressure-volume loop in, 358, 358*f*
ventricular ejection in, 360
Cardiac diseases, common sequelae of, 411–433
Cardiac dysrhythmias, 421–431
abnormal rates of sinus rhythm in, 424
abnormal site of impulse initiation in, 425–428, 428*b*
atrial dysrhythmias as, 425–426, 426*f*
escape rhythms as, 425, 425*f*
junctional dysrhythmias as, 426, 427*f*
ventricular dysrhythmias as, 427–428, 427*f*
analysis of, 423–424, 423*f*, 424*b*
normal sinus rhythm and, 423–424, 423*t*, 424*f*
conduction pathway disturbances in, 428–431, 431*b*
abnormal, 429, 429*f*
atrioventricular, 428–429, 428*f*–429*f*
intraventricular, 429–431, 430*f*
Cardiac dysrhythmias *(Continued)*
mechanisms of, 421–422
automaticity as, 421
reentry as, 422, 422*f*–423*f*
triggered activity as, 421–422, 422*f*
treatment of, 431, 432*t*
Cardiac electrophysiology, 369–372, 372*b*
Cardiac energy metabolism, 368–369, 369*b*
oxygen utilization in, 368–369
substrate utilization in, 369
Cardiac function, 354–381. *see also* Heart
alterations in, 382–410. *see also* Heart diseases
curves, 375, 375*f*
tests of, 376–380
cardiac catheterization/coronary angiography as, 379–380, 380*f*–381*f*
computed tomography as, 378
echocardiography as, 378, 379*f*
electrocardiography as, 372–374, 373*f*–374*f*, 374*b*, 376–378, 377*f*–379*f*
magnetic resonance imaging as, 378
nuclear cardiography as, 379
Cardiac muscle, 90, 91*f*. *see also* Myocardium
Cardiac myocyte(s), 363–366, 366*b*
action potential in, 369–370, 370*f*
contractile filaments
characteristics of, 365–366, 366*f*–367*f*
structure of, 363, 365*f*
contraction of, 366–367
calcium in, 367, 368*f*
overview of, 366
sliding filament/cross-bridge theory of, 366–367, 368*f*
energy metabolism in, 368–369, 369*b*
hypertrophy of, in heart failure, 415
oxygen utilization by, 368–369
relaxation of, energy of, 367, 367*b*–368*b*, 369*f*
resting potential in, 369
rhythmicity of
autonomic regulation of, 372
of myocardial cells, 370–371, 371*f*
sarcomeres in, 363, 365*f*
sarcoplasmic reticulum of, 363, 364*f*
structure of, 363, 363*f*–364*f*
substrate utilization by, 369
Cardiac output, 376*b*
determinants of, 374–376
heart rate and, 374–375
monitoring of, in shock, 447
stroke volume and, 375–376
workload and, 376
Cardiac resting potential, 369
Cardiac tamponade, 402
Cardiac valve(s), 355–356, 356*f*
Cardiac valve disorders
acquired, etiologies of, 394*t*
aortic, 397–398
regurgitation as, 397–398, 398*f*
stenosis as, 395*f*, 397, 397*f*
mitral, 394–397
prolapse, 396–397, 396*f*
regurgitation, 396, 396*f*
stenosis, 395–396, 395*f*
Cardiac workload, 376
Cardinal ligaments, in pelvic support, 674
Cardiogenic shock, 439–441, 441*b*
clinical manifestations of, 440
etiology of, 435*b*, 439–440
mechanical assist devices for, 441
Cardiogenic shock *(Continued)*
pathogenesis of, 439–440, 440*f*
pharmacotherapy for, 441
treatment of, 440
Cardiomyopathy(ies), 400–401, 400*f*
chronic ischemic, 393
classification of, 401*t*
definition of, 400
dilated, 400
hypertrophic, 400–401
restrictive, 401
Cardiovascular disorders, 383
complicating chronic kidney disease, management of, 605
hypertension causing, 345–346
Cardiovascular system
circulatory system in, 356–357, 357*f*, 358*b*. *see also* Circulatory system
heart in, 355–356, 355*f*–357*f*, 356*t*. *see also* Heart
HIV infection manifestations involving, 253
Carrier(s)
in contact hypersensitivity, 208
of genetic diseases, 95–96
membrane transport, 42–43, 43*f*
Carrión disease, 9
Cartilage
articular, structure and function of, 1011–1012, 1012*b*, 1012*f*
deterioration of, in osteoarthritis, 1043
Cartilaginous joints, 1007, 1009*f*
Caseous necrosis, 65, 66*f*
Caspases, initiator, in apoptosis, 68, 68*f*
Casts, urinary, in renal disease, 572
Catabolism, 36, 839
in physiologic stress, 845
Catamenial pneumothorax, 508–509
Cataracts, 947
Catecholamines
in glucose metabolism, 843, 843*t*
in stress and adaptation, 17–18, 21
tissues and organ effect of, 18*t*
Catheterization, cardiac, 379–380, 381*f*
Causative agents, 2
CCR5 inhibitors, 257
$CD4^+$ T lymphocytes
in contact hypersensitivity, 208
HIV infection and, 239–240, 240*f*
CD8+ T lymphocytes, activity of, HIV infection survival and, 243
CD95 ligand, in cell-mediated immunity, 181
Celiac disease, 735
Cell(s)
aging of, 74–75, 74*f*
death of, in HIV infection, 242–244, 242*f*
diversification of, in tissue differentiation, 86
division of. *see* Mitosis
growth and proliferation of, regulation of, 54–55, 57*f*
myocardial. *see* Cardiac myocyte(s)
structure and function of, 26–58, 27*f*, 57*t*
growth and proliferation regulation as, 54–55, 57*f*
intercellular communication and growth in, 48–55, 55*b*. *see also* Intercellular communication
intracellular receptor-mediated responses in, 53, 54*f*–55*f*

Cell(s) *(Continued)*
metabolism in, 34–39
organization of, 30–34
plasma membrane in, 27–29, 27*f*. *see also* Plasma membrane
Cell adaptation, 63–64, 63*f*, 64*b*
atrophy in, 63, 63*f*
dysplasia in, 63*f*, 64
hyperplasia in, 63, 63*f*
hypertrophy in, 63, 64*f*
metaplasia in, 63–64, 63*f*
Cell cycle, 54, 55*f*, 57*f*
cancer biology and, 54–55
Cell injury, 59–76
adaptive responses to, 63–64. *see also* Cell adaptation
etiology of, 68–73
chemical, 72, 72*t*
hypoxic, 68–70, 74*b*
immunologic, 70–72
infectious, 70–72
ischemic, 68–70, 70*f*
nutritional, 70, 71*t*
physical/mechanical, 72–73
reperfusion in, 69
irreversible, 64–68
reversible, 59–62
Cell-mediated immunity
cytotoxic T cells in, 181, 182*f*
mechanisms of, 178–181
T-helper cells in, 178–181, 181*f*
Cell memory, in tissue differentiation, 86
Cell signaling strategies, 48–49, 48*f*–49*f*
Cement burns, 1110
Central nervous system (CNS)
brain in, 853–861, 853*t*, 856*f*. *see also* Brain
spinal cord in, 861–863, 861*f*–863*f*. *see also* Spinal cord
structural organization of, 851–863
support and protection of, 851–853, 852*f*–855*f*, 852*t*
Central nervous system disorders. *see also* Encephalitis; Meningitis
infections as. *see also* Brain abscess
acute brain injury from, 911–913, 912*t*
Central venous pressure, 360
Centriacinar emphysema, 490, 491*f*
Centromere, 95, 95*f*
Cerebellar disorders, 924–925, 925*b*
Cerebellum, 860, 860*f*
in motor function, 885–886
Cerebral aneurysm, 909–910, 910*f*
Cerebral cortex, 853, 857*f*
Brodmann areas of, 854, 858*f*
Cerebral edema, 764, 896, 896*f*
Cerebral palsy, 922–923, 923*b*
Cerebrospinal fluid (CSF)
abnormal accumulation of, in hydrocephalus, 923–924, 923*f*
composition of, 851–852, 852*t*
production and circulation of, 852–853
Cerebrotendinous xanthomatosis, 779
Cerebrovascular accident (CVA), 905–906. *see also* Cerebrovascular disease and stroke
Cerebrovascular disease and stroke, 905–909
epidemiology of, 906
hemorrhagic stroke in, 906
ischemic stroke in, 906, 907*f*, 907*t*
stroke sequelae in, 908–909
treatment of, 906–908
Cerumen impaction, hearing loss from, 939
Cervical cancer, 679–680
screening guidelines for, 121*t*
Cervical dysplasia, in HIV infection, 252, 252*f*
Cervical plexus, 864–865, 865*f*–867*f*, 866*t*
Cervicitis, 690–691, 691*b*
Chancres, 692–693
Chancroid, 694, 694*f*
Channel proteins, membrane, 44, 44*f*
Chaperone proteins
in cell injury, 60, 62*f*
in translation, 82–83
CHD. *see* Coronary heart disease
Chemical injury, 1109–1110, 1110*b*
Chemicals
causing hypertension, 349*b*
nephrotoxic, 604*b*
teratogenic, 111–112
Chemokines, 169
Chemoreceptors, in control of respiration, 463, 464*f*
Chemotherapy
for cancer, 138
for hematologic neoplasms, 219
Cherry angioma, 1090, 1091*f*
Chest pain, 388, 721
Chest wall, deformities of, respiratory effects of, 511–513
Chewing, in nutrient ingestion, 707
Chickenpox, 1090, 1090*f*
Chigger, 1075, 1076*f*
Children
blood pressure classification in, 338*t*
burns in, physiologic changes in, 1095*t*, 1110, 1111*t*
cancer's warning signs in, 136*b*
diabetes mellitus in, 833–834, 834*b*
hypertension in, 345
joint disorders in, 1056–1057
respiratory system changes in, 459*b*
secondary hypertension in, 348*b*
skin disorders in, 1088–1090, 1088*f*
Chlamydia trachomatis, sexually transmitted infection from, 690
Chloride shift, 270
Cholangiopathy, progressive obliterative, 780
Cholangitis, primary sclerosing, 748, 773
Cholecystectomy, 748
Cholecystitis, 745–748
Cholecystoenteric fistula, 747
Cholecystokinin (CCK)
gastrointestinal motility and, 706
in schizophrenia, 978*t*
Cholelithiasis, 745–748, 747*f*
Cholestasis
extrahepatic, 780
intrahepatic, 780
Cholesterol, atherosclerosis risk and, 329
Cholesterol gallstone, formation of, 744–745, 747*f*
Cholinergic synapse, 919, 919*f*
Chondroma, 1036
Chondrosarcoma, 1037–1038
Chordae tendineae, 355, 357*f*
Chorionic villus sampling (CVS), in prenatal diagnosis, 114
Christmas factor, action of, 300*t*
Chromatids, 95, 95*f*, 98*f*
Chromatin, 78–79, 80*f*
Chromosome(s), 78, 95, 96*f*, 97*b*–98*b*
aberrant number of, 99–100
deletions in, 100, 101*f*
duplications in, 100, 101*f*
inversion in, 100, 101*f*
Philadelphia, in chronic myeloid leukemia, 221, 221*f*
rearrangement of, proto-oncogene activity and, 125, 126*f*
ring, 101*f*
structural abnormalities of, 100
translocations of, 100, 101*f*
Chromosome disorders, 99–102
autosomal, 100–101
cri du chat syndrome as, 101
trisomy 13 (Patau syndrome) as, 101
trisomy 18 (Edwards syndrome) as, 101
trisomy 21 (Down syndrome) as, 100–101, 102*f*, 102*t*
autosomal-dominant, 103–104, 105*t*
autosomal-recessive, 104–106, 106*t*
multifactorial, 110–111
polygenic, 110–111
sex, 101–102
Klinefelter syndrome as, 101, 103*f*
multiple X females and double Y males as, 102
Turner syndrome as, 102, 103*f*
Chronic active hepatitis, 772
Chronic bronchitis, 485–487, 486*f*–487*f*, 488*t*
Chronic cholecystitis, 748
Chronic cholelithiasis, 746–747, 747*f*
Chronic hyperglycemia, 826
Chronic ischemic cardiomyopathy, 393
Chronic kidney disease (CKD), 601–604, 604*b*
clinical management of, 604–605, 607*b*
dialysis in, 606
kidney transplant in, 606–607
in older adults, 607
complications of, 602–604
progression of, pathophysiology of, 602
risk factors for, 601–602, 601*b*
stages of, 602, 602*t*
target ranges for calcium and phosphorus in, 605*t*
therapeutic goals in, 605*b*
Chronic mucocutaneous candidiasis disease, 212
Chronic muscle pain, 1040
Chronic obstructive pulmonary disease (COPD)
type A, 488–489. *see also* Emphysema
type B, 486
Chronic open-angle glaucoma, 948–950, 950*f*
Chronic pancreatitis, 750–752, 752*f*
Chronic pericarditis, 403
Chronic persistent hepatitis, 772
Chronic tophaceous gout, 1055
Chronic venous insufficiency, 325, 333, 333*b*
Chronotropic effect, 372
Chylothorax, 510
Chyme, 708, 710*f*
Cilia, on airway epithelial cells, 452, 455*f*
Circadian rhythm(s)
of cortisol secretion, 793, 793*f*
depression related to, 980
in endocrine gland secretion, 790, 792*f*
test data and, 6, 6*f*
Circulation
cerebral, 853, 855*f*
portal, 757*f*
pulmonary, 457, 457*f*

Circulatory hypoxia, 468
Circulatory system, 314*f*–315*f*
  anatomy of, 356–357, 357*f*, 358*b*
  changes, in pregnancy, 667–668
  geriatric considerations on, 315, 317*b*
  hemodynamics of, 317–322
  organization of, 315–317, 317*b*
  primary functions of, 314
Cirrhosis, 773–774, 780*b*
  biliary, 773
  cryptogenic, 772
Cisternae, of Golgi apparatus, 32, 34*f*
Citric acid cycle, 36, 38*f*
Claudication, intermittent. *see* Intermittent claudication
Clear cell renal cell carcinoma, 580*f*
Climacteric, 668
Clinical manifestations, in pathophysiology, 3, 4*b*
Clitoris, 658, 660*f*
Cloning, DNA, 114–115
Closed fracture, 1027, 1028*f*
Clostridium, gas gangrene from, 65
Clot retraction, 299
  normal value and significance of, 305*t*
Clotting factors, 169, 298
  actions of, 300*t*
  activation of, intrinsic and extrinsic pathways of, 298
  coagulation disorders involving, 308–311
Cluster of differentiation (CD) markers, 166
CMV retinitis, in HIV infections, 253
Coagulation. *see also* Hemostasis
  common final pathway of, 301, 302*f*
  disorders of, 308–311, 311*b*
  disseminated intravascular, 310, 310*f*. *see also* Disseminated intravascular coagulation
  extrinsic pathway of, 301, 302*f*
  intrinsic pathway of, 299–300, 302*f*
Coagulation cascade, 299, 302*f*
Coagulation factors, 299, 301*f*. *see also* Clotting factors
Coagulative necrosis, 64, 66*f*
Coagulopathies, 308
Coarctation of aorta, 408, 408*f*
  hypertension in children from, 348–349
Cocci, 149, 151*f*
Coccygeal plexus, 864–865, 866*t*
Codominant expression, of traits, 96
Codons, RNA, for amino acids, 80, 82*t*
Cognitive behavioral therapy, for panic disorders, 991
Cognitive deficits, after stroke, 908–909, 909*b*
Cognitive functioning
  in depression, 979
  in schizophrenia, 976
Cold calorics test, 902
Cold injury, cellular, 72
Colitis, ulcerative, 729–730, 729*f*
  enteropathic arthritis secondary to, 1053
Collagen fibers
  in articular cartilage, 1011, 1012*f*
  in tendons, 1012–1013, 1012*f*
Collecting duct (tubule), of nephron, 555*f*–556*f*, 559
  epithelial cells of, 556*f*
  functions of, 556*t*
Colloids, for hypovolemic shock, 443
Colon
  diverticular disease of, 731–732, 732*b*, 732*f*
  functional anatomy of, 703, 703*f*–704*f*
  gastrointestinal motility of, 710–711
Colon cancer, 738–739, 739*f*, 739*t*
Colon polyps, 738, 738*f*
Colonic agenesis, 699
Colorectal cancer, 738–739, 739*t*
  screening guidelines for, 121*t*
Coma, in acute brain injury, 900, 901*f*
Comminuted fracture, 1026, 1026*f*
Common variable immunodeficiency disease (CVID), 212
Communication
  intercellular, 48–55. *see also* Intercellular communication
  neuronal, 872–880. *see also* Neuronal communication
  social, deficits in, in autism spectrum disorder, 997
Compact bone, 1003*f*
Compartment syndrome, 1024
  blood flow alterations from, 325
  complicating fractures, 1030
Compensated metabolic acidosis, 546
  laboratory values in, 547*t*
Compensated metabolic alkalosis, 548
Compensated respiratory acidosis, 547
  laboratory values in, 547*t*
Compensated respiratory alkalosis, 549
Compensation, 542
  renal, in acid-base balance regulation, 566
Compensatory pause, in premature ventricular complex, 427
Complement, 167–169, 168*f*
  membrane attack complex and, 168–169, 168*f*
  in type II hypersensitivity, 201
Complement cascade, 168, 168*f*
Complete fracture, 1027, 1027*f*
Compound fracture, 1027
Compression fracture, 1027, 1027*f*
Compulsion, in obsessive-compulsive disorder, 993
Computed tomography (CT)
  in cardiac function evaluation, 378
  renal, 573
Concentration, diminished, in major depression, 980–981
Concentric muscle contraction, 1017
Concussion, in traumatic brain injury, 903, 904*t*
Conduction aphasia, after stroke, 908
Conduction pathways disturbances, 428–431
  abnormal, 429, 429*f*
Conduction system, of heart, 371–372, 372*f*
Conductive hearing impairment, 939
Condyloid joint, 1010, 1011*f*
Condylomata acuminata, 694
Congenital adrenal hyperplasia, 795, 808, 808*f*
Congenital disorders, 111*t*
  environmentally induced, 111–113
    periods of fetal vulnerability for, 111, 112*f*, 113*b*
  of kidneys, 577–579, 579*b*
  of lower urinary tract, 615–617, 617*b*
  of penis and male urethra, 641–643
  in TORCH syndrome, 112, 113*f*
Congenital heart disease (CHD), 403–409, 409*b*
  acyanotic, 406–408
  cyanotic, 408–409
  etiology and incidence of, 404–405, 406*t*
  pathophysiology of, 405–406, 406*b*
Congestive cardiomyopathy, 400
Congestive heart failure (CHF), 411
Conn syndrome, 809–810
Connective tissue, 89–90, 89*f*–90*f*
  dense (supportive), 89, 89*f*
  hematopoietic, 89–90, 90*f*
  loose, 89, 89*f*
  types and locations of, 88*t*
Consciousness, 886–888
  level of
    in acute brain injury, 900, 900*t*, 901*f*
    Glasgow Coma Scale and, 900–901, 900*b*
Constipation, 723
Contact dermatitis, 1073–1074, 1074*f*
Contractile soft tissue injuries, 1023–1025, 1024*f*–1025*f*
Contractility, of heart, 375–376
  in shock, 447
Contraction
  muscle, 1015–1017. *see also* Muscle contraction
  myocardial, molecular basis of, 366–367
    calcium in, 367, 368*f*
    overview of, 366
    sliding filament/cross-bridge theory of, 366–367, 368*f*
Contralateral sensory transmission, 883
Convalescence, from disease, 3
Coombs test, for hemolytic disease of newborn, 284
Coping, adaptation and, 20–21
Copper deficiency syndromes, function and, 71*t*
Cor pulmonale, in right-sided heart failure, 418–419
Cornea, 942–943, 943*f*
Corneal reflex, in acute brain injury, 902, 902*b*
Corns, 1078
Coronary angiography, 379–380, 380*f*–381*f*
Coronary arteries, supplying heart, 360, 361*f*
Coronary artery disease. *see* Coronary heart disease
Coronary circulation, 360–362, 362*b*–363*b*
  regulation of flow in, 361–362
  vascular anatomy for, 360, 361*f*–362*f*, 362*t*
Coronary heart disease (CHD), 393*b*–394*b*
  acute coronary syndrome as, 388–393, 389*f*. *see also* Acute coronary syndrome
  angina pectoris in, 388. *see also* Angina pectoris
  atherosclerosis and, 327–329. *see also* Atherosclerosis
  chronic ischemic cardiomyopathy as, 393
  clinical features and management of, 388–393
  etiology of, 383
  family history of, atherosclerosis risk and, 329
  ischemia in, pathophysiology of, 385–387, 387*b*–388*b*
  myocardial infarction as, 388–389, 390*t*, 391*f*–393*f*. *see also* Myocardial infarction
  risk factors for, 383*b*
  risk of, recommended serum low-density lipoprotein levels to reduce, 383
  sudden cardiac arrest as, 393
Corpora cavernosa, 631, 631*f*
Corpus callosum, 856, 858*f*–859*f*
Corpus cavernosum, smooth muscle relaxation in, mechanism of, 646*f*

Corpus spongiosum, 631, 631*f*
Cortical bone, 1026
  healing in, 1028, 1029*f*
Corticospinal tract, 860, 885, 886*f*
Corticosteroids
  for autoimmune disorders, 198
  for type I hypersensitivity, 200–201
Cortisol
  biological actions of, 796, 796*b*
  in glucose metabolism, 843–844, 843*t*
  in lipid metabolism, 843*t*
  secretion of, diurnal rhythm of, 793, 793*f*
  in stress and adaptation, 18–19, 21
Cowper glands, 628*f*, 629
Cradle cap, 1087, 1088*f*
Cranial meninges, 851, 852*f*
Cranial nerve reflexes, in acute brain injury, 901–902, 902*f*
Cranial nerves, 860, 863–864
  origin and distribution of, 864*f*
  origins and functions of, 861*t*
Creams, in skin care, 1086
Creatine phosphate (CP), 369
Creatinine, serum, in renal function evaluation, 572
Creatinine clearance, in GFR estimation, 572
Crescentic glomerulonephritis, 589
Cri du chat syndrome, 101
Crigler-Najjar syndrome, 779
Crohn disease, 730–731, 730*f*
  enteropathic arthritis secondary to, 1053
Cross-bridge theory, of muscle contraction, 90, 92*f*, 366–367, 368*f*
Crossing over, during meiosis, 95, 97*f*
  chromosome structural abnormalities arising during, 100, 101*f*
Croup syndrome, 496
Crust, 1063*f*–1064*f*
Cryptocidins, 159
Cryptogenic cirrhosis, 772
Cryptorchidism, 647–648, 648*f*
Cryptosporidiosis, in HIV infection, 249
Crystalloids, for hypovolemic shock, 443
CSF-brain barrier, 853
Cultural considerations
  influencing normality, 5–6
  on mental health illness, 984
Cushing disease, 808
Cushing syndrome, 806*b*, 808–809, 809*f*–810*f*
CVA tenderness, in intrarenal disorders, 576
Cyanosis, 1082*b*
  in left-sided heart failure, 417–418
Cyanotic congenital defects, 408–409
Cyclic guanosine monophosphate (cGMP), 51, 53*f*
Cyclins, 54–55, 57*f*
Cyclosporine (Sandimmune), for autoimmune disorders, 198
Cyst(s)
  Bartholin, 677–678
  ovarian, 678–679
Cystic fibrosis (CF), 106, 108*f*, 494–495, 779
  clinical manifestations of, 495
  diagnosis of, 495
  etiology of, 494
  pathogenesis of, 494–495
  treatment of, 495
Cystic fibrosis transmembrane conductance regulator protein (CFTR), 106, 108*f*
Cystic kidney disease, 578–579, 578*t*
Cystitis, 619–621, 620*b*, 621*f*
  clinical manifestations of, 620
  diagnosis and treatment of, 620–621, 621*f*
  etiology and pathogenesis of, 620
  interstitial, 613–614
  risk factors for, 620, 620*b*
Cystocele, 675, 676*f*
Cystography, 611
Cystometry, 611
Cystourethrography, voiding, 611
Cytokines, 169, 170*t*
  in allergic asthma, 480
  immune
    in shock, 436, 437*t*
    in stress and adaptation, 19, 21
  in leukocyte differentiation, 162
  in mononuclear phagocyte system, 159–160
Cytolytic hypersensitivity, 199*t*. *see also* Type II hypersensitivity
Cytomegalovirus (CMV) retinitis, in HIV infection, 253
Cytoplasmic signaling pathways, in cell proliferation, 124–125, 125*f*
Cytosine nucleotides, methylation of, 86, 87*f*
Cytoskeleton, 30, 32*f*, 34*b*
  function of, 57*t*
Cytotoxic edema, 896
Cytotoxic hypersensitivity, 201. *see also* Type II hypersensitivity
Cytotoxic T cells, 167, 167*f*
  in cell-mediated immunity, 181, 182*f*
Cytotoxicity, cell-mediated, antibody-dependent, 166
Cytotoxins, for autoimmune disorders, 198

D

Data, reliability, validity, and predictive value of, 5, 6*b*
Dead space, in ventilation, 460
Death
  of cells, 64
  of organism, 75
  recurrent thoughts of, in major depression, 981
Decerebrate posturing, in acute brain injury, 901, 901*f*
Decorticate posturing, in acute brain injury, 901, 901*f*
Decubitus ulcers, 1078
  clinical description of, 1078*t*
  prevention of, 1078, 1078*b*
Deep partial-thickness burn, 1096*t*, 1097
Deep vein thrombosis, 333
  complicating fractures, 1030
Defecation, 710–711
Defensins, 159
Degranulation, of mast cells and basophils, 164
Dehydration
  clinical, 528, 528*f*, 529*b*
  in systemic disease, 1083
Dehydroepiandrosterone, in stress and adaptation, 19
Delayed fracture healing, 1029
Delayed hypersensitivity, 207
Deletions, chromosomal, 100, 101*f*
Delirium, in acute brain injury, 900
Delusions
  in bipolar disorder, 983
  in schizophrenia, 976
Dementia, 918–920
  clinical manifestations of, 920
  diagnosis and treatment of, 920, 920*b*
  etiology of, 918
  pathogenesis of, 918–920, 918*f*–919*f*
Dendrites, 91, 92*f*
Dendritic cells, 159, 165, 166*f*
Denver shunt, for ascites, 765
Deoxyribonucleic acid. *see* DNA
Depolarization, membrane, action potential triggered by, 45
Depressed fracture, 1027
Depression
  complicating chronic kidney disease, 604
    management of, 606
  major, 978–982, 979*t*. *see also* Major depressive disorders
  prevalence of, in women, 984, 985*t*
Dermal disorders, 1076–1081
  altered cell growth as, 1078–1080
  decubitus ulcers as, 1078. *see also* Decubitus ulcers
  pigmentation alterations as, 1080–1081
  scleroderma as, 1076–1077
  sunburn and photosensitivity as, 1077–1078
Dermatitis
  allergic contact, 207–209
  atopic, 1072–1073, 1073*f*
  contact, 1073–1074, 1074*f*
  seborrheic, 1069, 1070*f*
    in HIV infection, 1079, 1079*f*
Dermatomes, 576*f*, 864–865, 1104
  sensory, 957–958, 959*f*
Dermatomyositis, 1038
Dermatophyte infections, 1065, 1068*f*
Dermis
  age-related changes in, 1060–1061
  disorders of, 1076–1081. *see also* Dermal disorders
Desensitization, to stressor, 20–21
Desensitization therapy, for type I hypersensitivity, 201
Developmental disorders, 94–116, 111*b*
Developmental processes, 7
Diabetes insipidus, 568, 812–813
  clinical manifestations and diagnosis of, 813
  etiology and pathogenesis of, 812–813, 813*b*
  treatment of, 813
Diabetes mellitus, 815–837
  clinical manifestations of, 825–827, 827*b*
    acute hyperglycemia as, 825
    chronic hyperglycemia as, 826
    diabetic ketoacidosis as, 825–826, 826*t*
    nonketotic hyperglycemic hyperosmolar syndrome as, 826, 826*t*
  complications of, 825–827, 827*b*
    geriatric, 835
    nephropathy as, 826–827
    neuropathic, 827
    pediatric, 834
    pregnancy-related, 827
    retinopathy as, 826, 947–948
    vascular, 826–827
  diagnosis of, 822*b*
  eating disorders and, 828–829
  education about, 827–832, 828*f*
  excessive glycogen storage in, 60
  geriatric considerations in, 834–835, 835*b*
  gestational, 824, 824*t*
  obesity and, 828–829

