

Table 5-1 Overview of Epidemiologic Data Sources

Data Source	Nature of Data	Availability	Population Coverage	Strengths versus Limitations
Mortality statistics	Data from registration of vital events	Annually, from vital registration systems and political subdivisions	Complete	Useful for studying mortality overall and by specific cause
Medical data from birth records	Data on congenital anomalies, complications of pregnancy and childbirth, birth weight, etc.	Annually, from vital registration systems and political subdivisions	Complete	Routinely available; some aspects of morbidity may be incompletely reported (e.g., etiologic factors)
Reportable disease statistics	Statistics based on physician reports, new cases of notifiable communicable diseases	Weekly reports for the United States	Complete	Useful for detection of outbreaks of infectious diseases; some conditions not completely reported
Mass diagnostic and screening surveys	Data that result from diagnostic and screening tests for specified diseases	On an ad hoc basis	Variable	Unknown completeness
Disease registries	Statistics based on existing case registries of cancer, stroke, etc.	Continuous, from national, state, and local jurisdictions	Presumably complete for selected diagnosed diseases	Useful for studying incidence of diseases such as cancer; used to select cases in a case-control study
Morbidity surveys of general population	From the U.S. Health Examination Survey (HES) and the Household Interview Survey (HIS)	Continuous, data released for download	Complete for probability sample	Useful for epidemiologic research; some HIS data may not be accurate for self-reports; HES data contain more precise diagnoses
Health insurance statistics	Cases given medical care under prepaid insurance coverage	Not generally available, although some agencies conduct research on their own insured	Covered population	Useful for healthcare utilization studies; research on morbidity and mortality in selected populations

Life insurance statistics	Mortality data and results of physical examinations of those applying for	Not generally available	Insurance policyholders	Data for selected population
Hospital inpatient statistics	Cases treated in hospital; dependent on type of hospital	Generally not available without special approval	Not determinable	Diagnostic information may be of higher quality than that from other sources; difficult to relate cases to a population denominator
Hospital outpatient statistics	Patients in clinics and outpatient divisions (OPD) of hospitals	Generally not available without special approval	Not determinable	OPDs provide a large volume of care; OPD hospital records are sometimes not well developed; diagnostic data may be incomplete
Data on diseases treated in special clinics and hospitals	Dependent on nature of clinic or hospital; essentially medical care data	Generally not available without special approval	Not determinable	Counts of patients treated; difficult to determine prevalence rates; population denominator unknown
Data from public health clinics	Data from physical examinations of clients	Generally not available without special approval	Not determinable	Possible use for identification of cases of disease and for study of health services; population denominator unknown
Data from records of physicians' practices	Medical care provided in physicians' offices	Generally not available without special approval	Not determinable	May be useful for identification of cases; records may vary in completeness and quality; duplication of cases for patients who see multiple providers
Absenteeism data	Frequency of absenteeism from work or school	By special arrangement with school system or industry	Probably complete for selected population	Nonspecific indicator of disability in selected population; useful for assessing acute disease outbreaks
Data from school health programs	Findings of physical examinations of schoolchildren	Generally not available	Elementary and secondary school population	Uneven quality and completeness of data

continues