Diabetes mellitus *(Continued)*
pediatric considerations in, 833–834, 834*b*
as risk factor for chronic kidney disease, 602
screening for, 824, 825*b*
treatment of, 827–832, 828*f*, 832*b*–833*b*
efficacy of, assessment of, 831–832
exercise in, 829
geriatric, 835
nutrition in, 828, 829*b*
pediatric, 834
pharmacologic, 829–831, 829*t*, 832*t*
incretin enhancers, incretins, and amylins in, 830, 832*t*
insulin in, 830–831, 832*t*
oral antidiabetic agents in, 829–830, 829*t*
stress management in, 831
type 1, 821–823, 822*f*
type 2, 823, 823*f*
types of, 824
Diabetic ketoacidosis, 825–826, 826*t*
laboratory values in, 547*t*
pediatric, 834
Diabetic neuropathy, 965–966
Diabetic retinopathy, 947–948
Dialysis, in chronic kidney disease management, 606
Diapedesis, 172, 173*f*
Diaper rash, 1087, 1088*f*
Diarrhea, 723–724
chronic, in HIV infection, 249
pathophysiologic mechanisms of, 723–724
Diarthroses, 1007–1010, 1009*f*
Diastole, 358, 358*f*
Diastolic blood pressure, 338, 360
Diastolic dysfunction, with low ejection fraction, 412, 413*f*
DIC. *see* Disseminated intravascular coagulation
Diencephalon, 856–860, 859*f*
Diet, for hypertension, 346–348
Diffuse axonal injury, 903
Diffuse esophageal spasm, 721
Diffuse interstitial lung disease, 500–501, 500*f*
Diffuse scleroderma, 1077, 1077*f*
Diffusion, facilitated, through membrane proteins, 41
Diffusion capacity test, 471–472
DiGeorge syndrome, 211
Digestion, 712–715, 714*f*, 716*b*–717*b*
Digitalis, for heart failure, 421
Dilated cardiomyopathy (DCM), 400, 400*f*, 401*t*
Dipeptidylpeptidase-4 (DPP-4), for diabetes, 830
Diploid number, of chromosomes, 95
Disease(s)
aggregate factors of, 7–9
clinical course of, 3
clinical manifestations of, 3
convalescence from, 3
endemic, definition of, 7
epidemic, definition of, 7
etiology of, 2, 3*b*
exacerbation of, 3
incubation period of, 3
latent period of, 3
pandemic, definition of, 7
patterns of, in populations, 6–10, 7*f*
prodromal period of, 3
remission of, 3
signs of, 3
silent period of, 3
stages of, 3
Disease(s) *(Continued)*
subclinical stage of, 3
symptom of, 3
treatment implications for, 3–4
viral, 154*t*
Dislocations, 1030
Disorganized behavior, in schizophrenia, 976
Disorganized thinking, in schizophrenia, 976
Displaced fracture, 1027
Disseminated intravascular coagulation (DIC), 310, 310*f*
complicating shock, 449
Distal convoluted tubule, of nephron
epithelial cells of, 556*f*
functions of, 556*t*
structure of, 555*f*, 559
Distress, 20
Distributive shock, 443–447, 447*b*
etiology of, 435*b*
Disuse atrophy, 63
Disuse osteoporosis, 1035
Diuretic agents, in renal tubular fluid reabsorption, 569–570, 569*t*
Diurnal variations, in test data, 6, 6*f*
Divalproex, for bipolar disorder, 984
Diverticular disease, 731–732, 732*b*, 732*f*
Diverticulitis, 731–732, 732*b*
Diverticulosis, 732*b*
Diverticulum, 732*b*
DNA (deoxyribonucleic acid), 30, 77–78
double helix model of, 79, 79*f*
mutation of, 96–97, 99*f*
repair of, 96–97, 99*f*
replication of, 79–80, 81*f*–82*f*
sense strand of, 81
structure of, 78–83, 78*f*–80*f*
transcription of, 81–82, 83*f*
controls on, in genome regulation, 84, 86*f*
DNA cloning, 114–115
DNA virus, 150–151, 154*t*
Doll's eyes test, 902, 902*f*
Dominant alleles, 95–96, 98*f*
Dominant trait, 95–96, 98*f*
Dopamine
as neurotransmitter, 876*b*, 876*t*, 878*f*
in renal tubular fluid reabsorption, 568*t*
in schizophrenia, 972–973, 973*f*, 978*t*
Dopaminergic neurons, degeneration of, in Parkinson disease, 921, 921*f*
Dorsal column-medial lemniscal tracts, 883, 884*f*
Double Y males, 102
Down-regulation, of hormone receptor, 787–788
Down syndrome (trisomy 21), 100–101, 102*f*, 102*t*
Dromotropic effect, 372
Drug(s)
associated with aplastic anemia, 277*t*
asthma induced by, 479–480
bacterial resistance to, mechanisms of, 145
in cancer chemotherapy, 138
in diabetes treatment, 829–831, 829*t*, 832*t*
hypertension caused by, 349*b*
in hypertension treatment, 348, 348*b*
reactions to, antibody-mediated, anemia in, 275*t*–276*t*, 284–285, 284*t*
teratogenic, 111–112, 113*t*
Drug eruptions, 1074, 1074*f*
Dry gangrene, 65
Dry mouth, in HIV infection, 249
Duchenne muscular dystrophy, 1039
Ductus deferens, 630–631
Dukes classification, for colorectal cancer, 739*t*
Dumping syndrome, 735–736, 736*f*
Duodenum, 701, 702*f*
Duplications, chromosomal, 100, 101*f*
Dura mater, 851, 852*f*
Dura mater injuries, 1023
Dysfunctional uterine bleeding, 673
Dyskinesia, in cerebral palsy, 922
Dyslipidemias, HIV-associated, 253
Dysmenorrhea, 673
Dysphagia, 721
causes of, 722*f*
Dysplasia, 63*f*, 64
Dyspnea, in left-sided heart failure, 417
Dysrhythmias, cardiac, 421–431. *see also* Cardiac dysrhythmias
Dysrhythmogenic right ventricular disorder, 401*t*

E

Ear(s)
bacteria targeting, 153*f*
external, 937, 937*f*
inner, 937, 937*f*–938*f*
middle, 937, 937*f*
infections of, 938
structure and function of, 937–938, 937*f*
Eating disorders, diabetes mellitus and, 828–829
Eccentric contraction, muscle, 1017
Ecchymosis, in coagulation disorders, 302–303, 303*f*
Echocardiography, 378, 379*f*
Ectoparasites, infections caused by, 155*t*–156*t*
Ectopic testes, 648*f*
Ectopic ureter, 616
Eczematous dermatitis, 1072–1073, 1073*f*
Edema, 1082*b*
causes of, 528–529, 530*f*
cytotoxic, 896
formation of, in soft tissue healing, 1024, 1024*f*
insulin, 831
pathophysiologic mechanisms of, 321
vasogenic, 896
Edwards syndrome (trisomy 18), 101
Effusion
pericardial, 402
pleural, 510–511
Ejaculation, 639
premature, 645, 647*t*
in sperm transport, 639
Elastin, in tendons and ligaments, 1013
Elastosis, 1060
Elderly
acid-base imbalance in, 550*b*
blood pressure classification in, 339*t*
burns in, physiologic changes in, 1095*t*
changes in the eyes of, 946*b*
chronic kidney disease management in, 607
diabetes mellitus in, 834–835, 835*b*
fluid and electrolyte homeostasis and imbalances in, variations in, 537*b*
gastrointestinal system of, changes in, 717*b*
infection risk in, 144
nervous system changes in, 881, 881*b*–882*b*
orthostatic hypotension in, 350
pain in, 967
physiologic changes in. *see* Aging, physiologic changes in
presbycusis in, 940, 941*b*
renal function changes in, 570, 571*b*

Elderly *(Continued)*
restrictive lung disease in, age-related features contributing to, 501*t*
secondary hypertension in, 348*b*
skin changes in, 1059, 1060*f*
Electrical activity, of gastrointestinal smooth muscle, 705–706, 706*f*
Electrical asystole, 424, 425*f*
Electrical burn injury, 1107–1108, 1107*f*, 1108*b*–1109*b*
cellular, 73
Electrocardiography (ECG), 372–374, 373*f*–374*f*, 374*b*, 376–378, 377*f*–379*f*
Electrochemical gradient, 41
Electroconvulsive therapy (ECT), for depression, 981–982
Electrolyte(s)
absorption of, 715, 716*f*
serum concentrations of, normal, 529*t*
Electrolyte concentration
calcium, imbalances in, 533–534, 534*b*
magnesium, imbalances in, 534–535, 535*b*, 535*f*
phosphate, imbalances in, 535–536, 536*b*
potassium, imbalances in, 531–533, 532*b*–533*b*
Electrolyte homeostasis, 531*f*
abnormal loss of electrolytes in, 531
disorders of. *see* Electrolyte imbalances
distribution in, 531
excretion in, 531
geriatric considerations in, 537*b*
intake and absorption in, 529–530
pediatric considerations in, 537*b*
principles of, 529–531, 531*b*
Electrolyte imbalances, 531–536, 536*b*
complicating chronic kidney disease, 603
management of, 605
geriatric considerations in, 537*b*
in plasma calcium, 533–534, 534*b*
in plasma magnesium, 534–535, 535*b*, 535*f*
in plasma phosphate, 535–536, 536*b*
in plasma potassium, 531–533, 532*b*–533*b*
signs and symptoms of, summary of, 538*t*
Electrolyte pools, 531
Electromagnetic radiation, cell injury from, 73, 73*f*
Electromechanical coupling, in muscle contraction, 1015–1016, 1016*f*
Electromyography, 1016
Electron transport chain, 36–38, 39*f*
Electrophoresis, in HIV infection diagnosis, 246
ELISA, in HIV infection diagnosis, 245
Embden-Meyerhof pathway, in erythrocyte maintenance, 267
Embolectomy, 325
Embolus, 324–325
clinical manifestations of, 324–325
etiology and pathogenesis of, 324
materials causing, 325
pulmonary, 473–475, 474*b*–475*b*, 474*t*
treatment of, 325
Embryonic development, 657
events in, 666*t*
fetal membranes and placenta in, 663–664
in first month, 664
of gastrointestinal system, 698–699
implantation in, 663, 665*f*
of male reproductive system, 632–633, 633*b*
of pancreaticobiliary system, 743, 744*f*
Embryonic development *(Continued)*
in second month, 665, 666*f*
in third month, 665, 666*f*
tissue differentiation in, 86–92, 92*b*
Emergency esophagogastroduodenoscopy (EGD), 761
Emesis, repeated, laboratory values in, 547*t*
Emission, 639
Emphysema, 488–491, 491*b*
centriacinar, 490, 491*f*
clinical manifestations of, 490, 492*f*
diagnosis of, 490–491
distinguishing features of, 488*t*
etiology of, 488–489
panacinar, 490, 491*f*
paraseptal, 490
pathogenesis of, 489–490, 489*f*–491*f*
treatment of, 491
Encephalitis, 912
Encephalopathy
hepatic, 763–764
portal systemic, 763–764
Enchondroma, 1036
End-diastolic volume (EDV), 360, 375
End-stage renal disease (ESRD), nephron loss in, 602*t*
End-systolic volume (ESV), 360
Endemic disease, definition of, 7
Endocardial disease, 394–399, 394*f*, 399*b*
infective endocarditis as, 398–399, 399*f*
rheumatic heart disease as, 398
Endocardium, 356, 357*f*
Endocrine disorders, 799–814
adrenal medulla, 810–811, 811*b*
adrenocortical, 806–810, 806*b*, 807*f*–810*f*, 810*b*
antidiuretic hormone, 812, 813*b*–814*b*, 813*f*
basic concepts of, 799–801, 801*b*
categories of, 797–798, 798*b*
classification of, 800–801, 801*f*
diabetes and, 815–837. *see also* Diabetes mellitus
etiology of, 799–800, 800*f*
growth hormone, 801–802, 801*b*–802*b*, 802*f*
parathyroid gland, 811–812, 811*f*, 812*b*
primary or secondary, 800, 801*f*
thyroid hormone, 803–805, 803*f*–805*f*, 804*b*–806*b*, 804*t*
Endocrine function, of heart, 376
Endocrine signaling, 49, 50*f*, 784, 784*f*
Endocrine system, 783–798. *see also* Hormone(s)
disorders of, 799–814. *see also* Endocrine disorders
hormone structure and action in, 784–786, 786*b*
hypothalamic-pituitary, 788–793, 789*f*
physiology of, 783–798
Endocytosis, 39–40, 40*f*–41*f*, 44*b*
Endometrial cancer, 680
screening guidelines for, 121*t*
Endometriosis, 679
Endometrium, 658
Endomysium, 1014, 1014*f*
Endoplasmic reticulum (ER), 31–32, 34*f*
function of, 57*t*
Endorphins
in pain modulation, 959
in stress and adaptation, 19
Endoscopic band ligation, of gastroesophageal varices, 761, 762*f*
Endoscopic sclerosis, of gastroesophageal varices, 761, 762*f*
Endospore, bacterial, 145
Endosteum, 1003, 1003*f*
Endothelium, 88
vascular, 323
Endotoxins, 70
bacterial, 144–145
in septic shock, 444
Energy, of muscle relaxation, 367, 367*b*–368*b*, 369*f*
Energy failure, cellular, in acute brain injury, 892–893, 893*f*
Energy loss, in major depression, 980–981
Energy metabolism
during exercise, 818, 820*f*
in fasting state, 817, 819*f*
in fed state, 816–817, 819*f*
pathophysiology of
in type 1 diabetes mellitus, 822–823, 822*f*
in type 2 diabetes mellitus, 823, 823*f*
during stress, 818–820
Enkephalins, in stress and adaptation, 19
Enteric infections, sexually transmitted, 695
Enterocolitis, 731–732
Enteropathic arthritis, 1053–1054
Enuresis, 613
Envelope glycoproteins, in HIV infection, 242
Environmental factors
in congenital disorders, 111
influencing respiration, 464
Enzyme deficiencies, multisystem, 779
Enzyme-linked immunosorbent assay (ELISA), in HIV infection diagnosis, 245–246
Enzyme-linked receptors, 49–51, 50*f*, 53*f*
Enzymes, bacterial, 145
Eosinophils, 163–164, 164*f*, 172
characteristics of, 263*t*
functions of, 163*t*
Ependymal cells, 868, 873*f*
Epicardium, 356, 357*f*
Epidemic disease, definition of, 7
Epidemiology, concepts of, 6–10, 7*f*, 10*b*
disease patterns as, 6–10
prevention levels as, 9–10
Epidermal proliferation, 1078–1080
Epidermis
age-related changes in, 1060, 1060*b*
epithelial cells in, 87, 88*f*
Epididymis, 630–631, 630*f*
Epididymitis, 650
Epidural hematoma, in traumatic brain injury, 903–904, 904*f*
Epigenetics, 77–78
Epigenomic characterization, 2
Epiglottitis, 496
Epilepsy, 916
Epileptogenic focus, 916
Epimysium, 1014, 1014*f*
Epinephrine
in glucose metabolism, 843
in stress and adaptation, 17–18
for type I hypersensitivity, 200–201
Epiphyseal injury, 1026, 1027*f*
Epiphyseal plate, 1003, 1004*f*
Epispadias, 643, 644*f*
Epistaxis, 303–304
Epithalamus, 860
Epithelial cells, airway, cilia on, 452, 455*f*

Epithelium, 88
  as barrier to infection, 159
  bronchial, 452, 454*f*
  shape and layering variations in, 88–89, 88*f*
  types and locations of, 88*f*, 88*t*
Equilibrium, 938
Erectile dysfunction (ED), 645, 646*f*
Erection
  painful, persistent, 643
  in sperm transport, 639
Ergot alkaloids, for migraine, 964*t*
Erosion, 1063*f*–1064*f*
Erythema, 1082*b*
Erythroblastosis, 284
Erythroblastosis fetalis, 203
Erythrocyte sedimentation rate (ESR), in inflammation, 175
Erythrocytes, 89–90, 90*f*, 262, 262*f*, 268*b*
  carbon dioxide transport by, 270, 271*f*
  characteristics of, 263*t*
  destruction of, 268, 269*f*
    extrinsic, anemia related to, 284–285
  development of, 264*f*, 265
  disorders of, laboratory findings in, 275*t*–276*t*
  hypochromic, microcytic, 279–280
  inherited disorders of, anemia in, 279–284
  macrocytic, 278
  maintenance of, energy and, 267
  mature, 262*f*
  oxygen transport by, 268–270
  production of, 267–268, 267*f*
    decreased, anemia related to, 277–279
  structure and function of, 262–268
Erythropoiesis
  nutritional requirements for, 266–267
  regulation of, 267–268, 268*f*
Erythropoietin, 267–268, 270, 272*f*, 570
Escape rhythms, 425, 425*f*
Esophageal cancer, 737
Esophageal disorders, 724–725, 726*b*
Esophageal varices, 725
Esophagogastroduodenoscopy, emergency, 761
Esophagus
  Barrett, 725
  pain in, 721
Estrogen(s)
  in stress and adaptation, 19
  target organs and actions of, 660*t*
Ethnicity
  atherosclerosis risk and, 329–330
  in epidemiology, 7
  hypertension and, 344–345
Etiology, in pathophysiology, 2, 2*f*, 4*b*
Euglobulin lysis time, normal value and significance of, 305*t*
Eukaryotic cells, organelles of, 30. *see also* Organelles
Eustachian tube, 453
Ewing sarcoma, 1037*f*, 1038
Exacerbation, 3
Excision and grafting, in burn wound management, 1104–1105
Excitatory amino acids, in acute brain injury, 893–894, 894*f*
Excitatory postsynaptic potential (EPSP), 874
Excitotoxin, glutamate as, 893, 894*f*
Excoriation, 1063*f*–1064*f*
Excretion
  electrolyte, in maintaining homeostasis, 531
  fluid, in maintaining homeostasis, 524
Exercise
  asthma induced by, 479
  in diabetes mellitus treatment, 829
  energy metabolism during, 818, 820*f*
  in hypertension management, 346–348
Exhaustion stage, of general adaptation syndrome, 14*t*, 16
Exocrine pancreas, alterations in function of, 742–753
Exocytosis, 39–40, 40*f*, 44*b*
Exons, 82
Exophthalmos, in Graves disease, 805, 805*f*
Exotoxins, 70
  bacterial, 145
Expiratory reserve volume, 460*t*
External ear, 937, 937*f*
Extracellular fluid volume (ECV), 525–526
  deficit in, 525–526, 525*b*, 525*f*
    laboratory values in, 547*t*
  excess in, 526, 526*b*, 526*f*
Extracellular fluids, 522, 522*f*
Extracellular matrix, in intercellular communication, 48, 49*f*
Extrahepatic cholestatic conditions, 780
Extrahepatic ductopenia, 780
Extrapyramidal disorders, 860
Extrapyramidal tracts, injury to, in cerebral palsy, 922–923
Exudates, inflammatory, 175
Eye(s)
  aqueous and vitreous humor in, 943*f*
  bacteria targeting, 153*f*
  disorders of, 944–950. *see also* Visual impairments
  HIV infection manifestations involving, 253
  structure of, 942–943, 943*f*

F

Facial nerve (CN VII), 861*t*, 863, 864*f*
Facilitated diffusion, through membrane proteins, 41
Factor VIII alterations, in hemostatic disorders, 305*t*
Factor IX alterations, in hemostatic disorders, 305*t*
Fallopian tubes, 656–657, 657*f*
False aneurysms, 330–331, 331*f*
Familial adenomatous polyposis, 738–739
Fas ligand, in cell-mediated immunity, 181
Fascia of Buck, 631, 631*f*
Fasciae injuries, 1023
Fasciculi, in skeletal muscle, 1014, 1014*f*
Fast sodium channels, 45, 47*f*
Fast twitch muscle, 1017
Fasting glucose tolerance, impaired, 820
Fasting state, energy metabolism in, 817, 819*f*
Fat(s). *see also* Lipid(s)
  cancer risk and, 120–123
Fat emboli, 325
Fat emboli syndrome, complicating fractures, 1030
Fat necrosis, 64–65, 66*f*
  in breast, 682–683
Fatigue, in major depression, 980–981
Fatty donor liver, gross appearance of, 62*f*
Fatty liver, alcoholic, 773
Fed state, energy metabolism in, 816–817, 819*f*
Feedback mechanism
  in growth hormone regulation, 791–792
  in hormone secretion regulation, 787
  in hypothalamic-pituitary-testicular axis, 637*f*
Feelings of worthlessness, in major depression, 980–981
Female genital and reproductive function, 656–670
  alterations in, 671–688
    benign growths and aberrant tissue as, 678–679, 679*b*
    in breast disorders, 682–686, 686*b*. *see also* Breast disorders
    cancer, 679–680, 680*b*–681*b*
    inflammation and infection of, 676–678, 678*b*
    menstrual disorders, 672–673, 674*b*
    pregnancy disorders, 681, 682*b*
    in uterine position and pelvic support, 674–676, 676*b*
  hormones of. *see* Hormone(s), female reproductive
  in menopause, 668–669, 669*b*
  menstrual cycle and, 659–662, 661*f*, 662*b*
  in pregnancy, 663–668, 668*b*. *see also* Pregnancy
  structures of, 656–658, 658*b*
    breast as, 662–663, 663*b*
    external genitalia as, 658, 660*f*
    organization of, 656–658, 657*f*
    ovaries as, 657
    oviducts as, 658
    uterus as, 658, 659*f*
    vagina as, 658
Fertilization, 663
Fetal alcohol syndrome (FAS), 112
Fetal development, 664–665
  events in, 666*t*
  membrane formation in, 663–664
  in second trimester, 665, 667*f*
  in third trimester, 665
Fetus, vulnerable periods of organ development in, 111, 112*f*, 113*b*
Fever, nutritional considerations in, 846–848
Fiber, dietary, cancer risk and, 120–123
Fibrillation
  atrial, 426, 426*f*
  ventricular, 427–428, 427*f*
Fibrin clot, 299–301
  components of, 299
Fibrin split products
  alteration in, in hemostatic disorders, 305*t*
  normal value and significance of, 305*t*
Fibrin-stabilizing factor, action of, 300*t*
Fibrinogen, 261
  action of, 300*t*
  normal value and significance of, 305*t*
Fibrinolysis, 301, 303*f*
Fibrinous exudate, 175
Fibroblasts, in healing process, 174–175
Fibrocystic breast disease, 683–684, 683*f*
Fibromas, 579*t*
Fibromyalgia syndrome (FMS), 963, 965*f*, 1040
Fibrotic interstitial lung diseases, 500–504
Fight-or-flight response, 14, 15*f*, 17
Filamentous bacteria, 149
Filaments, 363, 1014, 1015*f*
  contractile, characteristics of, 365–366, 366*f*
Filtration, in fluid distribution, 522*f*, 523
First-degree atrioventricular block, 428, 428*f*
First-degree burns, 1096, 1096*t*

Fissure, 1063*f*–1064*f*
Fistula(s)
arteriovenous, 325
cholecystoenteric, 747
tracheoesophageal, types of, 453*f*
urethrorectal and vesicourethral, 642
Flail chest, respiratory effects of, 513
Flank pain, in intrarenal disorders, 578
Fleas, 1075, 1075*f*
Flora, normal microbial, 151–152, 154*t*
pathogenic potential of, 142, 147*b*
Fluid distribution, 522*f*, 523–524
Fluid excess, in systemic diseases, 1083
Fluid imbalance(s), 525–529, 529*b*
of body fluid concentration, 526–528, 527*b*–528*b*, 527*f*–528*f*
of both volume and concentration, 528, 528*f*, 529*b*, 530*f*
complicating chronic kidney disease, management of, 605
of extracellular fluid volume, 525–526, 525*b*–526*b*, 525*f*–526*f*. *see also* Extracellular fluid volume
geriatric considerations in, 537*b*
of interstitial fluid volume, 528–529, 530*f*
pediatric considerations in, 537*b*
signs and symptoms of, summary of, 538*t*
Fluid resuscitation, in burn shock, 1100, 1100*b*
Fluoride deficiency syndromes, function and, 71*t*
Flutter, atrial, 426, 426*f*
Foams, in skin care, 1086
Focal segmental glomerulosclerosis (FSGS), 590–591
Folate, in erythropoiesis, 266–267
Folate deficiency, anemia from, 278–279
Follicle-stimulating hormone (FSH)
release of, 792
target organs and actions of, 660*t*
Follicles, ovarian, 658*f*
Food additives, asthma induced by, 480
Foods, causing hypertension, 349*b*
Foot drop, after stroke, 908
Force-time relationship, in muscle contraction, 1017
Forced expiratory flow rate, 460*t*
Forced expiratory volume in 1 second (FEV_1), 471
Forced vital capacity (FVC), 471
Forehead ulcer, 1079, 1079*f*
Foreign body occlusion, hearing loss from, 939
Foreskin, 631
Fournier gangrene, 650
Fourth-degree burns, 1096*t*, 1097
Fracture, bone, 1026–1030. *see also* Bone fracture(s)
Fragile X syndrome, 109
Frameshift mutation, 97, 99*f*
Frank-Starling law of the heart, 375
Free radical theory, of aging, 75
Free radicals. *see* Oxygen radicals (free radicals)
"Frozen shoulder", 1022–1023
Full-thickness burns, 1096–1097, 1096*t*
Functional residual capacity, 460*t*, 461
Functional syncytium, of cardiac myocytes, 363, 364*f*
Fungal infections, 1065–1067
Fungal pneumonia, 514, 515*t*
Fungi
morphology of, 151*f*
pathogenic, 151–152, 154*t*
Fusiform aneurysms, 330–331, 331*f*
Fusion inhibitors, for HIV/AIDS, 257

G

G-protein-coupled receptors, 51, 52*f*, 784–785, 785*f*
G-protein-coupled signaling systems, 51, 52*f*
GABA, as neurotransmitter, 876*t*, 879
Gag reflex, 708
Gallbladder
alterations in function of, 742–753
structure and function of, 742–743
Gallbladder disorders, 744–748, 748*b*
cholecystitis as, 745–748
cholelithiasis as, 745–748, 747*f*
Gallstone, formation of, 744–745, 747*f*
Gangrene, 65
Fournier, 650
Gap junctions, in intercellular communication, 48, 48*f*
Gas(es). *see also* Carbon dioxide *entries;* Oxygen *entries*
blood. *see* Blood gases
intestinal, 723
Gas diffusion, in respiratory function, barriers to, 466–467, 466*f*
Gas exchange
alterations in, 451–477
alveolar units in, 454–456, 455*f*
ineffective
from true pulmonary shunt, 469, 470*f*
from ventilation-perfusion mismatch, 468–469, 470*f*
from ventilatory failure, 468
pulmonary system in, 452. *see also* Pulmonary system
Gas gangrene, 65
Gas transport
acid-base balance and, 268–270
in respiratory function, 466–468, 466*f*, 468*b*
Gas values, significant to blood oxygenation, 273*t*
Gastric carcinoma, 737, 738*f*
Gastric emptying, 708
Gastric filling and storage, 708
Gastric glands, 700–701, 702*f*
Gastric mucosa, 700–701, 702*f*
Gastrin, 712*t*
gastrointestinal motility and, 706
Gastritis, 726
Gastroenteritis, 726
Gastroesophageal reflux disease (GERD), 724–725
Gastroesophageal varices, 760–763, 761*f*
Gastrointestinal disorders, 720–741
esophageal, 724–725, 726*b*
manifestations of, 721–724, 724*b*
motility, 733–737, 735*b*
neoplasms as, 737–740, 739*b*
oral, 724, 726*b*
psychosocial aspects of, 739–740
Gastrointestinal function, 697–719, 718*f*
across life span, 717–718, 718*b*
digestion and absorption as, 712–715, 714*f*, 716*b*–717*b*
motility, 704–711, 711*b*. *see also* Gastrointestinal motility
secretory, 712, 712*b*
Gastrointestinal juices, secretion of, 712
Gastrointestinal motility, 704–711, 711*b*
of colon, 710–711
disorders of, 733–737, 735*b*
hormonal control of, 706
of ileocecal sphincter, 710
intestinal wall characteristics and, 704, 705*f*
mixing movements in, 707
motor functions of stomach in, 708–709
movement of nutrients in, 707–711
neural control of, 704–706
propulsive movements in, 706–707, 707*f*
of small intestine, 709–710, 710*f*–711*f*
smooth muscle contraction in, 706
Gastrointestinal system, 698*f*
disorders of, 720–741. *see also* Gastrointestinal disorders
of elderly, changes in, 717*b*
embryology of, 698–699
functional anatomy of, 699–703
esophagus in, 700
large intestine in, 703, 703*f*–704*f*
oral cavity and pharynx in, 699, 699*f*
small intestine in, 701–703, 702*f*–703*f*
stomach in, 700–701, 701*f*–702*f*
HIV infection manifestations in, 248–254
HIV replication in, 242–243
of infant, changes in, 716*b*
structure and organization of, 698–703, 704*b*
Gastrointestinal tract wall, alterations in, 726–732, 732*b*
Gate control theory, in pain modulation, 958
Gels, in skin care, 1086
Gender, in epidemiology, 7–8
Gender differences
in normality, 6
in posttraumatic stress disorder, 994
Gene(s), 77–78
cancer-critical, 123–124
cellular, altered expression of, in cancer, 117–118
tumor suppressor, 123–124, 127–129, 127*t*, 128*f*–130*f*
Gene amplification, proto-oncogene activity and, 125, 126*f*
Gene defects, associated with congenital heart disease, 404–405, 406*t*
Gene therapy, 113–115
for cancer, 138–139
General adaptation syndrome (GAS), 14–16
alarm stage of, 14, 14*t*, 15*f*
allostasis and, 14–16
exhaustion stage of, 14*t*, 16
resistance stage of, 14–16, 14*t*
General systemic disease, cardiomyopathies and, 401*t*
Generalized anxiety disorder (GAD), 991–992
Generalized seizures, 916–917, 916*b*
Genetic analysis, gene therapy and, 114
Genetic code, 80–81, 82*t*, 83*b*
Genetic disorders
chromosomal, 99–102. *see also* Chromosome(s); Chromosome disorders
developmental disorders and, 94–116, 111*b*
inheritance principles and, 95–97
Mendelian single-gene disorders as, 98–99, 102–107
nonmendelian single-gene disorders as, 107–110
prenatal diagnosis and counseling on, 113–114, 115*b*

Genetic engineering, 115
Genetic factors
in autoimmunity, 196, 197*t*
in type I hypersensitivity, 199
Genetic lipoprotein disorders, 385*t*
Genetic manipulation, 2
Genital duct system, development of, 633
Genital herpes, 692–693, 693*f*
Genital warts, 694
Genitalia, external
development of, 633, 636*f*
female, 658, 660*f*
male, 629–631, 633, 636*f*
Genome, 77–78
regulation of, 84
Genomic characterization, 2
Genomic imprinting, 110, 110*f*
Genotype, definition of, 95
Genotypic resistance testing, in monitoring HIV disease status, 246
Genu valgus or varus, in osteoarthritis, 1043, 1045*f*
Geographic location, in epidemiology, 9, 9*f*
Gerontology, 7
Gestational diabetes mellitus, 824, 824*t*
GFR. *see* Glomerular filtration rate
Ghrelin, gastrointestinal motility and, 706
Giant cell tumor, 1037
Gilbert syndrome, 779
Gingivostomatitis, herpetic, 724
Ginglymus joint, 1010, 1010*f*
Glands, age-related changes in, 1061, 1061*b*, 1062*f*
Glanzmann disease, 307
Glasgow Coma Scale, 900–901, 900*b*
Glaucoma, 948–950
Glial cells
astrocytes as, 853
in nervous system, 868–872
Globulin, serum, 261
Glomerular capillaries
hydrostatic pressure in, 561, 562*f*
oncotic pressure in, 561, 562*f*
Glomerular disorders. *see* Glomerulopathies
Glomerular filtration
factors affecting, 562–563
physics of, 560–561
regulation of, 560–564, 562*f*, 564*b*
glucose and amino acids in, 563–564
mesangial cells in, 564
tubuloglomerular feedback in, 563, 563*f*–564*f*
Glomerular filtration rate (GFR), 557
in chronic kidney disease, 602*t*
measures of, 572
Glomerulonephritis, 587–589
acute, 588, 588*f*
chronic, 589
crescentic, 589
immune complex, 206
rapidly progressive, 589
Glomerulopathies, 586–591, 586*b*, 591*b*
focal segmental glomerulosclerosis as, 590–591
glomerulonephritis as, 587–589. *see also* Glomerulonephritis
membranous nephropathy as, 590
minimal change disease as, 591
nephrotic syndrome as, 589–591
primary, 586, 586*b*
secondary, 586
Glomerulus, 559*f*
anatomy of, 587*f*
arterioles of, 553, 557*f*
functions of, 556*t*
membrane structure of, 557, 558*f*
structure of, 555–557, 555*f*
Glossopharyngeal nerve (CN IX), 861*t*, 863, 864*f*
Glucagon
in glucose metabolism, 843, 843*t*
in lipid metabolism, 843*t*
Glucocorticoids, 796. *see also* Cortisol
in stress and adaptation, 18, 18*t*
Gluconeogenesis, 36, 839–840
Glucose
in glomerular filtration regulation, 563–564
reabsorption of, 565, 566*f*
renal threshold for, 565
Glucose-6-phosphate dehydrogenase (G6PD) deficiency, 282–284
laboratory findings in, 275*t*–276*t*
Glucose intolerance disorders, 820–824, 824*b*–825*b*
atherosclerosis risk and, 329
classification of, 820, 821*b*
diabetes mellitus as, 815–837. *see also* Diabetes mellitus
pre-diabetes as, 820, 822*b*
Glucose metabolism
disorders of, 820–824. *see also* Glucose intolerance disorders
increased, in obsessive-compulsive disorder and, 992, 993*f*
regulation of, 816–820, 820*b*
exercise in, 818, 820*f*
hormonal, 816–817, 816*f*–819*f*
neural, 817–818
stress in, 818–820
Glucose tolerance, impaired, 820. *see also* Glucose intolerance disorders
Glutamate
excitotoxicity of, 893, 894*f*
as neurotransmitter, 876*t*, 877–879, 879*f*
in schizophrenia, 978*t*
Glycine, as neurotransmitter, 879
Glycogen storage, excessive, in diabetes, 60
Glycolipids, of plasma membrane, 28, 29*f*
Glycolysis, 36, 37*f*
Glycoproteins, envelope, in HIV infection, 242
Glycosaminoglycans, 87
intracellular accumulations of, in cell injury, 60
Glycosuria, 572
Golgi apparatus, 32, 34*f*
function of, 57*t*
Gonadotropins, 792
Gonads, development of, 633
Gonorrhea, 690
Goodpasture syndrome, 589
Gout, 1054–1055, 1055*f*
Gouty arthritis, 1054–1055, 1055*f*
Graft rejection, hyperacute, 204
Graft-*versus*-host disease, complicating bone marrow transplantation, 219
Grafting, in burn wound management, 1104–1105
Grand mal seizures, 916–917
Grandiosity, in bipolar disorder, 983
Granulocytes, 162
Granulocytopenia, in anemia, 277
Granuloma inguinale, 694
Granulomas, 174
Granulomatous hypersensitivity, 207
Graves disease, 204, 805, 805*f*
Great arteries, transposition of, 408–409, 409*f*
Greenstick fracture, 1026, 1026*f*
Growth factor receptors, in cell proliferation, 124, 125*f*
Growth factors, mitogen, in cell proliferation, 54, 57*f*, 124
Growth hormone (GH)
actions of, 790–792, 792*f*
in glucose metabolism, 843, 843*t*
in stress and adaptation, 19
Growth hormone (GH) deficiency, 801–802, 801*b*
Growth hormone (GH) disorders, 801–802, 801*b*–802*b*, 802*f*
Growth hormone (GH) excess, 801*b*, 802, 802*f*
Guanylin, in renal tubular fluid reabsorption, 568–569, 568*t*
Guillain-Barré syndrome, 931–932, 932*b*
respiratory effects of, 511, 512*t*
Guilt, inappropriate, in major depression, 980–981
Gustatory cortex, 952
Gut-associated lymphoid tissue (GALT), 162
Gynecologic manifestations, of HIV infection, 251–252

H

Habituation, to stressor, 20–21
Hageman factor, action of, 300*t*
Hairs
age-related changes in, 1060*b*, 1061
manifestations of systemic diseases in, 1084
Hairy cell leukemia, 224, 225*f*
Haldane effect, 270
Half-life, of hormone activity, 787
Hallucinations
in bipolar disorder, 983
in schizophrenia, 976
Haploid number, of chromosomes, 95
Hapten, in contact hypersensitivity, 208
Hapten mechanisms, of drug-induced immune hemolysis, 284
Hashimoto thyroiditis, 803–805
Haustral churning, in colon, 710
Haversian system, 1002–1003, 1002*f*
Headache, 962, 963*b*, 964*t*
Healing process
for burn wounds, 1106–1107, 1106*f*
in fracture, 1028–1029, 1029*f*
in soft tissue after trauma, 1024–1025, 1024*f*–1025*f*
Health, definition of, 9
Hearing, 937–942. *see also* Ear(s)
age-related changes in, 939, 941*b*
Hearing impairment
conductive, 939
general manifestations of, 938–939
interventions for, 942, 942*b*
sensorineural, 939–940
Heart
anatomy of, 355–356, 355*f*–357*f*, 356*t*
bacteria targeting, 153*f*
cardiac cycle and, 358–360, 358*f*, 360*b*. *see also* Cardiac cycle
chambers of, 356
normal pressures in, 356*t*
contractile capabilities of, 375–376

Heart *(Continued)*
coronary circulation in, 360–362, 362*b*–363*b*
regulation of flow in, 361–362
vascular anatomy for, 360, 361*f*–362*f*, 362*t*
electrophysiology of, 369–372, 372*b*
embryologic development of, 403–404, 403*f*–404*f*
endocrine function of, 376
Frank-Starling law of, 375
geriatric considerations on, 359*b*
layers of, 356, 357*f*
muscles of, 355–356. *see also* Cardiac myocyte(s); Myocardium
pressure work of, 376
pressures in, normal, 356*t*
specialized conduction system of, 371–372, 372*f*
valves of, 355–356, 356*f*
volume of blood in, 375, 375*f*
Heart diseases
congenital, 403–409, 409*b*
acyanotic, 406–408
cyanotic, 408–409
etiology and incidence of, 404–405, 406*t*
pathophysiology of, 405–406, 406*b*
coronary, 383–393. *see also* Coronary heart disease
endocardial, 394–399, 394*f*, 399*b*. *see also* Endocardial disease
myocardial, 399–401, 401*b*–402*b*. *see also* Cardiomyopathy(ies); Myocardial infarction; Myocarditis
pericardial, 402–403, 403*b*
rheumatic, 398
valvular, 394–399, 394*f*, 399*b*. *see also* Cardiac valve disorders
Heart failure, 412–421, 420*b*
backward, 417
biventricular, 420
class and stage of, 420, 420*t*
clinical manifestations of, 417–420, 418*f*, 420*t*
in left-sided heart failure, 417–418, 419*f*
in right-sided heart failure, 418–420, 419*f*–420*f*
compensatory mechanisms, remodeling, and progression of, 413–417, 414*f*, 417*b*
myocardial remodeling and progression in, 415–417, 415*f*–416*f*, 416*b*
preload increase as, 413–415, 415*f*
sympathetic nervous system activation in, 413, 414*f*
congestive, 411
definitions of, 411
diastolic dysfunction in, 412, 413*f*
forward, 417
in infants, 406*b*
pathogenesis and diagnosis of, 412, 412*b*–413*b*
systolic dysfunction in, 412
treatment of, 421, 421*b*
Heart rate, determinants of, 374–375
Heartburn, 700, 721
Heat injury, cellular, 72
Heat-shock proteins, in cell injury, 60, 62*f*
Heberden nodes, in osteoarthritis, 1043, 1044*f*
Helicobacter pylori
in gastritis, 726
in peptic ulcer disease formation, 727–728, 729*f*
Helminthic infections, 155*t*–156*t*
Hemangiomas, 579*t*
Hemarthrosis, in coagulation disorders, 302–303
Hematemesis, 303–304
Hematochezia, 303–304
Hematocrit, 319
Hematogenous osteomyelitis, 1031
Hematologic neoplasms
complications of, prevention and management of, 219–220, 220*b*
etiology of, 216–217
lymphoid, 216–217, 217*b*, 218*f*. *see also* Lymphoid neoplasms
manifestations of, 218*b*
myeloid, 216–217, 216*b*, 218*f*. *see also* Myeloid neoplasms
treatment principles for, 218–219
Hematologic values, age-related changes in, 262*t*
Hematoma(s)
in coagulation disorders, 302–303
formation of, in fracture healing, 1006, 1007*f*, 1028–1029, 1029*f*
intracranial, in traumatic brain injury, 903–905, 904*f*
Hematopoiesis, 263–265, 265*f*
bone marrow in, 160–161, 160*f*
definition of, 263
Hematuria, 303–304
Hemianopsia, 950
homonymous, after stroke, 908, 909*f*
Hemochromatosis, hereditary, 774
Hemodialysis, in chronic kidney disease management, 606
Hemodynamic monitoring, in shock, 448–449, 448*f*, 448*t*
Hemodynamics, circulatory, 317–322
Hemoglobin
molecular structure of, 265–266, 266*f*
in oxygen transport, 467, 469*f*
synthesis of, 265–266, 266*f*
in infants, 266, 267*f*, 294*b*
Hemolytic anemia, 281
Hemolytic disease of newborn, 203–204, 284
laboratory findings in, 275*t*–276*t*
Hemophilia, 308–309
Hemophilia A, 107, 109*f*
Hemophilic arthropathy, 1054
Hemoptysis, 303–304
Hemorrhage(s)
intracerebral, 906
splinter, in systemic disease, 1084, 1086*f*
subarachnoid, in traumatic brain injury, 904*f*, 905
Hemorrhagic exudate, 175
Hemorrhagic shock. *see also* Hypovolemic shock
classification of, 442, 442*t*
Hemorrhagic stroke, 906
Hemosiderin, intracellular accumulations of, 60–61
Hemostasis, 298–312, 301*b*
definition of, 298
disorders of
in hepatic disease, 310–311
laboratory value alterations in, 305*t*
platelet, 306–308, 306*b*
vascular, 304–306
evaluation of, 301–304, 304*b*
clinical assessment in, 301–304, 301*t*
laboratory tests in, 304, 305*t*
fibrin clot formation in, 299–301
fibrinolysis in, 301, 303*f*
Hemostasis *(Continued)*
platelets in, 299, 299*f*–300*f*
process of, 298–301
stages of, 298–299
Hemothorax, 510
Henderson-Hasselbalch equation, 542
Hepatic encephalopathy, 763–764
Hepatitis, 767–771, 772*b*–773*b*
alcoholic, 773–774
chronic, 772
ischemic, 780–781
viruses causing, comparison of, 769*t*
Hepatitis A virus infection, 767, 769*t*
in children, 778
diagnosis, treatment, and prevention of, 767
immunologic markers in, 768*t*
pathogenesis and clinical manifestations of, 767, 770*f*
Hepatitis B immunoglobulin (HBIG), 769–770
Hepatitis B virus infection, 767–771, 769*t*
in children, 778
diagnosis of, 768–769
in elderly, 781
immunologic markers in, 768*t*
pathogenesis and clinical manifestations of, 767–768
prevention of, 769–771
treatment of, 769
Hepatitis C virus infection, 769*t*, 771
in children, 778
in elderly, 781
immunologic markers in, 768*t*
pathogenesis and clinical manifestations of, 771
treatment of, 771
Hepatitis D virus infection, 769*t*, 771
immunologic markers in, 768*t*
pathogenesis and clinical manifestations of, 771
treatment and control of, 771
Hepatitis E virus infection, 769*t*, 771
in children, 779
immunologic markers in, 768*t*
pathogenesis and clinical manifestations of, 771
treatment of, 771
Hepatitis G virus infection, immunologic markers in, 768*t*
Hepatocellular carcinoma (HCC), 776
Hepatocellular failure, 755–759. *see also* Liver disorders
Hepatolenticular degeneration, 774–775
Hepatorenal syndrome, 766, 766*b*–767*b*
Herd immunity, 144
Hereditary hemochromatosis, 774
Hereditary hemorrhagic telangiectasia, 304*f*, 306
Hereditary nonpolyposis colorectal cancer syndromes, 738–739
Hereditary spherocytosis, 282
laboratory findings in, 275*t*–276*t*
Hering-Breuer reflex, 464
Hernia, hiatal, 725
in elderly, 717–718
Herniation, brain, in increased intracranial pressure, 897–899, 898*f*–899*f*
Herpes simplex virus (HSV), 1064–1065, 1066*f*
Herpes simplex virus encephalitis, 912
Herpes simplex virus skin/mucous membrane infections, 250, 250*f*
sexually transmitted, 692–693, 693*f*
Herpes zoster, 1065, 1067*f*
in HIV infection, 250, 250*f*
Heterochromatin, 86

Hiatal hernia, 725
in elderly, 717–718
Hilum, renal, 552
Hinge joints, 1010, 1010*f*
Hippocampus
pyramidal cells of, in schizophrenia, 974
reduced volume of, in posttraumatic stress disorder, 994
Hirschsprung disease, 734
Hirsutism, in systemic diseases, 1084
Histamine
in inflammatory response, 171
in type I hypersensitivity, 199
Histiocytes, 159
Histones, 78–79, 80*f*
Histotoxic hypoxia, 468
HIV-1 and HIV-2
binding of, HIV infection and, 239–242, 240*f*
differences between, 234
envelope of, 239, 239*f*
genome of, 239
replication of, in host cells, 242–243
structure of, 239, 239*f*
transmission of, 236–238
prevention of, 238
types of, 234–236
HIV/AIDS, 234–238, 238*b*–239*b*, 242*b*, 254*b*, 257*b*
global impact of, 234, 235*t*
HIV encephalopathy, 244
HIV infection, 233–258. *see also* AIDS (acquired immunodeficiency syndrome); HIV-1 and HIV-2; HIV/AIDS
classification system for
CDC, 245
pediatric, 245, 245*t*
clinical latency period in, 244
clinical manifestations of, 248–254
cardiovascular, 253
in children, 254
gastrointestinal, 249
gynecologic, 251–252, 252*f*
mucocutaneous, 250–251, 250*f*–252*f*
neurologic, 252–253
ocular, 253
pulmonary, 249–250, 250*f*
systemic, 248–249
diagnostic testing for, 245–246
epidemiology of, 234–238
etiology of, 239–242
HIV binding and, 239–242, 240*f*
pathogenesis of, 242–245
effect on immune cells at cellular level in, 242
progression of, to seroconversion to AIDS, 236*f*, 244–245, 244*f*
viral production and cell death in, 242–244, 242*f*–243*f*
progression of, monitoring of, 246–248, 247*b*–248*b*
seroconversion in, 244
HIV RNA, plasma level of, in monitoring HIV disease status, 246
Hodgkin disease, 227–229
clinical manifestations of, 227, 227*f*–228*f*
pathogenesis of, 227, 227*f*–228*f*
prognosis and treatment of, 227–229, 228*b*, 228*f*
staging of, 227–229, 228*b*, 228*f*
Hodgkin lymphoma, 227–229
Homeostasis, 12–13, 13*b*
acid-base, 541–543. *see also* Acid-base homeostasis
adaptive responses to stress and, 12–25. *see also* Adaptive response(s)
calcium, 1004
electrolyte, 531*f*. *see also* Electrolyte homeostasis
fluid and electrolyte, 521–540. *see also* Body fluid, homeostasis of
fluid volume, in blood pressure regulation, 341–342
Homeostatic systems, examples of, 13*b*
Homologous chromosomes, 95
Homonymous hemianopsia, after stroke, 908, 909*f*
Hormone(s)
action of, 799
mechanisms of, 784–786
activity of, amplification of, 786
classes of, chemical structure of, 784, 784*b*
counterregulatory, 817
female reproductive, 660*t*
changes in, in parturition, 665–667
gastrointestinal
motility and, 706
secretion of, 712, 712*t*, 713*f*
hypersecretion of, 797, 799–800, 800*f*
hyporesponsiveness to, 797–798
hyposecretion of, 797, 799–800, 800*f*
hypothalamic-pituitary, 788–793. *see also* Pituitary gland
mechanical, changes in, in parturition, 667
metabolism and excretion of, 787
parathyroid, regulation and actions of, 811–812, 811*f*
pharmacologic concentrations of, 787
receptor responses to, regulation of, 787–788
regulating sodium chloride and water reabsorption, 568*t*
regulation of, 786–788, 788*b*
reproductive, mental illness vulnerability and, 984
steroid, 794–796. *see also* Steroid hormones
structure and action of, 784–786, 784*f*, 786*b*
synthesis and secretion of, 786–787
thyroid, 793–794. *see also* Thyroid hormone(s)
Hormone resistance, 797–798
Host
characteristics of, infection risk and, 142–144, 143*t*
microorganisms and, relationship between, 142–145, 142*t*
Host-microbe relationship, for infectious processes in, 142–145, 142*t*
Human embryo, development of, 664–665
Human immunodeficiency virus (HIV), 233–258. *see also* AIDS (acquired immunodeficiency syndrome); HIV-1 and HIV-2; HIV/AIDS; HIV infection
Human leukocyte antigen (HLA) complex. *see* Major histocompatibility complex
Human microbiome, 142
Human papilloma virus infections, 694–695, 695*f*
Humoral immune system, dysfunction of, in HIV infection, 242
Humoral immunity
antibody functions in, 186–187, 187*f*
antibody structure in, 183–187
B-cell antigen recognition in, 181–183, 182*f*
Humoral immunity *(Continued)*
class switching and affinity maturation in, 185–186, 186*f*
mechanisms of, 181–183
Huntington disease, 104
Hyaline cartilage, 1003
Hydrocele, 648, 648*f*
Hydrocephalus, 852–853, 923–924
etiology of, 923–924, 923*f*
increased intracranial pressure from, 896, 897*f*
pathogenesis and clinical manifestations of, 924, 924*f*
treatment of, 924, 924*b*, 925*f*
Hydrofluoric acid, for chemical injury, 1109
Hydrolytic degradation, in cellular atrophy, 63
Hydropic swelling, in reversible cell injury, 60, 60*f*, 62*b*
Hydrostatic pressure, 321
in glomerular filtration, 561, 562*f*
increased, edema from, 528–529, 530*f*
Hyperacute graft rejection, 204
Hyperaldosteronism, 809–810
primary, hypertension in, 349
Hypercalcemia, 533–534, 534*b*
Hypercarbia, definition of, 452
Hypercortisolism, 806*b*, 808–809, 809*f*–810*f*
Hyperemesis gravidarum, 681
Hyperemia, 323
Hyperglycemia, in diabetes mellitus
acute, 825
chronic, 826
geriatric, 835
Hyperkalemia, 532–533, 533*b*
complicating chronic kidney disease, 603
Hyperkeratosis, 1078
Hyperlipidemia, atherosclerosis risk and, 384
Hypermagnesemia, 534–535, 535*b*, 535*f*
Hypermetabolic state
in burn, 1102
in systemic diseases, 1084
Hypermobility of joints, 1056
Hypernatremia, 527–528, 528*b*, 528*f*
Hyperopia, 944–945, 945*f*
Hyperparathyroidism, 811–812
secondary, complicating chronic kidney disease, 603
Hyperphosphatemia, 536, 536*b*
complicating chronic kidney disease, management of, 605
Hyperplasia, 63, 63*f*
Hypersecretion, of hormones, 797, 799–800, 800*f*
Hypersensitivity, 195, 198–209
delayed, 207–209
granulomatous, 207
mediators of, 164
tuberculin-type, 207
type I, 199–201, 199*t*, 200*f*
type II, 199*t*, 201–204, 202*f*
type III, 204–207, 205*f*, 205*t*
type IV, 207–209
types of, 198, 199*t*
Hypersensitivity pneumonitis, 502, 502*t*
Hypersomnia, in major depression, 980–981
Hypertension, 343–350, 350*b*
atherosclerosis risk and, 329
in childhood and adolescent, 345
classification of, 343–344
complicating chronic kidney disease, management of, 605
definition of, 343–344

Hypertension *(Continued)*
malignant, 349–350
outcomes of, 345–346
end-organ damage as, 345–346, 345*f*
portal
manifestations of, 759–763
pathophysiology of, 755, 761*f*
pregnancy-induced, 681
prevalence of, 343
primary, 344–348
pulmonary, 472–473, 473*b*
as risk factor for chronic kidney disease, 603
risk factors for, 344–345, 344*t*
secondary, 348–349, 348*b*–349*b*
subtypes of, 344
treatment interventions for, 346–348, 346*t*, 347*f*
Hypertensive cardiomyopathy, 401*t*
Hypertensive emergencies, 349–350
Hypertensive urgency, 349–350
Hyperthyroidism, 804–805, 804*b*, 804*t*, 805*f*
Hypertrichosis lanuginosa, in systemic diseases, 1084
Hypertrophic cardiomyopathy (HCM), 400–401, 400*f*, 401*t*
Hypertrophy, 63, 63*f*
cardiac muscle, 63, 64*f*
Hyperuricemia, asymptomatic, 1054
Hyperventilation, 468
causes of, 548*b*
as compensatory response to acid excess, 542, 543*t*
Hypocalcemia, 533, 534*b*
Hypochromic microcytic red blood cells, 279–280
Hypoglossal nerve (CN XII), 861*t*, 864, 864*f*
Hypoglycemia
complicating diabetes mellitus, 831
geriatric, 835
pediatric, 834
Hypogonadism, 648, 648*b*
Hypokalemia, 532, 532*b*
Hypomagnesemia, 534, 535*b*, 535*f*
Hypomania, in bipolar disorder, 983
Hypomenorrhea, 673
Hyponatremia, 526–527, 527*b*, 527*f*
Hypoosmolality, 527
Hypoparathyroidism, 812
Hypophosphatemia, 535–536, 536*b*
Hypophysis, 788
Hyporesponsiveness, to hormones, 797–798
Hyposecretion, of hormones, 797, 799–800, 800*f*
Hypospadias, 642, 643*f*–644*f*
Hypotension, orthostatic, 350, 351*b*
Hypothalamic-pituitary-adrenal (HPA) axis
impaired, in depression, 979
in response to stress, 14, 15*f*
Hypothalamic-pituitary dysfunction, 806
Hypothalamic-pituitary endocrine system, 788–793
Hypothalamic-pituitary-testicular axis, 633–637, 637*f*
Hypothalamus, 857, 859*f*
hormones of, 790–793, 791*f*–792*f*, 793*b*
Hypothyroidism, 803–804, 803*f*–804*f*, 804*b*, 804*t*
diagnosis of, 800–801, 801*f*
Hypotonic syndrome, 527
Hypoventilation, 468
causes of, 547*b*
as compensatory response to acid excess, 542–543, 543*t*
obesity, 514
Hypovolemic shock, 442–443
classification of, 442, 442*t*
clinical manifestations of, 442–443
etiology of, 435*b*, 442
pathogenesis of, 442, 442*f*
treatment of, 443
Hypoxemia, 461, 468–469
definition of, 452, 468
laboratory values in, 547*t*
Hypoxia, 468–469
in acute brain injury, 892–895
alveolar, 466
anemic, 468
cellular, in shock, 435, 436*f*
circulatory, 468
definition of, 468
histotoxic, 468
hypoxic, 468
Hypoxic hypoxia, 468

I

I bands, in sarcomere, 1015, 1015*f*
Iatrogenic condition, definition of, 2
Ichthyosis, 1078, 1079*f*
Icterus, in thalassemia, 280
Idiopathic condition, definition of, 2
Idiopathic inflammatory myopathy, 1038
IgA deficiency, 212
IgA nephropathy, 588
Ileocecal sphincter, 710
Ileocecal valve, 703, 711*f*
Ileum, 701
Illness
allostatic overload and, 21–23, 22*f*
chronic, infection risk and, 144
Immobility, nutritional considerations in, 848
Immobilization, fracture, 1028
Immune cells, evasion of, 145
Immune complex glomerulonephritis, 206
Immune complex reaction, 204. *see also* Type II hypersensitivity
Immune cytokines
in shock, 436, 437*t*
in stress and adaptation, 19, 21
Immune function, alterations in, 194–214
Immune-mediated disorders, 1046–1052
Immune response(s)
cell-mediated, mechanisms of, 178–181. *see also* Cell-mediated immunity
chemical mediators of, 167–169
chemokines as, 169
clotting factors as, 169, 169*f*
complement as, 167–169, 168*f*
cytokines as, 169, 170*t*
kinins as, 169, 169*f*
deficient, 210–213. *see also* Immunodeficiency disorders
primary immunodeficiency disorders as, 210–212
secondary immunodeficiency disorders as, 212–213
excessive, 195–209
autoimmunity as, 195–198. *see also* Autoimmunity
causes of, 195
hypersensitivity as, 198–209. *see also* Hypersensitivity
humoral, mechanisms of, 181–183. *see also* Humoral immunity
time phases in, 188*f*
Immune system
adaptive immunity in, 176–187. *see also* Adaptive immunity
changes in, 188*b*, 191*b*
components of, 159–169
chemical mediators, 167–169
epithelial barriers as, 159
leukocytes, 162–167
lymphoid, 160–162, 161*f*
mononuclear phagocyte system, 159–160, 159*f*
innate defense mechanism of, 169–175. *see also* Inflammatory response
integrated function and regulation of, 189–193, 190*f*
regulation of, 192
Immune thrombocytopenia (ITP), 306
Immunity
active, 187, 189*b*
herd, 144
passive, 187
Immunization
in infection prevention, 144
vaccines for, 189*b*
Immunodeficiency disorders
primary, 210–212
secondary, 212–213
Immunoglobulin(s). *see* Antibody(ies)
Immunoglobulin A (IgA), structure and properties of, 183, 185*f*
Immunoglobulin classes, properties of, 185*t*
Immunoglobulin D (IgD), structure and properties of, 183
Immunoglobulin E (IgE)
structure and properties of, 183, 186*f*
in type I hypersensitivity, 199
Immunoglobulin G (IgG), structure and properties of, 183
Immunoglobulin M (IgM), structure and properties of, 183–184
Immunomodulators, for autoimmune disorders, 198
Immunosuppression, infection risk and, 144
Immunosuppressive therapy, for autoimmune disorders, 197–198
Immunotherapy
allergen-specific, in allergic asthma management, 481
for cancer, 138
for type I hypersensitivity, 201
Impacted fracture, 1026, 1026*f*
Impaired fasting glucose tolerance, 820
Impaired glucose tolerance, 820
Impetigo, 1067–1069, 1068*f*
Implantation, of embryo, 663, 665*f*
Impotence, 645
Impulse initiation, abnormal site of, in dysrhythmias, 425–428
Inborn errors of metabolism, 780
phenylketonuria as, 106
Incomplete fracture, 1027, 1027*f*
Incontinence, urinary, 611–613
nocturnal, 613
risk factors for, 612*b*
Increased intracranial pressure (ICP)
acute brain injury and, 895–899, 896*b*, 896*f*
brain compression and herniation in, 897–899, 898*f*–899*f*

Increased intracranial pressure (ICP) *(Continued)*
etiology of, 895–897, 896*b*, 896*f*–897*f*
management of, 899, 899*b*–900*b*
manifestations of, 897, 898*f*
Incretin enhancers, for diabetes, 830, 832*t*
Incubation period, 3
Inert soft tissue injuries, 1021–1023, 1022*f*
Infant(s)
acid-base imbalance in, 550*b*
bleeding in, from vitamin K deficiency, 309
burns in, physiologic changes in, 1095*t*
fluid and electrolyte homeostasis and imbalances in, variations in, 537*b*
gastrointestinal system of, changes in, 716*b*
heart failure in, 406*b*
hemoglobin synthesis in, 266, 294*b*
infections in, risk of, 144
newborn
hemolytic disease of, 284
laboratory findings in, 275*t*–276*t*
integument in, 1089*b*
thermoregulation in, 840*b*
vision in, development of, 946*b*
pain in, 967
renal function in, 570, 571*b*
restrictive lung disease in, age-related features contributing to, 501*t*
in secondary hypertension, 348
skin disorders in, 1087–1088, 1088*f*
Infant respiratory distress syndrome, 507–508
Infection(s)
in AIDS patients, agents of, 248*b*
autoimmunity triggered by, 196
bacterial, 152*b*
by body area, bacteria associated with, 153*f*
bone, 1031–1032, 1031*f*, 1032*b*
central nervous system, acute brain injury from, 911–913, 912*t*
complicating hematologic neoplasms, prevention and management of, 218
definition of, 445*t*
fungal, 154*t*
gonococcal, 690
host characteristics influencing, 142–144, 143*t*
nongonococcal, 691
nutritional considerations during, 846–848
opportunistic, in HIV/AIDS, 246
oral, 724, 726*b*
parasitic, 155*t*–156*t*
sexually transmitted, 689–696. *see also* Sexually transmitted infections
transmission of, 147–149, 148*f*, 149*b*
chain of, 147, 147*f*
breaking, 148, 148*f*
emerging infectious diseases in, 148–149
environmental factors in, 143*f*, 148
immunization in prevention of, 148
impaired immune function in, 143–144
physical and mechanical barriers to, 142–143, 143*f*
routes of, 148
weapon of bioterrorism in, 149
viral, 154*t*
Infectious agents. *see* Pathogen(s)
Infectious arthritis, 1045
Infectious process(es), 141–157
Infective endocarditis, 398–399, 399*f*
Infertility, male, 649–650, 649*b*
Inflammation, 169–175
acute, mediators of, 172*t*
cardinal signs of, 169–171, 172*f*
chronic, 174
exudates, 175
leukocyte emigration in, 172, 173*f*
phagocytosis during, 172–174
systemic manifestations of, 175
vascular, blood flow alterations from, 325
Inflammatory bowel disease (IBD), 728–731
Inflammatory cardiomyopathy, 401*t*
Inflammatory myopathy, idiopathic, 1038
Inflammatory response
in allergic asthma, 480
cell-mediated, mechanisms of, 178–181
events in, 171, 171*f*
exudates in, 175
leukocyte emigration in, 172, 173*f*
mediators of, 172*t*
phagocytosis during, 172–174
systemic manifestations of, 175
vascular permeability in, increased, 171, 172*f*
Inflammatory skin conditions, 1069–1072, 1069*f*, 1070*t*
Inheritance
DNA mutation and repair in, 96–97
principles of, 95–97
Inhibitory postsynaptic potential (IPSP), 874
Inner ear, 937, 937*f*–938*f*
Inotropic drugs, for cardiogenic shock, 441
Inotropic effect, 372
Inotropic receptor, 874, 875*f*
Insomnia, in major depression, 980–981
Inspiratory capacity, 460*t*
Inspiratory reserve volume, 460*t*
Insulin
in diabetes mellitus treatment, 830–831, 832*t*
in glucose metabolism, 816, 816*f*–818*f*, 843, 843*t*
in lipid metabolism, 843*t*
Insulin edema, 831
Insulin-like growth factor-1 (IGF-1), secretion and activation of, 791, 792*f*
Integrase strand transfer inhibitors, 257
Integrins, 87, 172, 173*f*
Integumentary system, 1058–1092. *see also* Skin
age-related changes in, 1059–1061, 1060*b*, 1060*f*
evaluation of, 1061–1062, 1062*t*
manifestations of systemic diseases in, 1083–1085, 1086*b*
thermal burn effects on, 1094–1096
Intercalated cells (I cells), of collecting duct, 559
Intercellular communication, 48–55, 48*f*, 55*b*
cell signaling strategies in, 48–49, 48*f*–49*f*
cell surface receptor-mediated responses in, 49–53, 50*f*
Intercritical gout, 1055
Interest, diminished, in major depression, 980–981
Interferons, in cancer therapy, 138
Interleukin-1 (IL-1), in systemic effects of inflammation, 175, 175*f*
Interleukins, in cancer therapy, 138
Intermediate filaments, 30, 32*f*
Intermittent claudication
from arterial occlusion, 326–327
from arterial thrombosis, 324
pathophysiologic process of, 327*f*
in polycythemia vera, 292
International normalized ratio
alterations in, in hemostatic disorders, 305*t*
normal value and significance of, 305*t*
Interstitial cystitis/bladder pain syndrome (IC/BPS), 613–614
Interstitial fluid, 522, 522*f*
volume of, excess, 528–529, 530*f*
Intervertebral disks, 1007, 1023*f*
injuries to, 1023
Intestinal gas, 723
Intestinal obstruction, 733–734
Intestinal villi, 701–702, 702*f*
Intestines. *see also* Large intestine; Small intestines
bacteria targeting, 153*f*
Intraaortic balloon counterpulsation, for cardiogenic shock, 441
Intracellular accumulations, in cell injury, 60–62, 61*f*
Intracellular fluids, 522, 522*f*
Intracellular receptor-mediated responses, 53, 55*f*
Intracellular receptors, hormones with, 786, 786*f*
Intracerebral hemorrhage, 906
Intracranial hematomas, in traumatic brain injury, 903–905, 904*f*
Intracranial pressure, increased, in acute brain injury, 895–899, 896*b*, 896*f*
etiology of, 895–897, 896*b*, 896*f*
Intrahepatic cholestatic conditions, 780
Intramembranous ossification, 1003
Intrarenal disorders, 575–592
common manifestations of kidney disease, 575–577, 577*b*
congenital, 577–579, 579*b*
cystic kidney diseases, 578–579, 578*t*
renal agenesis and hypoplasia as, 577–578
infection as, 581–583, 582*b*, 582*f*
neoplasms, 579–580, 580*b*–581*b*
benign, 579, 579*t*
nephroblastoma (Wilms tumor), 580
renal cell carcinoma as, 579–580
obstruction of, 583–586, 583*f*, 584*t*, 586*b*
other diagnostic tests for, 576–577
pain in, 575–576
Intravenous urography/pyelography, 572–573
Intraventricular conduction defects, 429–431, 430*f*
Intrinsic factor, 700–701
lack of, in pernicious anemia, 278
Introns, 82
Intussusception, 734, 734*f*
Inulin, in GFR measurement, 572
Invasion, microbial, 144
Inversions, chromosomal, 100, 101*f*
Involuntary muscle, 90
Iodine deficiency syndromes, function and, 71*t*
Ion channel-linked receptors, 49–51, 50*f*
Ion channels, gating of, 44, 44*f*
Ipsilateral sensory transmission, 883
Iris, 943, 943*f*
Iron
functions and deficiency syndromes of, 71*t*
in hemoglobin synthesis, 266
Iron deficiency anemia, 279
laboratory findings in, 275*t*–276*t*

Iron deficiency syndromes, function and, 71*t*
Irritable bowel syndrome, 733
Ischemia
in acute brain injury, 892–895
from arterial occlusion, 325
atrophy from, 63
pathophysiology of, 385–387
Ischemic cardiomyopathy, 388, 401*t*
Ischemic hepatitis, 780–781
Ischemic pain, 966
Ischemic stroke, 906, 907*f*, 907*t*
Isochromosomes, 100, 101*f*
Isoimmunity, 201
Isometric contraction, of muscle, 1017
Isotonic saline, 525
Isovolumic contraction, in cardiac cycle, 358*f*, 359–360
Isovolumic relaxation, in cardiac cycle, 360

J

Jaccoud arthropathy, 1050
Janus kinase 3 (JAK3), 1049–1050
Jaundice, 755–759, 1082*b*
diagnostic tests for, 758–759
etiology and pathogenesis of, 755–758, 759*f*
evaluation of, 758, 760*t*
hepatic, 756–758
physiologic, of newborn, 778
posthepatic, 758
prehepatic, 756
Jejunum, 701
Joint(s), 1010*b*–1011*b*
bacteria targeting, 153*f*
contractures of, complicating burn wound healing, 1107
diarthroses, 1007–1010, 1009*f*
hypermobility of, 1056
structure and function of, 1006–1010, 1008*f*
synarthroses, 1007, 1008*f*
types of, 1006, 1008*f*
Joint capsule, 1010
Joint capsule injuries, 1022–1023, 1022*f*
Joint derangement, internal, 1023
Joint function disorders
acute rheumatic fever as, 1052
ankylosing spondylitis as, 1051–1052, 1052*f*
immune-mediated, 1046–1052
local, 1042–1046
Lyme disease as, 1045–1046
osteoarthritis as, 1042–1045. *see also* Osteoarthritis
pediatric, 1056–1057
postinfectious, 1052
reactive arthritis as, 1052
rheumatoid arthritis as, 1046–1050. *see also* Rheumatoid arthritis
scleroderma as, 1050–1051
secondary to other diseases, 1053–1056
systemic, 1046–1052
systemic lupus erythematosus as, 1050
Junctional dysrhythmias, 426, 427*f*
Junctional escape rhythm, 425, 425*f*
Junctional tachycardia, 426, 427*f*
Juvenile idiopathic arthritis (JIA), 1056–1057, 1057*b*
Juxtaglomerular apparatus, in tubuloglomerular feedback, 563, 563*f*
Juxtaglomerular cells, 342–343
in heart failure, 413

K

Kallidin, 169
Kallikrein, 169, 169*f*
Kaposi sarcoma, in HIV infection, 251, 251*f*, 1079
Keloid, 1063*f*–1064*f*
Keratinocytes, 1078, 1079*f*
Kernicterus, 755–756, 778
Ketoacidosis
diabetic, 825–826, 826*t*
laboratory values in, 547*t*
starvation, 545
Ketotifen, for HIV infection, 249
Kidney(s)
anatomy of, 552–554, 552*f*–555*f*, 554*b*
bacteria targeting, 153*f*
biopsy of, 573
blood supply to, 553–554, 553*f*–555*f*
chemicals toxic to, 604*b*
damage to, from hypertension, 346
diseases/disorders of
acute-on-chronic, 606
common manifestations of, 575–577, 577*b*
endocrine functions of, 570
failure of, chronic, anemia of, 277–278, 278*f*
laboratory findings in, 275*t*–276*t*
HIV infection manifestations involving, 253
injury to, acute, 593–600, 600*b*–601*b*
innervation of, 552–553
lymphatics in, 552–553
nephron structure and function in, 554–559. *see also* Nephron
transplantation of, for chronic kidney disease, 606–607
Kidney, ureter, and bladder (KUB)
roentgenography, 572
Kidney stones, 584
Kinins, 169, 169*f*
Klinefelter syndrome, 101, 103*f*
Knee joint, structures of, 1022*f*
Koilonychia, in systemic disease, 1084
Korotkoff sounds, 340, 340*t*
Krause corpuscle, 882*f*
Krebs cycle, 36, 38*f*
Kupffer cells, 159, 159*f*
Kwashiorkor, 845, 846*f*
Kyphoscoliosis, respiratory effects of, 511–512, 513*f*

L

L-selectins, 162–163
L-type channels, 370
Labia majora, 658, 660*f*
Lactation, 663
Lacteal, 715
Lactic acidosis, in ischemia-induced cell injury, 68–69
Lacunar infarcts, 906
Language deficits, after stroke, 908
Laparoscopic cholecystectomy, 748
Laplace, law of, 320, 320*f*
Large cell carcinomas, pulmonary, 476
Large intestine, functional anatomy of, 703, 703*f*–704*f*
Larynx, 453
Left anterior descending artery, 360
Left anterior fascicular block, 430
Left bundle branch block, 429–430, 430*f*
Left main coronary artery, 360
Left posterior fascicular block, 430
Left-sided heart failure, 417–418, 419*f*
Leiomyomas
renal, 579*t*
uterine, 678, 678*f*
Length-tension relationship, 375
in muscle contraction, 1017
Lens, of eye, 943, 943*f*
Leprosy, 1069
Leptomeninges, bacterial meningitis invading, 911
Leukemia
acute and chronic compared, 222*t*
acute lymphocytic and acute nonlymphocytic compared, 224*t*
hairy cell, 224, 225*f*
lymphoblastic, acute, 223–224, 224*f*, 224*t*
lymphoid, chronic, 223, 224*f*
myeloid
acute, 222, 222*f*
chronic, 221–222, 221*f*
Leukocytes, 89–90, 90*f*, 162–167
basophils as, 164
characteristics of, 263*t*
dendritic cells as, 165, 166*f*
emigration of, in inflammatory response, 172, 173*f*
eosinophils as, 163–164
lymphocytes as, 165–167
macrophages as, 164–165, 164*f*
mast cells as, 164, 164*f*
monocytes as, 164–165
neutrophils as, 160*f*, 162–163
proportions and functions, 163*t*
Leukoderma, 1080
Leukopenia, in cancer, 137
Leukoplakia, oral hairy, in HIV infection, 250–251, 251*f*
Leukotrienes, 171, 173*f*
LeVeen shunt, for ascites, 765, 766*f*
Lewy bodies, Parkinson disease, 921
Leydig cells
interstitial, 630*f*
in spermatogenesis, 637
Lice, 1075
Lichen planus, 1071, 1071*f*
Lichenification, 1063*f*–1064*f*
Life stress, panic disorder and, 990
Lifestyle alterations
in gastrointestinal disorders, stress of, 739–740
in hypertension, 346–348, 346*t*
Lifestyle considerations
in epidemiology, 8–9, 8*f*
in hypertension risk, 345
Ligament injuries, 1021, 1022*f*
Ligaments, structure and function of, 1012–1013, 1012*f*–1013*f*
Ligand(s), in intercellular communication, 49, 50*f*
Ligand-gated ion channels, 44, 44*f*
Limbic system, 855, 858*f*
Lipid(s)
absorption of, 715
digestion of, 713–714, 714*t*
intracellular accumulations of, in cell injury, 60
metabolism of, 840–841, 843*t*
Lipid bilayer, in plasma membrane, 27–28, 28*f*
Lipid core, 384
Lipidemias, HIV-associated, 253
Lipodystrophy, complicating diabetes mellitus, 831

Lipoid nephrosis, 591
Lipomas, 579*t*
Lipoproteins
  in atherosclerosis, 383, 384*f*
  atherosclerosis risk and, 329
Liquefactive necrosis, 64, 66*f*
Lithium, for bipolar disorder, 983
Lithium toxicity, 983
Lithotripsy, shock-wave, for renal calculi, 585–586
Liver
  bacteria targeting, 153*f*
  structure and function of, 755, 755*b*, 756*f*–757*f*
  transplantation of, 776–778, 777*b*
    evaluation of, 777
    management after, 777–778
  trauma to, 776
Liver abscess, 775–776
Liver disorders, 767–772
  advanced, complications of, 764–766
  age-related, 778–780
  alcoholic, 773–774
  cirrhosis as, 773–774, 780*b*
  congenital, 779–780
  geriatric considerations and, 780–781
  hemostatic disorders in, 310–311
  hepatitis as, 767–771
  laboratory tests in, 760*t*
  manifestations of, 755–766
    ascites as, 764–765, 765*f*
    gastroesophageal varices as, 760–763
    hepatic encephalopathy as, 763–764
    jaundice as, 755–759
    pathophysiology underlying, 758*t*
    spontaneous bacterial peritonitis as, 765–766, 766*f*
  pediatric considerations and, 778
  structural, 775–776
  toxic, 774–775
Liver spots, 1090, 1091*f*
Load-velocity relationship, in muscle contraction, 1017
Localized scleroderma, 1077, 1077*f*
Long-term memory, 888
Longitudinal fracture, 1026, 1026*f*
Loop diuretics, 569, 569*t*
Loop of Henle, 555*f*, 558–559, 560*f*–562*f*
  ascending
    epithelial cells of, 556*f*, 560*f*
    functions of, 556*t*
  countercurrent mechanism in, 559, 561*f*
  descending
    epithelial cells of, 556*f*
    functions of, 556*t*
Lotions, in skin care, 1086
Lou Gehrig disease, 928
Low blood pressure, 350–351, 351*b*
Low-density lipoproteins (LDLs)
  in atherosclerosis, 383, 384*f*
  atherosclerosis risk and, 327–329
Lower airway structures, 453–456, 455*f*–457*f*
Lower esophageal sphincter, 700
Lumbar plexus, 864–865, 865*f*–867*f*, 866*t*
Lund and Browder chart, in burn assessment, 1097, 1098*f*
Lung(s)
  bacteria targeting, 153*f*
  blood circulation in, 457, 457*f*
  blood flow in, 465–466
  blood vessels of, 465
  bronchopulmonary segments of, 454, 457*f*
Lung(s) *(Continued)*
  compliance of, in ventilation, 462
  fetal, development of, 452
  HIV infection manifestations in, 249–250, 250*f*
  mechanics of breathing and, 461, 462*f*
  ventilation in, 460–464. *see also* Ventilation
  ventilation-perfusion ratios in, 465–466, 466*b*
Lung cancer, screening guidelines for, 121*t*
Lung infection/inflammation, 514–518
Lung parenchyma
  disorders of, 500–504, 503*b*
    diffuse interstitial lung disease as, 500–501, 500*f*
    fibrotic interstitial lung diseases as, 500–504
    hypersensitivity pneumonitis as, 502, 502*t*
    occupational lung diseases as, 503–504, 503*t*–504*t*
    sarcoidosis as, 501–502
  loss of , pulmonary obstruction from, 488–491
Lung volumes and capacities, 460, 460*f*, 460*t*
Lupus erythematosus, 1069
  discoid, systemic compared with, 1070*t*
  systemic, 206–207, 1050
    discoid compared with, 1070*t*
Lusitropic effect, 372
Lusitropy, 360
Luteal phase, of menstrual cycle, 661, 661*f*
Luteinizing hormone (LH)
  release of, 792
  target organs and actions of, 660*t*
Lyme disease, 1045–1046, 1046*b*, 1076
Lymph, 322
Lymph nodes, 162*f*
  immune function of, 161
Lymphangiomas, 579*t*
Lymphatic drainage, blocked, edema from, 528–529, 530*f*
Lymphatic flow, 322
  control of, 323
Lymphatic flow alterations, 326, 333–334, 334*b*
  in lymphedema, 333–334, 334*f*
Lymphatic pumps, 322
Lymphatic system
  organization of, 315–317, 315*f*–316*f*, 317*b*
  structure of, 317, 318*f*
Lymphatic vessels
  dynamics of, 321–322
  flow alterations in, 326
  structure of, 317
Lymphatics
  immune function of, 161
  renal, 552–553
Lymphedema, 321–322, 326, 333–334
Lymphoblasts, 223–224, 224*f*
Lymphocyte(s), 165–167
  B, 167, 167*f*
  characteristics of, 263*t*
  functions of, 163*t*
  natural killer cells as, 166
  T, 166–167, 167*f*
  in type IV hypersensitivity, 207
Lymphogranuloma venereum, 692, 692*f*
Lymphoid leukemia, chronic, 223, 224*f*
Lymphoid neoplasms, 218*f*, 223–229, 223*f*, 226*b*
  acute lymphoblastic leukemia/lymphoma as, 223–224, 224*f*, 224*t*
  B-cell, T-cell, and NK-cell lymphoma as, 229, 230*t*, 231*b*
  chronic lymphoid leukemia as, 223, 224*f*
  classification of, WHO, 217*b*
Lymphoid neoplasms *(Continued)*
  complications of, prevention and management of, 219–220, 220*b*
  etiology of, 216–217
  hairy cell leukemia as, 224, 225*f*
  Hodgkin disease as, 227–229, 227*f*–228*f*, 228*b*
  manifestations of, 218*b*
  plasma cell myeloma as, 224–226, 225*f*
  treatment principles for, 218–219
Lymphoid system, 160–162, 161*f*
  primary organs of, 160–161, 169*b*
  secondary organs of, 161–162
Lymphoma
  B-cell, T-cell, and NK-cell (non-Hodgkin), 229, 230*t*, 231*b*
  Hodgkin, 227–229. *see also* Hodgkin lymphoma
  lymphoblastic, acute, 223–224
Lymphopoiesis, definition of, 263
Lysosomes
  in cellular atrophy, 63
  function of, 57*t*

M

M line, in sarcomere, 1015, 1015*f*
Macrocytic erythrocytes, 278
Macromolecules, membrane transport of, 39–40
Macrophages, 159, 159*f*
  antigen presentation function of, 165
  characteristics of, 263*t*
  cytokine secretion by, 169, 170*t*
  destruction of, in hemolytic disease of newborn, 203
  functions of, 163*t*
  impaired, by HIV infection, 242
  phagocytic function of, 165, 172–174
  receptors on surface of, 164–165, 165*f*
  secretory function of, 165, 166*f*
Macular degeneration, age-related, 948, 949*f*
Macule, 1063*f*–1064*f*
Magnesium, plasma, imbalances in, 534–535, 535*b*, 535*f*
Magnetic resonance imaging (MRI)
  in cardiac function, 378
  renal, 573
Major depressive disorders, 978–982
  clinical manifestation of, 980–981
  etiology and neurobiology of, 979–980, 979*t*, 980*f*
  risk factors for, 979*t*
  treatment of
    nonpharmacologic, 981–982
    pharmacologic, 981, 981*f*
Major histocompatibility complex (MHC), 176, 176*f*
  antigen presentation by, 176–178
    class I proteins in, 176*f*–178*f*, 177
    class II proteins in, 177–178, 179*f*
  genes for, autoimmunity and, 195, 197*t*
Malabsorption disorders, 735, 737*b*
  after surgical intervention, 735–737
Maladaptation, 20
Malaria, geographic distribution of, 9, 9*f*
Male genital and reproductive function, 626–640, 638*b*
  alterations in, 641–655
    of penis and male urethra disorder, 641–647, 647*b*
    of prostate disorders, 651–655, 655*b*
    of scrotum and testes disorders, 647–650, 651*b*

Male genital and reproductive function *(Continued)*
anatomy of, 626–631, 632*b*
auxiliary genital glands in, 628–629
external genitalia in, 629–631, 633, 636*f*
lower genitourinary tract in, 627–628, 627*f*
upper genitourinary tract in, 626–627, 627*f*
embryology of, 632–633, 633*b*
physiology of, 633–640, 640*b*
hypothalamic-pituitary-testicular axis in, 633–637, 637*f*
spermatogenesis in, 637, 638*f*
Malignancy
liver, 776
pulmonary, 475–476, 476*b*
Malignant cells, abnormal behaviors of, 118–119
Malignant hypertension, 349–350
Malignant melanoma, 1079–1080, 1080*f*
Malignant tumors
abnormal phenotype of, 118–119, 118*t*
characteristics of, 118, 118*t*
nomenclature for, 119*t*
Mallory-Weiss syndrome, 725
Malnutrition
cell injury from, 70
complicating chronic kidney disease, 603
disorders of, complicating chronic kidney disease
management of, 605
in HIV infection, 249
infection risk and, 143–144
Malunion, in fracture healing, 1029
Mammary duct ectasia, 682–683, 682*f*
Mammography, digital, 684
Mandibular nerve, 861*t*
Mania, bipolar disorder and, 983
Marasmus, 845, 845*f*
Marfan syndrome, 103–104, 106*f*
Margination, 172
Mass movement, 710
Mast cells, 164, 164*f*
autoimmunity and, 196
in type I hypersensitivity, 199, 200*f*
Mastitis, periductal, 682
Maternal age, Down syndrome frequency related to, 101, 102*t*
Mature sperm cell, anatomy of, 638*f*
Maxillary nerve, 861*t*
Mean arterial pressure (MAP), 338–339
Measles, 1089–1090
Measles-mumps-rubella (MMR), 1089
Mechanical injury, cellular, 72–73
Mechanically-gated ion channels, 44, 44*f*
Mediastinum, heart in, 355, 355*f*
Medical anthropology, 7
Medulla oblongata, 860
in control of respiration, 462, 464*f*
Medullary cavity, 1003, 1003*f*
Medullary pyramids, 860
Megacolon, 734
Megaloblastic dysplasia, in pernicious anemia, 278
Megaloblastic madness, in vitamin B_{12} deficiency, 278
Megaloblasts, 278
Meiosis, 95, 97*f*–98*f*
Meissner corpuscle, 882*f*
Melanoma, malignant, 1079–1080, 1080*f*
Melena, 303–304
Membrane(s), fetal, formation of, 663–664
Membrane attack complex, 168–169, 168*f*
Membrane calcium transporters, 42, 42*f*
Membrane channel proteins, 44, 44*f*
Membrane modification, in drug-induced immune hemolysis, 285
Membrane transport carriers, 42–43
Membrane transporters
ABC, 42, 43*f*
calcium, 42, 42*f*
Membranous nephropathy (MN), 590, 590*f*
Memory, 886–888
Menarche, 659
Mendelian single-gene disorders, 98–99, 102–107
autosomal-dominant, 103–104, 105*t*
Huntington disease as, 104
inheritance pattern of, 105*f*
Marfan syndrome as, 103–104, 106*f*
autosomal-recessive, 104–106, 106*t*
albinism as, 105–106
cystic fibrosis as, 106, 108*f*
inheritance pattern of, 107*f*
phenylketonuria as, 106
overview of, 102–103
sex-linked, 106–107
Meniere disease, 940
Meninges
bacteria targeting, 153*f*
cranial, 851, 852*f*
spinal, 851, 853*f*
Meningitis, 911–912, 912*t*
Meningocele, 927
Menisci, 1007–1009
Menopause, 668–669, 669*b*
Menorrhagia, 303–304, 673
Menstrual cycle, 659–662, 661*f*, 662*b*
Menstrual disorders, 672–673, 674*b*
Menstruation, 659
Mental illness, lifetime prevalence of, 971–972, 972*t*
Merkel corpuscle, 882*f*
Mesangium, glomerular, 557
in glomerular filtration regulation, 564
Mesencephalon, 860
Mesolimbic dopaminergic system, 972
Mesonephros, 632, 632*f*
Metabolic acidosis, 545–546, 545*b*
complicating chronic kidney disease, 603
management of, 605
in electrical injury, 1108
Metabolic alkalosis, 547–548, 548*b*
Metabolic bone disease, 1033–1036, 1034*b*, 1034*f*
Metabolic cardiomyopathy, 401*t*
Metabolic disorders, 838–849
Metabolic syndrome, 844, 844*b*
atherosclerosis risk and, 329
in hypertension, 345
Metabolism, 838–849
aging and, 846
anabolism in, 839
body fluids in, 522
catabolism in, 839
cellular, 34–39
citric acid cycle in, 36, 38*f*
glycolysis in, 36, 37*f*
oxidative phosphorylation in, 36–39, 39*f*
definition of, 838–839
enzyme pathways in, 838–839
genetics, epigenetics, and environment in, 841–842
Metabolism *(Continued)*
glucose, regulation of, 816–820, 820*b*. *see also* Glucose metabolism
inborn errors of, 780
phenylketonuria as, 106
maternal, during pregnancy, 667
nutrient, 839–841. *see also* Nutrient metabolism
pediatric considerations in, 840*b*
processes in, 838–839, 839*b*
rate of, 839, 839*t*. *see also* Basal metabolic rate
in starvation and physiologic stress, 844–846, 845*f*–846*f*, 846*b*, 847*t*
Metabotropic receptor, 874, 875*f*
Metal storage diseases, 774–775
Metanephric adenoma, 579*t*
Metanephros, 632, 632*f*
Metaplasia, 63–64, 63*f*
Metastasis, 132–135
of breast cancer, 684–685
in diagnosis of malignancy, 118
mechanisms of, 132, 133*f*–134*f*
patterns of spread in, 132–133
Metformin, for diabetes, 829–830
Methemoglobin, in erythrocyte cycle, 268
Methotrexate, for autoimmune disorders, 198
Micelle, 743, 745*f*
Microbiome, 142
Microcirculation, dynamics of, 321–322
Microglia, 868, 873*f*
Microglial cells, 159, 159*f*
β_2-Microglobulin test, in monitoring HIV disease status, 246
Micronutrients, infection risk and, 143–144
Microorganisms
characteristics of, virulence and invasiveness, 144
types of, 151*f*, 153*b*
Micropenis, 641–642
Microtubules, 30, 32*f*
Microvilli, 701–702, 702*f*
Micturition
contractions, 610–611
mechanism of, 610–611
physiology of, 610–611
Midbrain, 860
Middle ear, 937, 937f
infections of, 938
inflammation of, 941
Middle East respiratory syndrome, 516, 517*t*
Miglitol, for diabetes, 830
Migraine headache, 962, 963*b*
Milk ejection reflex, 663
Mineral disorders, complicating chronic kidney disease, 603
management of, 605
Mineral dusts, intracellular accumulations of, 60–61, 62*f*
Mineralocorticoids, 18–19
actions of, 796
Minimal change disease (MCD), 591
Minute ventilation, 460
Miscarriage, 681
Mitochondria, 34, 35*f*
function of, 57*t*
Mitochondrial gene mutations, 109
Mitogen growth factors, in cell proliferation, 54, 57*f*, 124
Mitosis
in cell proliferation, 54
stages of, 56*f*

Mitral valve, 355, 356*f*
Mitral valve disorders, 394–397
prolapse as, 396–397, 396*f*
regurgitation as, 396, 396*f*
stenosis as, 395–396, 395*f*
Mixed epithelial stromal neoplasm, renal, 579*t*
Mixing movements, 707
in small intestine, 709
Mobility, microbial, 145
Modification of Diet in Renal Disease study equation (MDRD), 572
Molecular mimicry theory, of autoimmunity, 195
Molecular therapy, for cancer, 138–139
Molecules, small, membrane transport of, 40–44
Moles, 1088*f*
Molluscum contagiosum, 694
Mönckeberg sclerosis, 326
Mongolian spots, 1087, 1088*f*
Monoamine oxidase inhibitors (MAOIs) , for major depression, 981
Monoclonal antibodies, in cancer therapy, 138, 139*f*
Monoclonal gammopathy of undetermined significance, in plasma cell myeloma, 225
Monocytes, 159, 159*f*
characteristics of, 263*t*
functions of, 163*t*
Monogenic traits, 95–96
Mononuclear phagocyte system, 159–160, 159*f*
Monosaccharides, dietary, 839
Monosomy, 99–100
Monosomy X, 102, 103*f*
Monro-Kellie hypothesis, 895
Mons pubis, 658, 660*f*
Mood, depressed, 980–981
Mood disorders
in elderly, 984
unified model of, 982*f*
Mood stabilizers, for bipolar disorder, 983
Morphogenesis, 663
Morphogens, 87
Mosquitoes, 1076
Motility disorders, 733–734, 735*b*
Motor deficits, after stroke, 908
Motor neurons, 883–884
Mouth disorders, 724–725, 726*b*
Mucocutaneous candidiasis, 1065–1067, 1068*f*
Mucocutaneous manifestations, of HIV infection, 250–251, 250*f*–252*f*
Mucopolysaccharidoses, 60
Mucosa-associated lymphoid tissue (MALT), 162
Mucosal disorders, 735
Multifactorial traits, 96, 110–111
Multipennate muscle, 1014
Multiple myeloma (plasma cell myeloma), 224–226, 225*f*–226*f*, 1038
Multiple organ dysfunction syndrome (MODS)
definition of, 445*t*
in shock, 449–450
Multiple sclerosis, 925–926, 926*b*, 926*f*, 927*t*–928*t*
Multiple X females, 102
Multipolar neurons, 868, 871*f*
Multisystem enzyme deficiencies, 779
Murmurs, 394
defining characteristics of, 395*t*
Muscle(s)
cardiac, 90, 91*f*
myoepithelial cells in, 91, 91*f*
skeletal, 90, 91*f*
smooth, 90, 91*f*
Muscle contraction
calcium in, 1015, 1016*f*
concentric, eccentric, and isometric contraction of, 1017
electromechanical coupling in, 1015–1016
fatigue and, 1017
mechanical principles of, 1017
mechanics of, 1015–1017, 1018*b*
mechanism of, 90
response to movement and exercise, 1017
sliding filament theory, 1015
temperature change and, 1017
twitch, 1016–1017
Muscle relaxation, energy of, 367, 367*b*–368*b*, 369*f*
Muscle tissue, 90–91, 91*f*–92*f*
types and locations of, 88*t*
Muscular dystrophies
cardiomyopathies and, 401t
respiratory effects of, 511, 512t
Muscular dystrophy, 1039, 1040*b*
Musculoskeletal disorders, rheumatic, 1042–1057. *see also* Joint function disorders
Musculoskeletal function, alterations in
bone and joint trauma as, 1026–1030
bone injuries and infections as, 1026–1030, 1030*b*, 1031*f*, 1032*b*
in bone structure and mass, 1032–1038
bone tumors as, 1036–1038
chronic muscle pain as, 1040
muscular dystrophy as, 1039, 1040*b*
myasthenia gravis as, 1039–1040
soft tissue injuries in. *see* Soft tissue injuries
Musculoskeletal system, 1001–1019
articular cartilage in, 1011–1012, 1012*b*, 1012*f*
changes in, 1018*b*–1019*b*
joints in, 1006–1010. *see also* Joint(s)
skeletal muscles in, 1013–1015, 1014*f*–1015*f*, 1015*b*
structure and function of bone, 1002–1006. *see also* Bone
tendons and ligaments in, 1012–1013, 1013*b*
Mutagens, 96
Mutation(s), genetic, 97, 99*f*
mitochondrial, 109
proto-oncogene activity and, 126*f*, 127
single-gene, 98–99, 102–107
triplet repeat, 107
Myasthenia gravis, 204, 204*f*, 1039–1040
respiratory effects of, 511, 512*t*
Mycoplasmal pneumonia, 514, 515*t*
Mycoses, 151
Myeloid leukemia
acute, 222, 222*f*
chronic, 221–222, 221*f*
myeloid cell maturation pathways and, 221*f*
Myeloid neoplasms, 218*f*, 221–222, 223*b*
classification of, WHO, 216*b*
complications of, prevention and management of, 219–220, 220*b*
etiology of, 216–217
treatment principles for, 218–219
Myeloma, plasma cell (multiple), 224–226, 225*f*
Myenteric plexus, in gastrointestinal motility, 704
Myocardial cells, rhythmicity of, 370–371, 371*f*
Myocardial diseases, 399–401, 401*b*–402*b*
cardiomyopathy as, 400–401, 400*f*
myocarditis as, 399–400, 399*b*
Myocardial hypertrophy, in heart failure, 415–417, 415*f*–416*f*
Myocardial infarction (MI)
clinical course of, 391–392, 392*f*–393*f*
diagnosis of, 389–390
electrocardiographic changes in, 390–391, 391*f*
location of, according to coronary artery affected, 390*t*
morphologic changes in, evolution of, 389, 390*t*
pericarditis and, 403
prognosis and treatment of, 392–393, 393*f*
serum biomarkers in, 391, 391*f*
silent, 390
Myocardial workload, in atherosclerosis, 386
Myocarditis, 399–400
etiologic agents of, 399*b*
postviral, 400
Myocardium, 355–356
chambers of, 356
normal pressures in, 356*t*
Myoclonic seizures, 916
Myoepicardial mantle, 403
Myoepithelial cells, 91, 91*f*
Myogenic autoregulation, GFR and, 562
Myoglobinuria, in electrical injury, 1108
Myopia, 944–945, 945*f*
Myosin filament, 1015
Myotonic dystrophies, 1039
Myxedema coma, 804, 804*f*

N

N-methyl-D-aspartate (NMDA) receptor antagonists, for Alzheimer disease, 920
Nails
age-related changes in, 1060*b*, 1061
manifestations of systemic diseases in, 1084–1085
Nasal cavity, 452, 453*f*
Natriuretic peptides, secretion of, 376
Natural killer cells, 166
Nebulin, 366
Necrosis, 64–65, 65*f*, 68*b*
Necrotizing enterocolitis, 731
Negative inotropes, 376
Neoantigen formation, in drug-induced immune hemolysis, 284–285
Neoplasia, 117–140
benign *versus* malignant growth in, 118–119, 118*t*–119*t*, 119*b*
carcinogenesis and, 129–132, 130*f*–132*f*, 131*b*
definition of, 117–118
epidemiology and risk factors for, 120–123, 120*f*, 121*t*, 122*f*–123*f*
genetic mechanisms of, 123–129, 124*t*, 125*f*, 127*t*
hematologic, classification of, 215–216
in HIV infection, 251
metastasis and, 132–135, 135*b*
Neoplasms
benign, specific, 683–684
of lower urinary tract, 617–618, 618*b*–619*b*
of penis, 647
of testis, 650, 651*f*
Nephrectomy, for benign renal neoplasms, 579
Nephric system, male, embryology of, 632, 632*f*
Nephritic syndrome, 587
Nephroblastoma (Wilms tumor), 580
Nephrogenic diabetes insipidus, 568
Nephrolithiasis, 584–586
clinical manifestations of, 585
contributing factors to, 584*b*
diagnosis and treatment of, 585–586, 585*b*

Nephrolithiasis *(Continued)*
etiology and pathogenesis of, 584–585
general interventions for, 585*b*
types of, 584–585, 585*t*
Nephroma, 579*t*
Nephron
diagram of, 544*f*
epithelial cells of, 554–555, 556*f*
loss of, in staging of chronic kidney disease, 602*t*
structure and function of, 554–559, 555*f*–556*f*, 556*t*
collecting duct (tubule) in, 555*f*–556*f*, 559
distal convoluted tubule in, 555*f*–556*f*, 559
glomerulus in, 555–557, 555*f*, 557*f*
loop of Henle in, 555*f*–556*f*, 558–559, 560*f*–562*f*
proximal convoluted tubule in, 555*f*–556*f*, 557–558, 560*f*
tubules of. *see* Renal tubules
Nephropathy
AIDS/HIV-associated, 253
diabetic, 826–827
IgA, 588
membranous, 590, 590*f*
Nephrotic syndrome, 587, 589–591
pathophysiologic process of, 589*f*
Nephrotoxins, 595
Nerve(s)
cranial, 860, 863–864. *see also* Cranial nerves
injuries to, 1023
parasympathetic, distribution of, 867, 869*f*
of spinal cord, 861–867, 863*f*, 865*f*–867*f*, 866*t*, 869*f*
supplying kidneys, 553
supplying lower genitourinary tract, male, 627–628
supplying lower urinary tract, 610
supplying penis, 631
sympathetic, distribution of, 867, 870*f*
Nerve roots, injuries to, 1023
Nervous system
aging in, 881, 881*b*–882*b*
autonomic, 867, 867*b*–868*b*, 869*f*–871*f*, 872*t*
central, 851–863. *see also* Central nervous system
development of, 880–881, 882*b*
HIV manifestations affecting, 252
injury to, 881–882, 882*b*
peripheral, 863–867. *see also* Peripheral nervous system
structural organization of, 851–867
structure and function of, 850–890
Nervous tissue, 91–92
Neural function
consciousness and memory in, 886–888, 889*f*, 890*b*
motor, 883–886, 886*b*, 886*f*
central control of, 885–886, 888*f*
motor neurons in, 883–884
spinal reflexes in, 884–885, 887*f*
neuronal communication in, 872–880. *see also* Neuronal communication
sensory, 882–883, 883*b*
pathways of, 883, 884*f*
receptors in, 882–883, 882*f*–883*f*
in somatosensory cortex, 883, 885*f*
sleep in, 888–890, 889*f*, 890*b*
Neural plasticity, 854, 868–872, 880–881
Neural stem cell, 868
proliferation, 868, 874*f*
Neural thread protein, 918
Neural tube defects, folate deficiency and, 278
Neuralgia, postherpetic, 966, 1065
Neurocrine signaling, 784, 784*f*
Neurodevelopmental disorders, 995–998
autism spectrum disorder as, 997–998, 998*b*
Neuroendocrine response, to stressor, 15*f*
Neurofibrillary tangles, 918, 918*f*
Neurogenic bladder, 614, 614*b*
Neurogenic shock, 444, 931
Neuroglia, 92, 868, 873*f*
Neurohormonal mediators, of stress and adaptation, 17–19
Neurohypophysis, 788
Neurologic function, chronic disorders of, 915–935
brain and cerebellar, 916–917, 916*b*–918*b*, 925*b*
spinal cord and peripheral nerve disorders as, 925–932
Neuromuscular disorders
affecting respiratory system, 511, 512*t*, 514*b*
cardiomyopathies and, 401*t*
Neuron(s)
action potentials of, 45, 47*f*
corticospinal, 885
dopaminergic, degeneration of, in Parkinson disease, 921, 921*f*
excitatory or inhibitory, 868
function of, 868
membrane potentials in, 874, 875*f*
motor, 883–884
neurotransmitters secreted by, 867, 871*f*
parts of, 91, 92*f*
preganglionic and postganglionic, 867, 869*f*
presynaptic and postsynaptic, in memory, 888, 889*f*
structure of, 868
types of, 868, 871*f*
Neuronal circuits, in neuronal communication, 880, 880*f*
Neuronal communication, 872–880
membrane potentials in, 872–874, 875*f*
neuronal circuits in, 880, 880*f*
neurotransmitters in, 875–880, 876*b*, 876*t*, 877*f*–879*f*
synaptic transmission in, 874, 875*f*–876*f*
Neuropathic osteoarthropathy, 1054
Neuropathy, diabetic, 827, 965–966
Neuropeptide neurotransmitters, 879
Neurotransmitter(s)
dysregulation of, in bipolar disorder, 982–983
in generalized anxiety disorder, 991–992
in neuronal communication, 875–880, 876*b*, 876*t*, 877*f*
pain transmission and, 957
preganglionic and postganglionic, 867, 871*f*
in schizophrenia, 978*t*, 980*f*
Neurovascular injury, in complications of fractures, 1030
Neutropenia, in hematologic neoplasms, 218
Neutrophilia, 163
Neutrophils, 160*f*, 162–163
characteristics of, 263*t*
functions of, 163*t*
toxicity of, to normal tissues, 163
Nevi, 1087, 1088*f*
Newborn heart, changes in, 405*b*
Nitric oxide (NO), 51
in coronary circulation, 362
excess production of, in shock, 436–437, 437*f*
as neurotransmitter, 880
NK cell lymphoma, 229, 230*t*, 231*b*
Nociception, 955–956
modulation of pain signals in, 955–956, 958–960, 960*b*, 960*f*–961*f*
perception of pain signals in, 956*f*, 958
transduction of stimuli in, 956, 956*f*
transmission of stimuli in, 956–958, 956*f*–957*f*, 956*t*, 959*f*
Nociceptors, 956
Nocturia, 611
Nodule, 1063*f*–1064*f*
Noise-induced hearing loss, 939
Non-Hodgkin lymphoma, 229, 230*t*, 231*b*
clinical differences, Hodgkin disease, 231*t*
Nonarticular rheumatism, 1056
Nondisjunction, 99–100, 100*f*
Nondisplaced fracture, 1027
Nonketotic hyperglycemic hyperosmolar syndrome, 826, 826*t*
Nonmendelian single-gene disorders, 107–110
from genomic imprinting, 110, 110*f*
from mitochondrial gene mutations, 109
from triplet repeat mutations, 107
Nonnucleoside reverse transcriptase inhibitors, 255–257
Nonopiates, for migraine, 964*t*
Nonpsychotic illnesses
anxiety disorders as, 989–995
neurobiology of, 989–1000
neurodevelopmental disorders as, 995–998
Nonsteroidal antiinflammatory drugs (NSAIDs)
for migraine, 964*t*
for osteoarthritis, 1043–1045
Nonulcerative lesions, sexually transmitted infections and, 694–695
Nonunion, in fracture healing, 1029
Norepinephrine (NE)
in glucose metabolism, 843
as neurotransmitter, 867, 871*f*, 876*t*, 878*f*
schizophrenia and, 978*t*
in stress and adaptation, 17–18
Normality, in health and disease, 4–6
data reliability, validity and predictive value and, 5
individual factors influencing, 5–6
statistical, 4–5, 4*f*
Nuclear cardiography, in cardiac function, 379
Nuclear pores, 30, 33*f*
Nucleic acid hybridization techniques, 114, 115*f*
Nucleolus, 30, 33*f*
ribosome assembly in, 82–83
Nucleoside reverse transcriptase inhibitors, 255
Nucleosomes, 78–79, 80*f*
Nucleotide reverse transcriptase inhibitors, 255
Nucleotides, 78, 78*f*
Nucleus, 30–31, 33*f*
function of, 57*t*
Nucleus tractus solitarius, 709
Nutrient metabolism, 839–841, 841*b*
carbohydrates, 839–840, 843*t*
lipids, 840–841, 843*t*
proteins, 841
regulation of, 841–844, 844*b*
genetics, epigenetics, and environment in, 841–842
hormonal, 843–844, 843*t*

Nutrients, movement of, through GI tract, 707–711
Nutrition, diabetes mellitus treatment, 828, 829*b*
Nutritional considerations, 846–848, 848*b*
  in aging, 846
  in burns, 848
  in cancer, 848
  in immobility, 848
  in infection, sepsis, and fever, 846–848
  in surgery, 848
  in trauma, 848
Nutritional disorders, 838–849
Nutritional status
  cell injury and, 70
  infection risk and, 143–144
  maintaining, in hematologic neoplasms, 219
Nutritional support, for burns, 1105–1106
Nystagmus
  in acute brain injury, 902
  vertigo and, 938

O

Oat cell carcinoma, pulmonary, 476
Obesity, 844, 844*b*
  allostatic overload and, 22
  atherosclerosis risk and, 329
  burn injuries in, 1110–1111
  diabetes mellitus and, 828–829
  in hypertension, 345
  respiratory effects of, 513–514
Oblique fracture, 1026, 1026*f*
Obsessions, in obsessive-compulsive disorder, 993
Obsessive-compulsive disorder, 992–994, 993*f*
Obstructive pulmonary disorders, 478–498. *see also* Pulmonary disorders, obstructive
Obstructive shock, 441, 441*b*
  clinical manifestations of, 441
  etiology of, 435*b*, 441
  pathogenesis of, 441
  treatment of, 441
Obstructive sleep apnea (OSA), hypertension in, 349
Occupational lung diseases, 503–504, 504*b*
  asthma as, 479
  clinical manifestations of, 504
  diagnosis of, 504
  etiology of, 503, 503*t*
  pathogenesis of, 503–504, 504*t*
  treatment of, 504
Ocular manifestations, of HIV infection, 253
Oculocephalic test, 902, 902*f*
Oculomotor nerve (cranial nerve III), 860, 861*t*, 863, 864*f*
Oculovestibular response, 902
Odynophagia, 721
Ogilvie syndrome, 733
Ohm's law, 319
Ointments, in skin care, 1086
Olfactory nerve (CN I), 860, 861*t*, 863, 864*f*
Olfactory system, 952
Oligoarticular onset, of juvenile idiopathic arthritis, 1057*b*
Oligodendrocytes, 868, 873*f*
Oligomenorrhea, 673
Oligosaccharides, dietary, 839
Oliguric phase, of acute kidney injury, 599–600
Oncocytoma, renal, 579*t*
Oncogenes, 124
  activating proto-oncogenes to become, 125–127
Onycholysis, in systemic diseases, 1084, 1085*f*
Oogonia, 657
Open cholecystectomy, 748
Open fracture, 1027, 1028*f*
Ophthalmic nerve, 861*t*
Opiates, for migraine, 964*t*
Opioids
  pain modulation and, 959–960, 960*t*
  receptor affinity of, 961*t*
Optic nerve (CN II), 860, 861*t*, 863, 864*f*, 943, 944*f*
Oral cavity
  in elderly, changes in, 700*b*
  functional anatomy of, 699, 699*f*
  infections of, 724, 726*b*
Oral hairy leukoplakia, in HIV infection, 250–251, 251*f*
Oral mucosa, 1082*b*
Organ of Corti, 937
Organelles
  cytoskeleton and, 30
  endoplasmic reticulum as, 31–32, 34*f*
  Golgi apparatus as, 32, 34*f*
  lysosomes in, 33–34
  mitochondria in, 34, 35*f*
  nucleus as, 30–31, 33*f*
  peroxisomes in, 33–34
Organism(s), death of, 75
Orthopnea, in left-sided heart failure, 417
Orthostatic hypotension (OH), 350, 351*b*
Osmolality, regulation of, kidneys in, 567–570
Osmoreceptors, 789
Osmotic diarrhea, 723–724
Osmotic diuretics, 569, 569*t*
Osmotic pressure, interstitial fluid, increased, edema from, 528–529, 530*f*
Ossicles, 937
Ossification, in fracture healing, 1006
Osteoarthritis, 1042–1045
  clinical manifestations of, 1043, 1044*f*–1045*f*
  etiology and pathogenesis of, 1043, 1043*f*–1044*f*
  treatment of, 1043–1045
Osteoarthropathy, neuropathic, 1054
Osteoblast, 1002, 1002*f*
Osteoclasts, 1002, 1002*f*
Osteocyte, 1002, 1002*f*
Osteoid, 1002
Osteoid osteoma, 1037, 1037*f*
Osteomalacia, 1035
Osteomyelitis, 1031–1032
  clinical manifestations of, 1031
  etiology and pathogenesis of, 1031, 1031*f*
  in fracture healing, 1029
  healing complications in, 1031–1032
  treatment of, 1032
Osteon, 1002–1003, 1002*f*
Osteonecrosis, in fracture healing, 1029
Osteophyte spur formation, in osteoarthritis, 1043, 1044*f*
Osteoporosis, 1033–1035, 1034*b*, 1034*f*
Osteosarcoma, 1037, 1037*f*
Ostium secundum, 403
Otitis media, 941–942, 941*t*
  acute, 941–942
  chronic, 942
Otosclerosis, hearing loss in, 939
Ototoxic medications, 938
  hearing loss from, 939
Ovarian cancer, 680
Ovarian cysts, 678–679
Ovarian follicles, 658*f*
Ovary(ies), 657
  development of, 633, 635*f*
Overactive bladder syndrome (OAB), 613
Overdrive suppression, 371–372
Overflow incontinence, 612
Oviducts, 658
Ovulation, 661
Oxidative phosphorylation, 36–39, 39*f*
Oxygen
  partial pressure of, 268
  utilization of, in cardiac energy metabolism, 368–369
Oxygen consumption ($\dot{V}O_2$), 269–270, 273*t*
Oxygen delivery, 269, 273*t*
  in shock, 447
Oxygen-hemoglobin dissociation curve, 269, 271*f*
Oxygen radicals (free radicals)
  in ischemic injury, 69
  in shock, 435, 436*f*
Oxygen transport, 268–270, 270*f*, 271*b*–272*b*, 467, 469*f*
  alterations in, 259–297
Oxygenation
  alveolar, 461
  tissue, impaired, in shock, 435–437, 436*f*–437*f*, 437*t*
Oxyhemoglobin dissociation curve, 467, 469*f*
Oxytocin
  in parturition, 667
  physiologic actions of, 790
  release of, 790
  in stress and adaptation, 19

P

P wave, 373–374, 373*f*
p24 antigen, in monitoring HIV disease status, 246
P53 gene, 128, 130*f*
Pacemakers, in heart failure, 421
Pacinian corpuscle, 882*f*
Paget disease, 1035–1036, 1036*f*
Pain, 955–970
  abdominal, 721–723
  acute, 961–962
    physiologic responses to, 962*t*
  in cancer, 136
  cancer-related, 964
  chronic, 962–963
  complicating chronic kidney disease, 603–604
    management of, 605–606
  esophageal, 721
  "growing", 1056
  in hematologic neoplasms, management of, 219–220
  in intrarenal disorders, 575–576
  ischemic, 966
  neuropathic, 964–966
  perception and integration of, altering of, 968–969, 968*t*, 969*b*
  peripheral transmission of, interruption of, 968
  physiologic responses to, 967
  physiology of, 955–960
    modulation of pain signals in, 956*f*, 958–960, 960*b*, 960*f*–961*f*
    perception of pain signals in, 955–956, 956*f*, 958

Pain *(Continued)*
transduction of stimuli in, 956, 956*f*
transmission of stimuli in, 956–958, 956*f*–957*f*, 956*t*, 959*f*
referred, 966, 966*b*–967*b*, 967*f*
treatment modalities for, 967–969
types of, 960–967
in the young and the elderly, 967
Pain threshold, 958
Pain tolerance, 958
Pallor, in systemic diseases, 1082*b*, 1083
Panacinar emphysema, 490, 491*f*
Pancreas
disorders of, 748–752
embryology of, 744
functional anatomy of, 744, 744*b*, 746*f*
Pancreatic ascites, 750
Pancreatic cancer, 752
Pancreatic duct, 742–743
Pancreatic enzyme replacement, 752
Pancreatic pseudocyst, 750
Pancreaticobiliary system
embryology of, 743, 744*f*
structure and function of, 742–743, 743*f*
Pancreatitis, 748–752, 752*b*
acute, 748–750, 749*b*, 750*f*, 751*b*
chronic, 750–752, 752*f*
Pancytopenia, 277
Pandemic, definition of, 7, 148
Panic disorder, 990–991
Panmyelosis, 286
Papillary muscles, 355, 357*f*
Papule, 1063*f*–1064*f*
Parabolic profile of laminar flow, 320, 320*f*
Paracrine signaling, 49, 50*f*, 784, 784*f*
Paradoxical kinesia, 921
Paraesophageal hernia, 725, 725*f*
Paralytic ileus, 709
Parametritis, 676
Paranasal sinuses, 452–453
Paraneoplastic syndromes, in cancer, 137
Paraphimosis, 644, 644*f*
Paraseptal emphysema, 490
Parasites, pathogenic, 152–153, 155*t*–156*t*, 156*f*
Parasitic infestations, 1075–1076
Parasitic worms, morphology of, 150*f*
Parasympathetic nerves
in autonomic regulation of rhythmicity, 372
distribution of, 867, 869*f*
Parasympathetic nervous system
in gastrointestinal motility, 705
in glucose metabolism regulation, 817–818
Parathyroid gland disorders, 811–812, 811*f*, 812*b*
Parathyroid hormone, regulation and actions of, 811–812, 811*f*
Parkin, Parkinson disease and, 921
Parkinson disease, 920–922
clinical manifestations and treatment of, 921–922, 922*b*, 922*f*
etiology of, 920–921
pathogenesis of, 921, 921*f*
Parkinsonism, 920–921
Parkland formula, for fluid resuscitation in burn shock, 1099, 1100*b*
Paroxysmal focal atrial tachycardia, 425–426, 426*f*
Paroxysmal nocturnal dyspnea, in left-sided heart failure, 417
Partial pressure, of blood gases, 269–270, 273*t*
Partial seizures, 916*b*, 917
Partial-thickness burns, 1096–1097, 1096*f*, 1096*t*
Parturition, 665–667
Passive immunity, 187
Passive transport carriers, 43, 43*f*
Patau syndrome (trisomy 13), 101
Patent ductus arteriosus, 407–408, 407*f*
Pathogen(s)
bacterial, 146*t*–147*t*, 149, 150*f*
causing infections in AIDS patients
characteristics of, 144–145
in disease transmission, 147, 148*f*
teratogenic effects of, 94–95
types of, 149–153
Pathogenesis, in pathophysiology, 2–3, 4*b*
Pathology, definition of, 1
Pathophysiology
definition of, 1
framework for, 2–4
historical background on, 1–2
introduction to, 1–11, 4*b*
Pauciarticular onset, of juvenile idiopathic arthritis, 1056
Pavementing, 172
Peak expiratory flow rate, 460*t*
Pedigree, 102–103, 104*f*
Pelvic floor muscle training, for urinary incontinence, 612
Pelvic inflammatory disease (PID), 676–677, 690–691, 691*b*
spread of, 677*f*
Pelvic support, alterations in, 674–676, 676*b*
Pemphigus, 1072
Pemphigus vulgaris, 1072, 1072*f*
Penile disorders, 641–647, 647*b*
acquired, 643–645
congenital anomalies of, 641–643
erectile dysfunction as, 645, 646*f*
infectious, 645–647
micropenis as, 641–642
neoplastic, 647
Peyronie disease as, 644
phimosis and paraphimosis as, 644, 644*f*
premature ejaculation as, 645, 647*t*
priapism as, 643–644
Penis, 628*f*, 631, 631*f*
erection of, 639
Peptic ulcer disease, 726–728, 727*f*
clinical manifestations and diagnoses of, 728
etiology and pathogenesis of, 726–728, 728*f*–729*f*
treatment of, 728
Perforins, in cell-mediated immunity, 181
Perfusion, pulmonary, 465, 465*f*
Periaqueductal gray (PAG) area, in pain modulation, 958, 960*f*
Pericardial diseases, 402–403, 403*b*
categories of, 402*b*
Pericardial effusion, 402
Pericardial sac, 356, 357*f*
Pericardiocentesis, 402
Pericarditis, 402–403, 402*b*
Pericardium, 356, 357*f*
Periductal mastitis, 682
Perimenopause, 668
Perimysium, 1014, 1014*f*, 1023
Periosteum, 1003, 1003*f*
Peripartal cardiomyopathy, 401*t*
Peripartum cardiomyopathy, 400
Peripheral nerve disorders, 925–932
amyotrophic lateral sclerosis as, 928
Bell palsy as, 932, 932*b*, 933*f*
Guillain-Barré syndrome, 931–932, 932*b*
multiple sclerosis as, 925–926, 926*b*, 926*f*, 927*t*–928*t*
spina bifida as, 927, 928*b*, 929*f*
spinal cord injury as, 928–931, 929*b*, 930*f*, 931*b*, 932*t*
Peripheral nervous system (PNS), 863–867
cranial nerves in, 860, 863–864
origin and distribution of, 864*f*
origins and functions of, 861*t*
spinal nerves in, 861–867, 863*f*, 865*f*, 869*f*
plexus, 864–865, 866*f*–867*f*, 866*t*
Peristalsis, 706–707, 707f
Peritoneal dialysis, in chronic kidney disease management, 606
Peritoneal shunt, for hydrocephalus, 924, 925*f*
Peritonitis, bacterial, spontaneous, 765–766, 766*f*
Permissiveness, in hormone receptor responses, 788
Pernicious anemia, 278, 715
laboratory findings in, 275*t*–276*t*
Peroxisomes, function of, 57*t*
Persistent asthma, in type IV hypersensitivity, 209
Persistent disorders, 978–982
Petechiae, 1082*b*
in coagulation disorders, 301–302, 303*f*
Petite mal seizures, 916
Peyer patches, immune function of, 162
Peyronie disease, 644
pH, of blood, maintaining, renal tubules in, 565
Phagocytosis, 39
bacterial protections against, 145
in inflammatory response, 172–174, 174*f*
Pharynx, functional anatomy of, 699
Phase 1, in cardiac action potential, 370
Phase 3, in cardiac action potential, 370
Phase 4, in cardiac action potential, 370
Phenome-genome system of classification, in cardiomyopathies, 400
Phenotype, definition of, 95
Phenylketonuria (PKU), 106
Pheochromocytoma, 811
hypertension from, 349
Philadelphia chromosome, in chronic myeloid leukemia, 221, 221*f*
Phimosis, 644, 644*f*
Phlebitis, 325
blood flow alterations from, 325
Phlebotomy, for polycythemia vera, 292
Phosphate, plasma, imbalances in, 535–536, 536*b*
Phosphodiester bonds, 78, 78*f*
Phospholipids, of plasma membrane, 27–28, 28*f*–29*f*
Photoreceptors, of retina, 943
Photosensitivity, 1077–1078
Phototherapy, for depression, 981–982
Physical injury, cellular, 72–73
Physiology, definition of, 1
Pia mater, 851, 852*f*
Pickwickian syndrome, 514
Pigmentation alterations, 1080–1081
Pigments, intracellular accumulations of, in cell injury, 60–61
Pineal gland, 859*f*, 860
"Pink puffers", 490
Pinocytosis, 39

Pituitary gland, 788
  anatomy of, 859*f*
  anterior, 859*f*
    endocrine cell types of, 789*t*
    hormones of, 790–793, 791*f*, 793*b*
  posterior, 859*f*
    hormones of, 788–790, 790*f*
Pityriasis rosea, 1071
Pivot joint, 1010, 1010*f*
Placenta
  development of, 663–664
  functions of, 663–664
Placenta previa, 681
Plantar warts, 1065*f*
Plaque, 1063*f*–1064*f*
  atherosclerotic, 327
Plasma
  blood, composition of, 260–261
  composition of, 851–852, 852*t*
Plasma cell myeloma (multiple myeloma), 224–226, 225*f*
Plasma cells
  antibody-producing, 181
  malignant, in plasma cell myeloma, 225
Plasma colloid osmotic pressure, 321
Plasma D-dimer assay, normal value and significance of, 305*t*
Plasma membrane, 27–29, 27*f*, 29*b*
  channel proteins of, 44, 44*f*
  functions of, 39–46, 57*t*
    active transport pumps in, 41–42
    macromolecule transport in, 39–40
    small molecule transport in, 40–44
  lipid bilayer of, 27–28, 28*f*
  permeability barrier of, disruption of, in necrosis, 64, 65*f*
  potentials initiated in, 45–46, 45*f*
    action, 45–46, 46*f*–48*f*
    resting, 45
  proteins of, 29, 30*f*
  structure of, 27, 28*f*
  tight junctions of, 29
  transporters in
    ABC, 42, 43*f*
    calcium, 42, 42*f*
Plasma membrane-bound molecules, in intercellular communication, 48*f*
Plasma proteins, 260–261
Plasma thromboplastin antecedent, action of, 300*t*
Plasma thromboplastin component, action of, 300*t*
Plasticity
  neural, 854, 868–872, 880–881
  synaptic, 990
Plateau phase, in cardiac action potential, 370
Platelet aggregation, normal value and significance of, 305*t*
Platelet count
  alterations in, in hemostatic disorders, 305*t*
  normal value and significance of, 305*t*
Platelet deficiency, in leukemia, management of, 219
Platelet disorders, 306–308
  qualitative, 307–308
  thrombocytopenia as, 306–307
  thrombocytosis as, 307
Platelets, 262, 299, 299*f*–300*f*
  characteristics of, 263*t*
  in hemostasis, 298, 300*t*
Pleasure, diminished, in major depression, 980–981
Plethysmography, in atherosclerosis diagnosis, 330
Pleural effusion, 510–511
Pleural space disorders, 508–511, 511*b*
Plexuses, 864–865, 865*f*–867*f*, 866*t*
Plicae circulares, 701
Pneumoconiosis, lung, occupational, 503, 504*t*
Pneumocystis jiroveci (carinii) pneumonia, in HIV patient, 250, 250*f*
Pneumonia, 514–516, 516*b*
  clinical manifestations of, 515
  diagnosis of, 515–516
  differentiating features of types of, 515*t*
  etiology of, 514
  pathogenesis of, 514–515, 516*f*
  treatment of, 516
Pneumonitis, hypersensitivity, 502, 502*t*
Pneumothorax, 508–510
  clinical manifestations of, 509, 510*f*
  diagnosis of, 509, 510*f*
  etiology of, 508–509
  pathogenesis of, 509, 509*f*
  treatment of, 510
Podocytes, glomerular, 555–557, 557*f*–558*f*
Point mutation(s), 97, 99*f*
  proto-oncogene activity and, 126*f*, 127
Point of maximal impulse (PMI), 355
Poiseuille's law, 318
Poisoning, acetaminophen, 775, 775*f*
Poliomyelitis, respiratory effects of, 511, 512*t*
Polyarteritis nodosa, 1074
Polycystic kidney disease
  autosomal dominant, 578–579, 578*t*
  autosomal recessive, 578
  normal *versus*, 579*f*
Polycythemia, 286–293, 293*b*
  classification of, 274*b*, 286
  definition of, 272
  differentiation of, 292*f*
  laboratory findings in, 275*t*–276*t*
  relative, 293
  secondary, 292–293
Polycythemia vera, 286–292
  clinical manifestations of, 286–292
  course and prognosis of, 292
  etiology and pathogenesis of, 286
  evolution of, 292, 293*f*
  laboratory features of, 275*t*–276*t*, 286, 292*f*
  treatment of, 292
Polygenic traits, 96, 110–111
Polymenorrhea, 673
Polymorphism, in major histocompatibility complex genes, 176
Polymorphonuclear leukocytes (PMNs), 162
Polymyositis, 1038
Polypeptides, as neurotransmitter, 876*b*
Polyps, colon, 738, 738*f*
Polyradiculoneuropathy, 931
Polysaccharides, dietary, 839
Polysomy, 99–100
Pons, 860
  in respiratory center, 462, 464*f*
Populations, disease patterns in, 6–10, 7*f*
Porcelain gallbladder, 748
Port-wine stain hemangiomas, 1087
Portal circulation, 757*f*
Portal hypertension
  manifestations of, 759–763
  pathophysiology of, 755, 761*f*
Portal systemic encephalopathy, 763–764
Portosystemic shunting, for gastroesophageal varices, 762–763, 764*f*
Positive inotropes, 376
Positron emission tomography, in cardiac function, 379
Postherpetic neuralgia, 966, 1065
Postinfectious acute glomerulonephritis, 588
Postmortem autolysis, 75
Postoliguric phase, of acute kidney injury, 600
Postsynaptic neurons, in memory, 888
Posttraumatic stress disorder (PTSD), 994–995, 995*b*
  stress hormones in, 22
Postural hypotension, 350
Postviral myocarditis, 400
Potassium
  plasma, imbalances in, 531–533, 532*b*–533*b*
  secretion of, renal tubules in, 566, 568*f*
Potassium equilibrium potential, 45
Potassium rectification, 370
Potassium-wasting diuretics, 569
Prader-Willi syndrome, 110, 110*f*
Pramlintide, for diabetes, 830
Pre-diabetes, 820, 822*b*
Predictive value, of test, 5
Preductal coarctation, 408
Preeclampsia-eclampsia, 681
Prefrontal cortex, serotonin and, depression and, 979, 980*f*
Pregnancy, 663–668, 668*b*
  early human development in, 663, 664*f*
  fetal membranes and placenta in, 663–664
  hypertension in, 349
  implantation in, 663
  medication administration categories during, 113*t*
  parturition and, 667–668
  weight gain and nutrition during, 668
Pregnancy disorders, 681, 682*b*
Pregnancy-induced hypertension (PIH), 681
Prehypertension, definition of, 343–344
Preload, 337–338, 375, 375*f*
  increased, in heart failure, 413–415, 415*f*
  in shock, 447
Premature atrial complexes (PACs), 425–426, 426*f*
Premature ejaculation, 645, 647*t*
Premature junctional beats, 426
Premature ventricular complexes (PVCs), in ventricular dysrhythmias, 427, 427*f*
Prenatal diagnosis and counseling, 113–114, 115*b*
Preprohormones, 786
Prepuce, 629*f*, 631
Presbycusis, 940, 941*b*
Presbyesophagus, 700, 717*b*
Presbyopia, 944–945
Pressure, affecting blood flow, 319, 319*f*
Pressure-sensitive receptors, 341
Pressure sores, 1078
  clinical description of, 1078*t*
  prevention of, 1078, 1078*b*
Pressure-volume loop, 358, 358*f*
Pressure work, of heart, 376
Presynaptic neurons, in memory, 888
Prevention, levels of, 9–10
Priapism, 643–644
Prickly heat, 1088
Primary biliary cirrhosis, 773
Primary cardiomyopathy, 400

Primary hypertension, 344–348
  end-organ damage in, 345–346, 345*f*
  outcomes of, 345–346
  risk factors for, 344–345, 344*t*
  subtypes of, 344
  treatment interventions for, 346–348, 346*t*, 347*f*
Primary immunodeficiency disorders, 210–212, 210*t*, 213*b*
Primary sclerosing cholangitis, 748, 773
Principal cells (P cells), of collecting duct, 559
Prion, morphology of, 150*f*
Pro-hormones, 786
Proaccelerin, action of, 300*t*
Proconvertin, action of, 300*t*
Prodromal period, 3
Prodromal phase, of acute kidney injury, 599
Prodrome, in seizure disorders, 917
Progesterone, target organs and actions of, 660*t*
Programmed senescence theory, of aging, 74–75
Progressive familial intrahepatic cholestasis, 779
Progressive obliterative cholangiopathy, 780
Prolactin
  secretion and activation of, 792
  in stress and adaptation, 19
Prolapse, mitral valve, 396–397, 396*f*
Proliferative phase, of menstrual cycle, 659–661, 661*f*
Pronephros, 632, 632*f*
Proprioceptors, in control of respiration, 464, 464*f*
Propulsion, 709, 710*f*
Propulsive movements, in gastrointestinal motility, 706–707, 707*f*
Prostaglandins
  in glomerular filtration regulation, 563
  in inflammatory response, 171, 173*f*
  in nociception, 956, 956*f*
Prostate, 628–629, 629*f*
  bacteria targeting, 153*f*
Prostate cancer, 652–655, 654*f*
  screening guidelines for, 121*t*
Prostate hyperplasia, benign, 651–652, 651*f*
Prostatitis, 652, 653*f*
Prosthetic joint infections, 1045
Protease inhibitors, 257
Protein(s)
  absorption of, 715
  amino acids forming, chemical structures of, 85*f*
  digestion of, 714–715, 714*t*
  gene activator, 84, 86*f*
  gene repressor, 84
  membrane, 29, 30*f*
  metabolism of, 841
  plasma, 260–261
  transmembrane, 31*f*
  transport, 31*f*
  trimeric G, 53*t*
Protein-energy malnutrition, metabolism in, 845, 845*f*–846*f*
Protein-energy wasting, complicating chronic kidney disease, 603
Protein kinase receptors, 785, 785*f*
  enzyme-linked, 49–51, 51*f*
Proteinuria, in glomerular basement membrane dysfunction, 557
Proteoglycans, in articular cartilage, 1011
Prothrombin, action of, 300*t*
Prothrombin time
  alterations in, in hemostatic disorders, 305*t*
  normal value and significance of, 305*t*
Proto-oncogenes, 124–127, 124*t*, 129*b*
  activation of
    to become oncogenes, 125–127
    effects on growth signaling pathways, 125*f*
    mechanisms of, 125, 126*f*
  overactivity of, 125, 126*f*
Protozoa, 152–153
  morphology of, 150*f*
Protozoan infections, 155*t*–156*t*
Proximal convoluted tubule, of nephron, 555*f*, 557–558
  epithelial cells of, 556*f*
  functions of, 556*t*
  membrane transporters of, 560*f*
Pruritus, in burn wound healing, 1106–1107
Pseudocyst, pancreatic, 750
Pseudomembranous colitis, 731
Psoriasis, 1069–1071, 1071*f*
Psoriatic arthritis, 1053
Psychomotor agitation
  in bipolar disorder, 983
  in major depression, 980–981
Psychomotor retardation, in major depression, 980–981
Psychosis
  definition of, 972
  lifetime prevalence of, 972*t*
  major depression and, 981
Psychotherapeutic interventions, for depression, 981–982
Psychotic illnesses
  cultural considerations on, 984
  geriatric considerations in, 984, 985*b*
  major depressive and persistent disorders, 978–982, 979*t*
  neurobiology of, 971–988
  schizophrenia as, 972–977
  women and, 984, 985*t*
Ptyalin, 699
Pulmonary artery events, in cardiac cycle, 358*f*, 360
Pulmonary atresia, 408
Pulmonary disorders
  bronchial provocation tests for, 472
  diagnostic tests for, 471–472, 472*b*
  lung cancer as, 475
  obstructive, 478–498
    acute bronchitis as, 483–485
    acute tracheobronchial obstruction as, 495–496
    from airway lumen obstruction, 491–496, 496*b*
    asthma as, 479–482, 479*f*–480*f*, 482*f*–485*f*. *see also* Asthma
    bronchiectasis as, 491–493, 492*f*–494*f*
    bronchiolitis as, 493–494
    chronic bronchitis as, 485–487, 486*f*–487*f*, 488*t*
    from conditions in wall of lumen, 479–487
    croup syndrome as, 496
    cystic fibrosis as, 494–495
    epiglottitis as, 496
    related to loss of lung parenchyma, 488–491
Pulmonary disorders *(Continued)*
  pulmonary function testing in, 471–472, 472*f*
  restrictive, 499–520, 500*t*
    acute respiratory distress syndrome as, 504–506, 505*b*, 506*f*–507*f*
    age-related features contributing to, 501*t*
    amyotrophic lateral sclerosis as, 511, 512*t*
    ankylosing spondylitis and, 512–513
    atelectatic disorders as, 504–508
    from chest wall deformities, 511–513
    diffuse interstitial lung disease as, 500–501, 500*f*
    fibrotic interstitial lung diseases as, 500–504
    flail chest and, 513
    Guillain-Barré syndrome as, 511, 512*t*
    hypersensitivity pneumonitis as, 502, 502*t*
    infant respiratory distress syndrome as, 507–508
    infection/inflammation of lung and, 514–518
    kyphoscoliosis and, 511–512, 513*f*
    lung parenchyma disorders as, 500–504, 503*b*
    Middle East respiratory syndrome as, 516, 517*t*
    muscular dystrophies as, 511, 512*t*
    myasthenia gravis as, 511, 512*t*
    neuromuscular, 511, 512*t*, 514*b*
    obesity and, 513–514
    occupational lung diseases as, 503–504, 503*t*–504*t*
    pleural effusion as, 510–511
    pleural space disorders as, 508–511, 511*b*
    pneumonia as, 514–516, 516*f*
    pneumothorax as, 508–510, 509*f*–510*f*
    poliomyelitis as, 511, 512*t*
    pulmonary tuberculosis as, 516–518, 517*f*–518*f*
    sarcoidosis as, 501–502
    severe acute respiratory syndrome as, 516, 517*t*
Pulmonary edema, acute cardiogenic, in heart failure, 418
Pulmonary embolism, 473–475, 474*b*–475*b*, 474*t*
  in complicating fracture, 1030
  hypoxemia from, laboratory values in, 547*t*
Pulmonary function testing, 471–472, 472*f*
Pulmonary hypertension, 472–473, 473*b*
Pulmonary infarction, in pulmonary embolism, 474
Pulmonary malignancies, 475–476, 476*b*
Pulmonary stenosis, 408
Pulmonary system, 458*b*
  age-related variations of, 457, 459*t*
  blood circulation in, 457, 457*f*
  blood flow in, 465–466, 465*f*
  blood vessels of, 465. *see also* Pulmonary vasculature
  development of, 452, 453*f*
  functional alterations in, 468–470, 470*b*–471*b*. *see also* Respiratory disorders
  functional anatomy of, 452–457
  lower airway structures of, 453–456, 455*f*–457*f*
  upper airway structures of, 452–453, 453*f*–455*f*
  ventilation and, 460–464. *see also* Ventilation
  ventilation-perfusion ratios in, 465–466, 466*b*
Pulmonary vascular disorders, 472–475, 475*b*
  pulmonary hypertension as, 472–473, 473*b*
  pulmonary venous thromboembolism as, 473–475, 474*b*–475*b*, 474*t*

Pulmonary vascular resistance, 337–338
Pulmonary vasculature, 465
  alterations in, 472–475, 475*b*
Pulmonary venous thromboembolism, 473–475, 474*b*–475*b*, 474*t*
Pulmonic valve, 355, 356*f*
Pulse pressure, 338
Pulsus paradoxus, 402
Punnett square, 95–96, 98*f*
Pupil reflex, in acute brain injury, 901–902
Purines, 78, 78*f*
  as neurotransmitter, 876*b*, 879–880
Purkinje cells (fibers), 371
Purpura
  in coagulation disorders, 302, 303*f*
  vascular, 304–306
Purulent exudate, 175
Pustular psoriasis, in type IV hypersensitivity, 209
Pustule, 1063*f*–1064*f*
Pyelography, intravenous, 572–573
Pyelonephritis, 581, 583*b*
  acute, 582, 583*f*
  chronic, 582–583
Pylorus, 700, 701*f*
Pyrimidines, 78, 78*f*
Pyrosis, 721

Q

Q wave, 373–374, 373*f*–374*f*
QRS complex, 373–374, 373*f*–374*f*

R

R wave, 373–374, 373*f*–374*f*
Racial differences, in HIV/AIDS rates, 234
Radiation
  electromagnetic, cell injury from, 73*f*
  etiology of hematologic neoplasms, 217
  teratogenic effects of, 112–113
Radiation therapy
  for cancer, 138
  for lymphoma, 227–229, 229*f*
Radiculopathy, 957–958
Radionuclide studies, renal, 573
Radionuclide voiding cystography, 611
Raphe magnus, 958, 960*f*
Rapid eye movement (REM) sleep, 889
Rapidly progressive glomerulonephritis (RPGN), 589
Rarefaction, of long bones, in thalassemia, 280
Rashes, 1082*b*
Raynaud disease, 1083
Raynaud syndrome, 330
Rb (retinoblastoma) tumor suppressor gene, 127, 128*f*–129*f*
RBCs (red blood cells), 262, 262*f*. *see also* Erythrocytes
Reabsorption, of glucose, in transport across renal tubules, 565, 566*f*
Reactive arthritis, 1052
Reactive hyperemia, 323
Reactive oxygen molecules. *see also* Oxygen radicals (free radicals)
  in ischemic injury, 69
Rebound hypoglycemia, 736
Receptor(s)
  cell surface, 49–53, 50*f*
    in ligand signaling, 49
    responses mediated by, 49–53
Receptor(s) *(Continued)*
  growth factor, in cell proliferation, 124
  hormone
    agonists and antagonists, 788
    down-regulation and up-regulation of, 787–788
    specificity and affinity of, 787
  inotropic, 874, 875*f*
  metabotropic, 874, 875*f*
  sensory, 882–883, 882*f*–883*f*
β_1-Receptor down regulation, in systolic function, 412
Recessive alleles, 95–96, 98*f*
Recessive trait, 95, 98*f*
Recombinant DNA technology, 114–115
Recruitment, in muscle contraction, 1016
Rectocele, 676, 676*f*
Rectum, 703
Reed-Sternberg cells, in Hodgkin disease, 227, 227*f*
Reentry, in dysrhythmias, 422, 422*f*–423*f*
Referred pain, 388, 966, 966*b*–967*b*, 967*f*
Reflexes, spinal, 884–885, 887*f*
Refraction, errors of, 944–945, 945*f*
Refractory septic shock, definition of, 445*t*
Regional enteritis, 730
Regurgitation, 394
  aortic, 397–398, 398*f*
  mitral, 396, 396*f*
Rehabilitation, after burn injury, 1106–1107, 1107*b*
Reiter syndrome, 1052
Relative refractory period, 370
Relaxation, muscle, energy of, 367, 367*b*–368*b*, 369*f*
Reliability, of data, 5
Remission, 3
Renal agenesis, bilateral, 578
Renal calculi, 584–586
Renal cell carcinoma, 579–580
  risk factors for, 580*b*
  staging system for, 581*f*
Renal compensation, in acid-base balance regulation, 566
Renal cortical adenomas, 579*t*
Renal disorders
  acute injury as, 593–600, 600*b*–601*b*
  hypertension and, in children, 348
  laboratory profile for, 597*t*–599*t*
Renal failure
  acute, 593–594
    in shock, 449
  chronic, 601
Renal function, 551–574
  age-related changes in, 570, 570*b*–571*b*
  blood volume and osmolality regulation and, 567–570
  in burn injury, 1102
  endocrine, 570, 570*b*
  glomerular filtration in, 560–564. *see also* Glomerular filtration; Glomerular filtration rate
  nephron in, 554–559. *see also* Nephron
  tests of, 570–573, 573*b*
    diagnostic, 572–573
    urine and blood studies in, 570–572
  tubular transport in, 565–566. *see also* Renal tubules
Renal hypoplasia, congenital, 577–578
Renal parenchyma, 552, 553*f*
Renal system, in acid-base homeostasis, 543, 543*t*, 544*f*
Renal tubules
  in blood volume and osmolality regulation and, 567–570, 568*t*, 569*f*, 570*b*
  collecting, 555*f*–556*f*, 559. *see also* Collecting duct (tubule), of nephron
  distal convoluted, 555*f*–556*f*, 559. *see also* Distal convoluted tubule, of nephron
  necrosis of, acute, 595
  proximal convoluted, 555*f*–556*f*, 557–558. *see also* Proximal convoluted tubule, of nephron
  transport across, 565–566, 565*f*, 566*b*–567*b*
    in acid-base balance regulation, 565–566, 567*f*
    glucose reabsorption and, 565, 566*f*
    potassium secretion and, 566, 568*f*
Renin, production of, juxtaglomerular cells in, 563
Renin-angiotensin-aldosterone system (RAAS)
  in blood pressure regulation, 342–343, 342*f*
  in heart failure, 413–415, 414*f*
  in shock, 437–438
Renin-angiotensin mechanism, of aldosterone secretion regulation, 797*f*
Reninomas, 579*t*
Reperfusion injury, 69, 435–436
  in acute brain injury, 894–895
Reperfusion therapies, for myocardial infarction, 392
Replication, DNA, 79–80, 81*f*–82*f*
Replication origins, 79–80
Repolarization, of membrane, 45
Reproductive system, male, 626–640
Residual volume, 460*t*
Resistance, to blood flow, 318
Resistance stage, of general adaptation syndrome, 14–16, 14*t*
Respiration
  age-related variations of, 457, 459*t*
  alterations in, 468–470, 470*b*–471*b*. *see also* Respiratory disorders
  gas diffusion and transport in, 466–468, 466*f*, 467*t*, 468*b*
  pulmonary anatomy and, 452–457. *see also* Pulmonary system
  pulmonary blood flow and perfusion in, 465–466
  ventilation in, 460–464. *see also* Ventilation
Respiratory acidosis, 546–547, 547*b*
Respiratory alkalosis, 548–549, 548*b*
Respiratory center, 462
Respiratory disorders, 468–470. *see also* Pulmonary disorders
  acute respiratory failure as, 469–470, 471*b*
  hypoventilation and hyperventilation as, 468
  hypoxemia and hypoxia as, 468–469, 470*f*
Respiratory distress syndrome, 508*b*
  acute, 504–506, 505*b*, 506*f*–507*f*
  infant, 507–508
Respiratory function, 451–477
Respiratory membrane, 466–467, 466*f*
Respiratory system
  in acid-base homeostasis, 542–543, 543*t*
  changes, in pregnancy, 668
Resting membrane potential (RMP), 45, 46*b*, 46*f*
Resting potential, cardiac, 369
Restrictive cardiomyopathy (RCM), 400*f*, 401, 401*t*

Restrictive pulmonary disorders, 499–520. *see also* Pulmonary disorders, restrictive
Resuscitation, acute, for burn shock, 1099–1101, 1100*b*
Retardation, psychomotor, in major depression, 980–981
Reticulocyte, 265, 265*f*
Reticulocytosis, in hemolytic disease of newborn, 284
Retina, 943
Retinal detachment, 947, 947*f*
  visual field defects in, 951*f*
Retinitis, CMV, in HIV infection, 253
Retinopathy, 947–948
  diabetic, 826
  HIV-associated, 253
Retrovirus, in proto-oncogene activation, 125, 126*f*
Reye syndrome, 779
Rheumatic disorders, 1042–1057. *see also* Joint function disorders
Rheumatic fever, acute, 1052, 1053*b*
Rheumatic heart disease, 398
Rheumatism, nonarticular, 1056
Rheumatoid arthritis (RA), 1046–1050
  clinical manifestations of, 1046–1048, 1048*f*–1049*f*
  diagnosis of, 1048–1049, 1049*t*
  etiology and pathogenesis of, 1046, 1047*f*
  treatment of, 1049–1050
Rheumatologic manifestations, of HIV infection, 254
Rhus dermatitis, 1073–1074, 1074*f*
Rhythmicity
  autonomic regulation of, 372
  of myocardial cells, 370–371, 371*f*
Ribosomes, 31–32, 33*f*, 82–83, 84*f*
Rickets, 1035
Rifle classification, for acute kidney injury staging, 594*t*
Right bundle branch block, 429, 430*f*
Right-sided heart failure, 418–420, 419*f*–420*f*
Ring chromosome, 100, 101*f*
Risk factor(s)
  definition of, 2
  stressors and, 16–17
RNA (ribonucleic acid)
  HIV, plasma level of, in monitoring HIV disease status, 246
  messenger (mRNA), synthesis of, transcription in, 81, 83*f*
  types of, 81, 81*b*
RNA virus, 150–151, 154*t*
Robertsonian translocation, 100, 101*f*
Rocky Mountain spotted fever (RMSF), 1076
Roseola infantum, 1089
Rostral pons, pain modulation and, 958, 960*f*
Rough endoplasmic reticulum, 31–32
Rubella, 1089
Ruffini endings, 882*f*
Ryanodine receptors, 363, 364*f*

S

S wave, 373–374, 373*f*–374*f*
Saccular aneurysms, 330–331, 331*f*
Sacral plexus, 864–865, 865*f*–867*f*, 866*t*
Saddle joint, 1010
Saline deficit, 525
Saline excess, 526
Saline imbalances, 525
Salivary amylase, 699
Salivary glands, 699, 699*f*
Salpingitis, 690–691, 691*b*
Saltatory conduction, 874
Sarcoidosis, 501–502
Sarcolemma, in skeletal muscle, 1014
Sarcoma
  Ewing, 1037*f*, 1038
  Kaposi, in HIV infection, 251, 251*f*, 1079
Sarcomeres, 363, 365*f*, 1015, 1015*f*
Sarcoplasm, 1014
Sarcoplasmic reticulum (SR), 363, 364*f*
  in muscle contraction, 1016
Scabies, 1075, 1075*f*
Scale, 1063*f*–1064*f*
Scar, 1063*f*–1064*f*
Scarlet fever, 1090
Schistocytes, in drug-induced immune hemolysis, 285, 285*f*
Schizophrenia, 972–977
  in elderly, 984
  etiology and neurobiology of, 972–976
    dopamine effects in, 972–973, 973*f*
    genetic effects in, 974
    gestational effects in, 974, 974*f*
    neurologic effects in, 975–976, 975*f*–976*f*
  negative symptoms of, 976
  positive symptoms of, 976
  subtypes of, 976
  treatment of
    nonpharmacologic, 977, 978*b*
    pharmacologic, 976–977, 977*f*, 978*t*
Sclerodactyly, 1051
Scleroderma, 1050–1051, 1076–1077
  diffuse, 1077, 1077*f*
  localized, 1077, 1077*f*
Sclerosing cholangitis, primary, 748, 773
Scoliosis, 1032–1033, 1033*b*, 1033*f*
Scotoma, 950
Scrotal disorders, 647–650, 651*b*
  acquired, 648–650
  hydrocele as, 648, 648*f*
  infectious, 650
  spermatocele as, 649
Scrotum, 629, 629*f*
Sebaceous glands
  age-related changes in, 1060*b*, 1061
  hyperfunction of, in systemic diseases, 1084
Seborrheic dermatitis, 1069, 1070*f*
Second-degree atrioventricular block, 428, 428*f*–429*f*
Second-degree burns, 1096–1097, 1096*f*, 1096*t*
Second messengers, in G-protein-coupled receptor signaling, 784–785
Secondary active transport, 41
Secondary cardiomyopathy, 400
Secondary hypertension, 348–349, 348*b*–349*b*
Secondary immunodeficiency disorders, 212–213, 213*b*
Secretin, 712, 712*t*
  gastrointestinal motility and, 706
Secretory diarrhea, 723–724
Segmentation contractions, in small intestine, 709, 711*f*
Seizure disorder, 916–917, 916*b*–918*b*
Selectin and chemokine receptors, 172
Selective serotonin reuptake inhibitors (SSRIs), for major depression, 981
Selenium deficiency syndromes, function and, 71*t*
Sellar joint, 1010
Semilunar valves, 355
Seminal vesicles, 629, 629*f*
Seminiferous tubules, 630, 630*f*
Sengstaken-Blakemore tube, 763*f*
Senile keratoses, 1091*f*
Senile lentigines, 1091*f*
Senile purpura, 1090
Sensitivity, of test, 5
Sensorineural hearing impairment, 939–940
Sensory deficits, after stroke, 908
Sensory dermatomes, 957–958, 959*f*
Sensory function
  alterations in, 936–954
  hearing and balance as, 937–942
  smell and taste as, 952–953
  vision as, 942–952
Sensory receptors, 882–883, 882*f*–883*f*
Sepsis
  definitions related to, 445*t*
  nutritional considerations in, 846–848
Septic arthritis, 1045
Septic shock, 444–447
  clinical manifestations of, 445–446
  definition of, 445*t*
  etiology of, 444
  immune cytokines in, 436–437
  pathogenesis of, 444–445, 446*f*
  treatment of, 446–447
Septicemia, 444
Septum primum, 403
Sequela, 3
Seroconversion, in HIV infection, 244
Serosanguineous drainage, 175
Serotonin
  depression and, 979, 980*f*
  as neurotransmitter, 876*b*, 876*t*, 877, 878*f*
  pain modulation and, 958
  schizophrenia and, 978*t*
Serotonin receptor agonists, for migraine, 964*t*
Serotonin syndrome, 981
Serous exudate, 175
Sertoli cells, 630
  in spermatogenesis, 637
Serum albumin, 261
Serum amylase activity, increased, causes of, 751*b*
Serum globulin, 261
Severe acute respiratory syndrome (SARS), 516, 517*t*
Severe combined immunodeficiency disorders, 210–211
Sex chromosome disorders, 101–102
Sex hormones, in stress and adaptation, 19
Sex-linked disorders, 106–107, 107*t*, 108*f*
Sexual maturity ratings (SMRs), five Tanner stages of, 660*f*
Sexually transmitted infections, 689–696, 690*t*
  enteric, 695
  epidemic of, 689
  with localized lesions, 694–695, 695*b*
  with nonulcerative lesions, 694–695
  organisms causing, 690*b*
  with systemic involvement, 691–693, 693*b*
  with ulcerative lesions, 694
  urethritis, cervicitis, salpingitis and pelvic inflammatory disease as, 690–691, 691*b*
Shared epitope, 1046
Shingles, 1065, 1067*f*
Shock, 434–450
  anaphylactic, 443–444, 443*b*
  arterial oxygen content in, 447–448

Shock *(Continued)*
assessment of, 447–449, 449*b*
blood flow in, distribution of, 448
burn, 1099–1101, 1100*f*
cardiac output in, 447
cardiogenic, 439–441, 441*b*
clinical findings in different types of, 435*t*
compensatory mechanisms of, 437–439, 438*f*–439*f*
complications of, 449–450, 450*b*
distributive, 443–447, 447*b*
etiology of, 435*b*
hemodynamic monitoring in, 448–449, 448*f*, 448*t*
hemorrhagic, 442*t*
hypovolemic, 434–435, 442–443, 442*f*, 443*b*
impaired tissue oxygenation in, 435–437, 436*f*–437*f*, 437*t*
neurogenic, 444, 931
obstructive, 441, 441*b*
pathogenesis of, 434–439, 439*b*
septic, 444–447, 445*t*, 446*f*
spinal, 931
stages of, 437–439, 438*f*
early, 438, 438*f*
progressive, 438–439
refractory, 439
types of, 439–447
Shock-wave, lithotripsy, for renal calculi, 585–586
Short-bowel syndrome, 736–737
Short-term memory, 888
Shunt
LeVeen and Denver, for ascites, 765, 766*f*
portosystemic, for gastroesophageal varices, 762–763, 764*f*
Sick cell syndrome, in burns, 1102–1103
Sick sinus syndrome, 424
Sickle cell anemia, 281–282, 281*f*
clinical manifestations of, 281, 283*t*
complications of, 282*b*
course and prognosis of, 282
etiology and pathogenesis of, 281, 281*f*
laboratory findings in, 275*t*–276*t*, 281, 283*f*
treatment of, 281–282
vasoocclusive consequences of, 283*t*
Signaling pathways
in cancer development, 139
in cell proliferation, 124–125, 125*f*
Silent period, 3
Simple fracture, 1027
Single-gene disorders
Mendelian, 98–99, 102–107
nonmendelian, 107–110
Single-gene traits, 95–96
Sinoatrial (SA) node, 371
Sinus arrest, in dysrhythmias, 424, 425*f*
Sinus arrhythmia, in dysrhythmias, 424, 425*f*
Sinus bradycardia, in dysrhythmias, 424, 424*f*
Sinus rhythm, in dysrhythmias
abnormal rates of, 424, 424*b*
normal, 423–424, 423*t*, 424*f*
Sinus tachycardia, in dysrhythmias, 424, 424*f*
Sinuses, paranasal, 452–453
Situational differences, in normality, 6
Skeletal muscle(s), 90, 91*f*
composition of, 1014–1015, 1014*f*
contractile apparatus of, 1014–1015, 1015*b*, 1015*f*
contraction of, 1015–1017. *see also* Muscle contraction
Skeletal muscle(s) *(Continued)*
diseases of, 1038–1039
idiopathic inflammatory myopathy as, 1038
structure and function of, 1013–1015
Skin
age-related changes in, 1059–1061, 1060*b*, 1060*f*
bacteria targeting, 153*f*
contractures of, complicating burn wound healing, 1107
dark
special characteristics of, 1081, 1081*b*
tips for assessing, 1082*b*
effects of sunlight on, 1077–1078
manifestations of systemic diseases in, 1083–1085, 1086*b*
thermal burn effects on, 1094
Skin cancer, 1063–1064
Skin disorders
in adolescence and young adulthood, 1090
allergic, 1072–1074
in childhood, 1088–1090, 1088*f*
developmental considerations for, 1087–1090
geriatric considerations of, 1090, 1091*f*, 1092*b*
in infancy, 1087–1088, 1088*f*
infectious, 1064–1069
bacterial, 1067–1069
fungal, 1065–1067
viral, 1064–1065
inflammatory, 1069–1072
parasitic infestations as, 1075–1076
selected, 1063–1064
treatment of
corticosteroids in, 1086–1087
delivery system selection for, 1087, 1087*b*
intralesional, 1086–1087
topical, 1086
Skin lesions. *see also* Skin disorders
descriptors for, 1062, 1065*t*
primary and secondary, 1061, 1062*b*, 1063*f*–1064*f*
Skin mesher, in burn wound management, 1105
Skin substitutes, in burn wound management, 1105
Sleep, 888–890, 889*f*, 890*b*
Sliding filament theory, of muscle contraction, 90, 366–367, 368*f*, 1015
Sliding hernia, 725, 725*f*
Slow channels, 370
Slow twitch, 1017
Small cell carcinoma, pulmonary, 476
Small intestines
functional anatomy of, 701–703, 702*f*–703*f*
gastrointestinal motility of, 709–710, 710*f*–711*f*
inflammation of, 726–728
neoplasms of, 737–738
Small nuclear ribonucleoproteins, 82, 86*b*
Smell, sense of, 952–953
disorders of, 952, 953*b*
Smoke inhalation, in burn injury, 1102
Smoking
atherosclerosis risk and, 327–329
cancer risk and, 120, 122*f*–123*f*
cessation of, in chronic bronchitis management, 487
Smooth endoplasmic reticulum, 31–32
Smooth muscle, 90, 91*f*
gastrointestinal
contraction of, 706
electrical activity of, 705–706, 706*f*
Socioeconomic factors, in epidemiology, 8–9, 8*f*
Sodium intake, elevations of, in blood pressure regulation, 342
Sodium ion-driven carriers, 42–43
Sodium-potassium ion pump, 41–42, 42*f*
in ion-gradient maintenance, 42
Soft tissue healing, after trauma, 1024–1025, 1024*f*–1025*f*
Soft tissue injuries, 1021–1025, 1025*b*
bursae injuries, 1023
contractile, 1023–1025, 1024*f*–1025*f*
fasciae injuries, 1023
inert, 1021–1023
internal joint derangement as, 1023
joint capsule injuries as, 1022–1023, 1022*f*
ligament injuries as, 1021, 1022*f*
nerve, nerve roots, or dura mater injuries, 1023, 1023*f*
Solar elastosis, 1060
Somatic death, 75, 75*b*
Somatosensory cortex, 883, 885*f*
Somatosensory receptors, 882*f*
Spasticity
after stroke, 908
in cerebral palsy, 922
Spatial summation, 874, 876*f*
Specificity
receptor, 787
of test, 5
Spermatocele, 649
Spermatogenesis, 637, 638*f*
Spermatozoa
anatomy of, 637
transport of, 637–640
acrosome reaction in, 639–640
capacitation in, 639
erection, emission, and ejaculation in, 639
Spherocytes, 282
Spherocytosis, hereditary, 282
laboratory findings in, 275*t*–276*t*
Sphincter(s)
ileocecal, 710
internal, 610
lower esophageal, 700
Sphygmomanometer, in blood pressure, 339–340
Spina bifida, 927, 928*b*, 929*f*
Spina bifida cystica, 927
Spina bifida occulta, 927
Spinal accessory nerve (CN XI), 861*t*, 864, 864*f*
Spinal cord, 861–863, 861*f*
cross-section of, 861, 862*f*
modulating pain transmission at, 968
nerves of, 861–863, 863*f*
tracts of, 861, 862*f*
Spinal cord and peripheral nerve disorders, 925–932
amyotrophic lateral sclerosis as, 928
Bell palsy as, 932, 932*b*, 933*f*
Guillain-Barré syndrome, 931–932, 932*b*
multiple sclerosis as, 925–926, 926*b*, 926*f*, 927*t*–928*t*
spina bifida as, 927, 928*b*, 929*f*
spinal cord injury as, 928–931, 929*b*, 930*f*, 931*b*, 932*t*
Spinal cord injury, 928–931, 929*b*, 930*f*, 931*b*, 932*t*
Spinal meninges, 851, 853*f*
Spinal nerves, 861–867, 863*f*, 865*f*, 869*f*
plexus, 864–865, 866*f*–867*f*, 866*t*
Spinal reflexes, 884–885, 887*f*

Spinal shock, 931
Spinothalamic tract, pain transmission and, 957
Spiral bacteria, 149
Spiral fracture, 1026, 1026*f*
Spirochetes, 149, 151*f*
Spirometry, 471, 472*f*
Spleen, immune function of, 161
Spliceosomes, 82
Splinter hemorrhages, 1084, 1086*f*
Spontaneous bacterial peritonitis, 765–766, 766*f*
Sprue, tropical, 735
Squamous cell carcinoma, of skin, 1079–1080, 1080*f*
Squamous cell epidermoid carcinoma, pulmonary, 475
Stable angina, 386, 388, 388*f*
Starling's hypothesis, 321
Starvation, metabolism in, 845, 845*f*–846*f*
Starvation ketoacidosis, 545
Statistical normality, 4–5, 4*f*
Status asthmaticus, treatment of, 481–482
Status epilepticus, 917
Stem cell transplantation
  in cancer therapy, 139–140
  for sickle cell anemia, 282
Stem cells
  neural, 868
    proliferation, 868, 874*f*
  tissue, malignant potential of, 119
Stenosis, 394
  aortic, 395*f*, 397, 397*f*, 408
  mitral, 395–396, 395*f*
  pulmonary, 408
Steroid hormones, 794–796, 796*b*
  actions of, on target cells, 795–796, 796*b*
  synthesis and secretion of, 794–795, 795*f*–797*f*
Stethoscope, in blood pressure, 339
Stevens-Johnson Syndrome, in type IV hypersensitivity, 209
Still disease, adult-onset, 1055–1056, 1056*b*
Stomach
  functional anatomy of, 700–701, 701*f*–702*f*
  inflammation of, 726–728
  motor functions of, 708–709
Stomatitis, 724
Stop codons, 80
Strabismus, 945
Strain, muscle, 1024
Strangulation, 733
Stratified epithelium, 87–89
Strawberry hemangiomas, 1087
Stress
  autoimmunity triggered by, 197
  as a concept, 13–17, 17*b*
  energy metabolism during, 818–820
  management of, in diabetes mellitus treatment, 831
  neurohormonal mediators of, 17–19. *see also* Neurohormonal mediators, of stress and adaptation
  physical and behavioral indicators of, 21*b*
  physiologic, atherosclerosis risk and, 329
Stress fracture, 1026, 1026*f*
Stress hormones, 22
Stress-induced analgesia, 959–960
Stress proteins, in cell injury, 60, 62*f*
Stress response, glucocorticoids effects on, 18*t*
Stress response pathways, of brain, overbuilt, panic disorder and, 990
Stress urinary incontinence (SUI), 612
Stressor(s)
  definition of, 13
  desensitization to, 20–21
  habituation to, 20–21
  response to, 15*f*
  risk factors and, 16–17
Stretch reflex, 885, 887*f*
Striation, 363, 1014
Stroke, 905–909
  hypertension and, 346
Stroke volume (SV), 360
  determinants of, 375–376
Stuart-Prower factor, action of, 300*t*
Subacute infective endocarditis, 398
Subarachnoid hemorrhage, in traumatic brain injury, 904*f*, 905
Subclinical stage, of disease, 3
Subcutaneous tissue, age-related changes in, 1060–1061
Subdural hematoma, in traumatic brain injury, 904, 904*f*
Subfalcine hernia, 897–899, 899*f*
Subluxation, 1030
Substantia gelatinosa, pain transmission and, 957, 957*f*
Substrate, utilization of, in cardiac energy metabolism, 369
Subthalamic nucleus, 860
Sudden cardiac arrest (SCD), 393
Sulfonylureas, for diabetes, 829*t*, 830
Summation, 874, 876*f*
  in response to repetitive stimulation, 1016
Sunburn, 1077–1078
Superficial fungal infections, 1065, 1068*f*
Superficial partial-thickness burn, 1096–1097, 1096*f*, 1096*t*
Supersaturation, urinary, in stone formation, 584
Suprachiasmatic nuclei, 343
Supraventricular tachycardia, 426, 427*f*
Surgery
  in burn, 1103–1104
  nutritional considerations in, 848
Swallowing, 707–708
  neural control of, 708, 708*f*
Swan-neck deformity, in rheumatoid arthritis, 1047, 1048*f*
Sweat glands, age-related changes in, 1060*b*, 1061
Swimmer's itch, 1076
Sympathetic nerves
  in autonomic regulation of rhythmicity, 372
  distribution of, 867, 870*f*
Sympathetic nervous system (SNS)
  activation of, in heart failure, 413, 414*f*
  in blood pressure regulation, 341
  in gastrointestinal motility, 705
  in glucose metabolism regulation, 817–818
  stimulation of, in shock, 437
  supplying kidneys, 553
Symphysis joint, 1007, 1009*f*
Synaptic plasticity, 990
Synaptic signaling, 49, 50*f*
Synaptic transmission, in neuronal communication, 874, 875*f*–876*f*
Synarthroses, 1007
  cartilaginous, 1007, 1009*f*
  fibrous, 1007, 1008*f*
Synchondrosis joint, 1007
Syncytium
  functional, GI tract as, 704
  T cell, in HIV infection, 243
Syndesmosis joint, 1007
Syndrome, definition of, 3
Syndrome of inappropriate antidiuretic hormone secretion (SIADH), 813–814, 813*b*, 813*f*
Synostosis, 1006–1007
Synovial fluid, 1010
Synovial joint, 1009–1010, 1009*f*
Synoviocytes, 1010
α-Synuclein gene, Parkinson disease and, 921
Syphilis, 691–692
  etiology of, 691
  pathogenesis of, 691–692, 691*f*–692*f*
  skin lesions in, 1069
  treatment of, 692
Systemic blood pressure
  determinants of, 337–338
  normal fluctuations in, 343
Systemic inflammatory response syndrome (SIRS), 444
  definition of, 445*t*
Systemic lupus erythematosus (SLE), 206–207, 1050
Systemic sclerosis, 1050
Systemic steroids, for skin disorders, 1087
Systemic vascular resistance (SVR), 337–338
Systole, 358
Systolic blood pressure, 338, 350, 360
Systolic dysfunction, with low ejection fraction, 412

T

T-cell lymphoma, 229, 230*t*, 231*b*
T-cell receptors, in cell-mediated immunity, 178–181, 181*f*
T helper cells, 166–167, 167*f*, 180*f*
T lymphocytes (T cells), 159, 166–167, 167*f*
  cytotoxic, 167, 167*f*, 181
  disorders of, 210–212
  theories of autoimmunity involving, 196
T wave, 373–374, 373*f*
Tachycardia
  atrial, paroxysmal focal, 425–426, 426*f*
  junctional, 426, 427*f*
  sinus, 424, 424*f*
  supraventricular, 426, 427*f*
  ventricular, 427, 427*f*
Tactile hair, 882*f*
Tamm-Horsfall protein, 572
Tamponade, cardiac, 402
Tar burns, 1110
Taste, sense of, 952–953
  disorders of, 952, 953*b*
Tectorial membrane, 937, 938*f*
Telangiectasia
  in coagulation disorders, 303, 304*f*
  hemorrhagic, hereditary, 304*f*, 306
Telomerase
  loss of expression of, in aging, 74–75
  stress and, 23
Telomeres
  shortening of, with age, 74–75, 74*f*
  stress and, 23
Temperature change, in muscle contraction, 1017
Temporal summation, 874, 876*f*
Tendon
  injury to, 1023–1024
  structure and function of, 1012–1013, 1012*f*–1013*f*
Tendon strains, 1024
Tension pneumothorax, 508–509

Tentorial herniation, of brain, 899
Teratogens, 111–113, 113*t*
  definition of, 111
Testicular disorders, 647–650, 651*b*
  acquired, 648–650
  congenital, 647–648
    cryptorchidism as, 647–648, 648*f*
  hypogonadism as, 648, 648*b*
  infectious, 650
  male infertility as, 649–650, 649*b*
  neoplastic, 650, 650*f*
  testicular torsion as, 649, 649*f*
Testicular torsion, 649, 649*f*
Testis(es), 629–630, 630*f*
  bacteria targeting, 153*f*
  development of, 633
Testosterone, in stress and adaptation, 19
Tetralogy of Fallot, 408, 408*f*
Thalamus, 856–857, 859*f*
Thalassemia, 279–281, 280*f*
  laboratory findings in, 275*t*–276*t*
Thalidomide
  for HIV infection, 249
  teratogenicity of, 111
Thermal injury, 1094–1107. *see also* Burn injury(ies)
Thermoregulation, in newborns, 840*b*
Thiazide-like diuretics, 569, 569*t*
Thiazolidinedione drugs, for diabetes, 830
Thick filaments, 363, 365*f*
Thin filaments, 363, 365*f*
Third-degree atrioventricular block, 428–429, 429*f*
Third-degree burns, 1096*t*, 1097
Thoracic pump, 322
Threshold, pain, 958
Thrill, 320
Throat, bacteria targeting, 153*f*
Thrombasthenia, 307
Thrombin time
  alterations in, in hemostatic disorders, 305*t*
  normal value and significance of, 305*t*
Thromboangiitis obliterans (Buerger disease), 330
Thrombocytes, 89–90, 299
Thrombocytopenia, 306–307, 308*b*
  in cancer, 137
  in hematologic neoplasms, 217–218
  in leukemia management, 219
  in liver disease, 310–311
Thrombocytosis, 307
Thromboembolus, 324
  pulmonary venous, 473–475, 474*b*–475*b*, 474*t*
Thrombophlebitis, 324
Thrombosis, 323–324
  arterial, 324
  deep vein, 333
  etiology of, 323–324
  risk factors associated with, 324*b*
  venous, 324
Thrombus, 323–324
Thrush, 1067
Thymus gland, defects in, autoimmunity from, 195
Thyroglobulin, 786–787
Thyroid gland, hormone synthesis in, 793, 794*f*
Thyroid hormone(s), 793–794, 794*b*
  actions of, on target cells, 793–794, 795*f*
  physiologic actions of, 795*t*
  synthesis and secretion of, 793, 794*f*
Thyroid hormone disorders, 803–805, 803*f*–805*f*, 804*b*–806*b*, 804*t*
Thyroid-stimulating hormone (TSH), secretion of, 793
Thyroid storm, 805
Thyroxine (T_4)
  structures of, 795*f*
  synthesis and secretion of, 793
Ticks, 1076
Tidal volume, 460*t*
Tight junctions, of plasma membrane, 29
Time variations, in test data, 6, 6*f*
Tinea, 1065, 1068*f*
Tinea pedis, 152
Tissue(s)
  differentiated, 87–92, 88*t*
  differentiation of, 86–92
    cell diversification in, 86
    cell memory in, 86
    mechanisms of development in, 86–87
  stem cells in, malignant potential of, 119
Tissue oxygenation, impaired, in shock, 435–437, 436*f*–437*f*, 437*t*
Tissue pressure hypothesis, of autoregulation, 323
Tissue thromboplastin, action of, 300*t*
Titin, 363
Tolerance
  pain, 958
  self, failure of, in autoimmunity, 195
Toll-like receptors, 196
Tomography
  computed
    in cardiac function, 378
    renal, 573
  positron emission, in cardiac function, 379
Tonic-clonic seizures, 916–917
Tonsillar herniation, of brain, 898*f*, 899
Tonsils, immune function of, 161
Tophaceous gout, chronic, 1055
Topical skin treatments, 1086
Topical steroids, for skin disorders, 1087
TORCH syndrome, congenital disorders in, 112, 113*f*
Total body surface area (TBSA), in burn assessment, 1097, 1098*f*, 1099*t*
Total body water, 522, 523*f*
Total lung capacity, 460*t*
Tourniquet test, normal value and significance of, 305*t*
Toxemia, 681
Toxic epidermal necrolysis, 209
Toxic liver disorders, 774–775
Toxic metabolic agents, 775
Toxic reactions, in cardiomyopathies, 401*t*
Toxins, microbial, 144–145
Trabecular bone, 1003
Trace elements and deficiency syndromes, functions and, 71*t*
Trachea, 454, 456*f*
  shape variations of, 454, 457*f*
Tracheobronchial obstruction, acute, 495–496
Tracheobronchial tree, 453
Tracheoesophageal fistulas, 699
  types of, 453*f*
Traction alopecia, 1081, 1082*f*
Traits
  codominant expression of, 96
  dominant and recessive, 95–96, 98*f*
  multifactorial, 96, 110–111
Traits *(Continued)*
  polygenic, 96, 110–111
  sickle cell, 281
  single-gene (monogenic), 95–96
Transcellular fluids, 522
Transcription, 81–82, 83*f*
  controls on, in genome regulation, 84, 86*f*
Transcription factors, activation of, in cell proliferation, 125, 125*f*
Transfusion reactions, 202–203, 202*f*, 286, 290*t*
Transfusion therapy
  for anemias, 286, 287*t*–289*t*
  blood components for, 287*t*–289*t*
Transient hypogammaglobulinemia, 212
Transient ischemic attack (TIA), 906
Transjugular intrahepatic portosystemic shunting (TIPS), 762–763
Translation, 82–83, 83*f*–85*f*
Translocations, chromosomal, 100, 101*f*
Transmembrane proteins, 29, 30*f*–31*f*
Transmission, of infection, 147–149, 148*f*
Transplantation
  bone marrow
    for aplastic anemia, 277
    for leukemias, 219
  kidney, for chronic kidney disease, 606–607
  liver, 776–778, 777*b*
    evaluation of, 777
    management after, 777–778
  stem cell, for sickle cell anemia, 282
Transport proteins, in plasma membrane, 29, 31*f*, 41
Transporters associated with antigen processing, 177, 177*f*
Transposition, of great arteries, 408–409, 409*f*
Transverse fracture, 1026, 1026*f*
Trauma
  liver, 776
  nutritional considerations in, 848
  posttraumatic stress disorder and, 994
Traumatic alopecia, 1081*b*
Traumatic brain injury, 902–905
  epidemiology, 902–903
  primary, 903–905
  secondary, 905
  treatment of, 905, 905*b*
  types of, 903
Tremors, in Parkinson disease, 921, 922*f*
Triaxial joints, range of motion of, 1010, 1011*f*
Tricarboxylic acid cycle, 36, 38*f*
Tricuspid atresia, 409
Tricuspid valve, 355, 356*f*
Tricyclic antidepressants (TCAs), 981
Trifascicular block, 430–431
Trigeminal nerve (CN V), 861*t*, 863, 864*f*
Trigeminal neuralgia, 965
Triggered activity, in dysrhythmias, 421–422, 422*f*
Trigone, 627, 627*f*
Triiodothyronine (T_3)
  structures of, 795*f*
  synthesis and secretion of, 793
Triplet repeat mutations, 107
Trisomy 13 (Patau syndrome), 101
Trisomy 18 (Edwards syndrome), 101
Trisomy 21 (Down syndrome), 100–101, 102*f*, 102*t*
Trochlear nerve (CN IV), 860, 861*t*, 863, 864*f*
Trochoid joint, 1010, 1010*f*
Trophoblast, 663
Tropical sprue, 735

Tropomyosins, 366
Truncus arteriosus, 409, 409*f*
Tubercles, 61–62
Tuberculin-type hypersensitivity, 207
Tuberculosis (TB)
bone and joint, 1032
pulmonary, 516–518, 517*f*–518*f*, 518*b*
risk factors for, 143–144
Tubular backleak, 596
Tubules, seminiferous, 630, 630*f*
Tubuloglomerular feedback, in glomerular filtration regulation, 563, 563*f*–564*f*
Tumor(s), 1063*f*–1064*f*, 1079
benign, characteristics of, 118, 118*t*
bone, 1036–1038, 1037*f*, 1038*b*
malignant
angiogenesis in, 133
characteristics of, 118, 118*t*
grading and staging of, 134–135, 135*t*, 136*f*
surgical removal of, 137
terminology related to, 118, 119*t*
Tumor markers, 133, 135*t*
Tumor suppressor genes, 123–124, 127–129, 127*t*, 128*f*–130*f*
Tunica albuginea, 629–630, 631*f*
Turner syndrome, 102, 103*f*
Twitch contraction, 1016–1017
Type I anaphylactic reactions, 443
Type I hypersensitivity, 199–201, 209*b*
clinical manifestations of, 200
etiology of, 199
pathogenesis of, 199–200, 200*f*
prevention of, 201
treatment of, 200–201
Type I interferon, 165
Type II hypersensitivity, 201–204, 209*b*
etiology of, 201–202
in Graves disease, 204
in hemolytic disease of newborn, 203–204
in hyperacute graft rejection, 204
in myasthenia gravis, 204, 204*f*
pathogenesis of, 201–202, 202*f*
in transfusion reaction, 202–203, 203*t*
Type III hypersensitivity, 204, 209*b*
diseases associated with, 205*t*
etiology of, 204–205
granulomatous disease associated with, 207*t*
immune complex glomerulonephritis, 206
pathogenesis of, 205–206, 208*f*
in systemic lupus erythematosus, 206–207
tissue deposition in, 206
Type IV hypersensitivity, 207–209, 209*b*
allergic contact, 207–209
granulomatous, 207
tuberculin-type, 207
Typhlitis, 731

U

U wave, 373–374
Ubiquitin-proteasome pathway
in cell injury, 60, 62*f*
in cellular atrophy, 63
Ulcer(s), 1063*f*–1064*f*, 1078
arterial, 324, 327*f*
decubitus, 1078
clinical description of, 1078*t*
prevention of, 1078, 1078*b*
peptic, 726–728, 727*f*
Ulcerative colitis, 729–730, 729*f*
enteropathic arthritis secondary to, 1053
Ulcerative lesions, sexually transmitted infections and, 694
Ultrasonography
lower urinary tract, 611
in prenatal diagnosis, 114
renal, 573
Uncal herniation, of brain, 899
Unclassified cardiomyopathy, 401*t*
Uncompensated metabolic acidosis, 545
Uncompensated metabolic alkalosis, 548
laboratory values in, 547*t*
Uncompensated respiratory acidosis, 547
laboratory values in, 547*t*
Uncompensated respiratory alkalosis, 548
laboratory values in, 547*t*
Uniaxial joint, range of movement of, 1010, 1010*f*
Unilateral renal agenesis (URA), 578
Unipennate muscle, 1014
Unipolar neurons, 868, 871*f*
United States Centers for Disease Control and Prevention (CDC), 144
Up-regulation, of hormone receptor, 787–788
Upper airway structures, 452–453, 453*f*–455*f*
Uremic syndrome
complicating chronic kidney disease, 603
pathophysiology of, 600*t*
Ureteral ectopy, 616
Ureterocele, 616–617
Ureteropelvic junction obstruction (UPJO), 616
Ureters, male, 627
Urethra
bacteria targeting, 153*f*
male, 628, 628*f*
Urethral disorders, male, 641–647, 647*b*
acquired, 648–650
congenital anomalies of, 641–643
epispadias as, 643, 644*f*
hypospadias as, 642, 643*f*–644*f*
urethral strictures as, 644–645
urethral valves as, 642, 642*f*
urethrorectal and vesicourethral fistulas as, 642
Urethral pressure profilometry, 611
Urethral strictures, 644–645
Urethral valves, 642, 642*f*
Urethritis, 619, 690–691, 691*b*
Urethrorectal fistulas, 642
Urge incontinence, 612
Urinalysis
abnormal findings in kidney disorders, 576, 577*t*
in renal function evaluation, 571–572
Urinary incontinence (UI), 611
nocturnal, 613
Urinary system, changes, in pregnancy, 668
Urinary tract. *see also* Kidney(s); Renal *entries*
lower
congenital, 615–617, 617*b*
diagnostic tests for, 611
disorders of, 609–625
functional anatomy of, 609–610
innervation of, 610
neoplasms, 617–618, 618*b*–619*b*
obstruction of, 622, 622*b*
role of, 609
symptoms and syndromes, 611–614, 614*b*
obstruction of, 583*f*
structure of, 552*f*
Urinary tract disorders, lower, 609–611, 611*b*
inflammatory/infectious of, 619–621, 621*b*
cystitis as, 619–621, 620*b*, 621*f*
urethritis as, 619
voiding dysfunction as, 611
Urinary tract infection (UTI)
risk factors for, 581, 582*b*
upper, 581
urethritis as, 619
Urine
normal composition of, 571*t*
supersaturation of, in stone formation, 584
Urodilatin, in renal tubular fluid reabsorption, 568–569, 568*t*
Urodynamic testing, 611
Uroflowmetry, 611
Urogenital sinus, female, 633, 634*f*
Urography, intravenous, 572–573
Uroguanylin, in renal tubular fluid reabsorption, 568–569, 568*t*
Urolithiasis
bladder (vesical), 622
lower urinary tract, 622
Uromodulin, 572
Urosepsis, 582
Urothelial carcinoma, bladder, 617
Uterine leiomyomas, 678, 678*f*
Uterine prolapse, 674–675, 674*f*
Uterine retrodisplacement, 675
Uterus, 658
bacteria targeting, 153*f*
position of, alterations in, 674–676, 676*b*
retrodisplacement of, 675, 675*f*

V

Vagina, 658
bacteria targeting, 153*f*
Vaginal cancer, 680
Vagus nerve (CN X), 861*t*, 863, 864*f*
Validity, of data, 5
Valves, cardiac, 355–356, 356*f*
Valvular cardiomyopathy, 401*t*
Valvular diseases, 394–399, 394*f*, 399*b*
Valvular incompetence, 325, 332
Varicose veins, 325, 332–333, 332*f*
Vas deferens, 630–631, 630*f*
Vasa recta, of nephron, 553, 554*f*
Vascular dementia, 919
Vascular disorders
altered hemostasis and coagulation in, 304–306
complicating diabetes mellitus, 826–827
Vascular endothelium, 323
Vascular fluids, 522, 522*f*
Vascular permeability, increased, in inflammation, 171, 172*f*, 172*t*
Vascular purpura, 304–306
Vasculitis, 1074
blood flow alterations from, 325
Vasoconstriction
in coronary circulation, 361
hypoxic, 466
Vasodilation, in coronary circulation, 361
Vasogenic edema, 896
Vasomotor center, in blood flow control, 322
Vasomotor tone, 322
Vasopressor, for anaphylactic shock, 444
Vasospasm
atherosclerosis and, 387
blood flow alterations from, 325
Vasovagal response, 372

Vectors, 373
of pathogen, 148
Vehicle-borne transmission of infection, 148
Veins
anatomy of, 316, 316*f*
compliance of, 321
systemic, primary, 314*f*
thrombosis in, 324
valves of, incompetence of, 332, 332*f*
varicose, 332–333, 332*f*
Venous blood flow alterations, 332–333, 333*b*
in chronic venous insufficiency, 333
in deep vein thrombosis, 333
in valvular incompetence, 332, 332*f*
in varicose veins, 332–333, 332*f*
Venous blood oxygen content, 269, 273*t*
Venous drainage, of heart, 360, 362*f*
Venous insufficiency, chronic, 325, 333, 333*b*
Venous pump, 322
Venous return, 375
Ventilation, 460–464, 464*b*–465*b*
airway resistance and, 461–462, 462*f*
alveolar, 461
dead space in, 460
distribution of, 462, 463*f*
lung compliance in, 462
lung volumes and capacities in, 460, 460*f*, 460*t*
mechanics of breathing and, 461, 462*f*
minute, 460
neurologic control of, 462–464, 464*f*
wasted, 460
Ventilation-perfusion ratios, 465–466, 466*b*
Ventricles
of brain, 851–852, 854*f*
of heart, 355
impedance to ejection from, 376
Ventricular assist devices (VADs), for cardiogenic shock, 441
Ventricular dysrhythmias, 427–428
Ventricular ejection, in cardiac cycle, 360
Ventricular escape rhythm, 425, 425*f*
Ventricular fibrillation, 427–428, 427*f*
Ventricular hypertrophy, remodeling and progression of, in heart failure, 415, 416*f*
Ventricular septal defect, 407, 407*f*
Ventricular stroke volume, preload on, 375, 375*f*
Ventricular tachycardia, 427, 427*f*
Ventriculoperitoneal shunt, for hydrocephalus, 924, 925*f*
Vermiform appendix, 703
Verrucae, 1064, 1065*f*, 1079
Vertebral column, 1023, 1023*f*
Vertigo, 938
in Meniere disease, 940
Very-low-density lipoproteins, in atherosclerosis, 383, 384*f*
Vesicles, 1063*f*–1064*f*, 1064
Vesicoureteral reflux, primary, 615–616, 615*f*
Vesicourethral fistulas, 642
Vesicourethral unit, 633, 634*f*–635*f*
Vestibulocochlear nerve (CN VIII), 861*t*, 863, 864*f*
Viral hepatitis, acute, 767
Viral infections, 1064–1065
Viral provirus, 240, 241*f*
Virchow's triad, 473–474, 474*b*
Virions, 241, 241*f*
Virulence, microbial, 144
Virus(es)
cell injury from, 70
in etiology of hematologic neoplasms, 216–217
human immunodeficiency, 234
morphology of, 150*f*
pathogenic, 150–151, 154*t*
specific diseases caused by, 154*t*
Vision, 942–952
development of, in newborn, 946*b*
in elderly, 946*b*
pathways of, 943–944, 944*f*
Visual field deficits, 950, 951*f*
Visual impairments
age-related, 945–947
amblyopia as, 945–947
cataracts and, 947
errors of refraction as, 944–945, 945*f*
general manifestations of, 944
glaucoma as, 948–950
interventions for, 950–952, 952*b*
macular degeneration as, 948, 949*f*
retinal detachment as, 947, 947*f*
retinopathy as, 947–948
strabismus as, 945
visual field deficits as, 950, 951*f*
Vital capacity, 460*t*
Vitamin B_{12}, in erythropoiesis, 266–267
Vitamin B_{12} (cobalamin) deficiency, anemia from, 278–279
Vitamin D, active, renal synthesis of, 570
Vitamin K deficiency, bleeding in infancy from, 309
Vitamins
in erythropoiesis, 266–267
functions and deficiency syndromes of, 71*t*
Vitiligo, 1080, 1082*f*
Voiding cystourethrography (VCUG), 611
Voiding dysfunction, 611
Voltage-gated ion channels, 44, 44*f*, 47*f*
in membrane potentials, 873
Volume work, of heart, 375
Volvulus, 734, 734*f*
Vomiting, 709, 723
von Willebrand disease, 309
Vulvar cancer, 680
Vulvovaginitis, 677

W

Wallerian degeneration, 881–882
Warts, 1064, 1065*f*, 1079
genital, 694
Wasted ventilation, 460
Wasting syndrome, in HIV infection, 249
Water absorption, 715, 716*f*
Water imbalances, 526
Water intoxication, 527
WBCs (white blood cells), 162–167, 262. *see also* Leukocytes
Weight gain, in pregnancy, 668
Wells rules, of pulmonary embolism, 474–475, 475*b*
Wernicke aphasia, after stoke, 908
Western blot test, in HIV infection diagnosis, 246
Wet dressings, in skin care, 1086
Wheal, 1063*f*–1064*f*
White blood cells (WBCs), 162–167. *see also* Leukocytes
malignant disorders of, 215–232
Wilson disease, 774–775
Wiskott-Aldrich syndrome, 211
Withdrawal reflex, 885, 887*f*
Wolff-Parkinson-White syndrome, 429, 429*f*
Wolff's law, 1004–1005
Women, mental illness and, 984
Workload, cardiac, 376
World Health Organization (WHO), in disease control and prevention, 143–144
Worthlessness, feelings of, major depression and, 980–981
Wounds
burns
healing of, 1106–1107, 1106*f*
management of, 1103, 1104*t*
repair, 1024–1025, 1025*f*–1026*f*

X

X-linked agammaglobulinemia (XLA), 212
X-linked disorders, 106–107, 107*t*, 108*f*
Xerostomia, in HIV infection, 249

Y

Yeast infections, 1065–1067, 1068*f*
Yolk sac, 663

Z

Z disks, 363
Z lines, in sarcomeres, 1015, 1015*f*
Zinc deficiency syndromes, function and, 71*t*
Zona fasciculata, 795, 796*f*
Zona glomerulosa, 795, 796*f*
Zona reticularis, 795, 796*f*
Zygote, 658

Prefixes and Suffixes Commonly Used in Medical Terminology

Prefix	Meaning	Suffix	Meaning
a-	Without; not	-al, -ac	Pertaining to
a[d]-	Toward	-algia	Pain
all[o]-	[An]other; different	-aps, -apt	Fit; fasten
an-	Without; not	-arche	Beginning; origin
ante-	Before	-ase	Signifies an enzyme
anti-	Against; resisting	-blast	Sprout; make
auto-	Self	-centesis	A piercing
bi-	Two; double	-cide	To kill
circum-	Around	-clast	Break; destroy
co-, con-	With; together	-crine	Release; secrete
contra-	Against	-ectomy	A cutting out
de-	Down from; undoing	-emesis	Vomiting
dia-	Across; through	-emia	Refers to blood condition
dipl-	Twofold; double	-flux	Flow
dys-	Bad; disordered; difficult	-gen	Creates; forms
ectop-	Displaced	-genesis	Creation; production
ef-	Away from	-gram	Something written
em-, en-	In, into	-graph(y)	To write; draw
endo-	Within	-hydrate	Containing H_2O (water)
epi-	Upon	-ia, -sia	Condition; process
eu-	Good	-iasis	Abnormal condition
ex-, exo-	Out of; out from	-ic, -ac	Pertaining to
extra-	Outside of	-in	Signifies a protein
hapl-	Single	-ism	Signifies "condition of"
hem-, hemat-	Blood	-itis	Signifies "inflammation of"
hemi-	Half	-lemma	Rind; peel
hom(e)o-	Same; equal	-lepsy	Seizure
hyper-	Over; above	-lith	Stone; rock
hypo-	Under; below	-logy	Study of
infra-	Below; beneath	-lunar	Moon; moonlike
inter-	Between	-malacia	Softening
intra-	Within	-megaly	Enlargement
iso-	Same; equal	-metric, -metry	Measurement; length
macro-	Large	-oid	Like; in the shape of
mega-	Large; million(th)	-oma	Tumor
mes-	Middle	-opia	Vision; vision condition
meta-	Beyond; after	-oscopy	Viewing
micro-	Small; millionth	-ose	Signifies a carbohydrate (especially sugar)
milli-	Thousandth	-osis	Condition; process
mono-	One (single)	-ostomy	Formation of an opening
neo-	New	-otomy	Cut
non-	Not	-penia	Lack
oligo-	Few; scanty	-philic	Loving
ortho-	Straight; correct; normal	-phobic	Fearing
para-	By the side of; near	-phragm	Partition
per-	Through	-plasia	Growth; formation
peri-	Around; surrounding	-plasm	Substance; matter
poly-	Many	-plasty	Shape; make
post-	After	-plegia	Paralysis
pre-	Before	-pnea	Breath; breathing
pro-	First; promoting	-(r)rhage, -(r)rhagia	Breaking out; discharge
quadr-	Four	-(r)rhaphy	Sew; suture
re-	Back again	-(r)rhea	Flow
retro-	Behind	-some	Body
semi-	Half	-tensin, -tension	Pressure
sub-	Under	-tonic	Pressure; tension
super-, supra-	Over; above; excessive	-tripsy	Crushing
trans-	Across; through	-ule	Small; little
tri-	Three; triple	-uria	Refers to urine condition

Adapted from Patton KT, Thibodeau GA: *Anatomy & physiology,* ed 9, St Louis, 2016, Mosby.

Word Roots Commonly Used in Medical Terminology

Root	Meaning	Root	Meaning
acro-	Extremity	men-, mens-, (menstru-)	Month (monthly)
aden-	Gland	metr-	Uterus
alveol-	Small hollow; cavity	muta-	Change
angi-	Vessel	my-, myo-	Muscle
arthr-	Joint	myc-	Fungus
asthen-	Weakness	myel-	Marrow
bar-	Pressure	myx-	Mucus
bili-	Bile	nat-	Birth
brachi-	Arm	natr-	Sodium
brady-	Slow	nephr-	Nephron; kidney
bronch-	Air passage	neur-	Nerve
calc-	Calcium; limestone	noct-, nyct-	Night
capn-	Smoke; carbon dioxide	ocul-	Eye
carcin-	Cancer	odont-	Tooth
card-	Heart	onco-	Cancer
cephal-	Head; brain	ophthalm-	Eye
cerv-	Neck	orchid-	Testis
chem-	Chemical	osteo-	Bone
chol-	Bile	oto-	Ear
chondr-	Cartilage	ov-, oo-	Egg
chrom-	Color	oxy-	Oxygen
corp-	Body	path-	Disease
cortico-	Pertaining to cortex	ped-	Children
crani-	Skull	phag-	Eat
crypt-	Hidden	pharm-	Drug
cusp-	Point	phleb-	Vein
cut(an)-	Skin	photo-	Light
cyan-	Blue	physio-	Nature (function) of
cyst-	Bladder	pino-	Drink
cyt-	Cell	plex	Twisted; woven
dactyl-	Fingers; toes (digits)	pneumo-	Air; breath
dendr-	Tree; branched	pneumon-	Lung
dent-	Tooth	pod-	Foot
derm-	Skin	poie-	Make; produce
diastol-	Relax; stand apart	pol-	Axis; having poles
dips-	Thirst	presby-	Old
ejacul-	To throw out	proct-	Rectum
electr-	Electrical	pseud-	False
enter-	Intestine	psych-	Mind
eryth(r)-	Red	pyel-	Pelvis
esthe-	Sensation	pyo-	Pus
febr-	Fever	pyro-	Heat; fever
gastr-	Stomach	ren-	Kidney
gest-	To bear; carry	rhino-	Nose
gingiv-	Gums	rigor-	Stiffness
glomer-	Wound into a ball	sarco-	Flesh; muscle
gloss-	Tongue	scler-	Hard
gluc-	Glucose; sugar	semen-, semin-	Seed; sperm
glutin-	Glue	sept-	Contamination; separation
glyc-	Sugar (carbohydrate); glucose	sigm-	Greek S or Roman S
hepat-	Liver	sin-	Cavity; recess
hist-	Tissue	son-	Sound
hydro-	Water	spiro-, -spire	Breathe
hyster-	Uterus	stat-, stas-	A standing; stopping
iatr-	Treatment	syn-	Together
kal-	Potassium	systol-	Contract; stand together
kary-	Nucleus	tachy-	Fast
kerat-	Cornea	therm-	Heat
kin-	To move; divide	thromb-	Clot
lact-	Milk; milk production	tom-	A cut; a slice
lapar-	Abdomen	tox-	Poison
leuk-	White	troph-	Grow; nourish
lig-	To tie; bind	tympan-	Drum
lip-	Lipid (fat)	varic-	Enlarged vessel
lys-	Break apart	vas-	Vessel; duct
mal-	Bad	vesic-	Bladder; blister
melan-	Black	vol-	Volume

Adapted from Patton KT, Thibodeau GA: *Anatomy & physiology,* ed 9, St Louis, 2016, Mosby.