CHAPTER 12

Toxicity testing and risk assessment

Chapter outline

In this final chapter you will learn about the testing of chemicals for toxicity and the assessment of risk from chemicals:

- Introduction: evaluation of toxicity
- Human exposure and epidemiology
- Acute toxicity tests
- Sub-chronic toxicity tests
- Chronic toxicity tests
- In vitro toxicity tests
- Risk assessment and interpretation of toxicological data
  - risk assessment
  - hazard identification
  - dose–response assessment
  - exposure assessment
  - risk characterization

Introduction

In most countries, drugs (including veterinary medicines), food additives and contaminants, industrial chemicals, pesticides and cosmetics, to which humans and other living organisms in the environment may be exposed, have to be evaluated for toxicity. The regulations can vary between countries, however, and it is not within the scope of this book to discuss the regulations in any detail. More information may be gained from the references in the Bibliography. The purpose of Regulatory Toxicology is to ensure that the benefits of chemical substances intended for use by humans outweigh the risks from that use.

Evaluation of toxicity

The toxicity of a chemical can be determined in one of three ways:

a by observing human (or animal or plant) populations exposed to a chemical (**epidemiology**);

b by administering the chemical to animals or plants under controlled conditions and observing the effects (*in vivo*);

c by exposing cells, subcellular fractions or single-celled organisms to the chemical (*in vitro*).

The exposure of humans to chemicals may occur accidentally through the environment, as part of their occupation or intentionally, as with drugs and food additives. Thus, chemical accidents, if thoroughly documented, may provide important information about the toxicity of a chemical in humans. Similarly, exposure of humans to chemicals at work may, if well monitored and recorded, provide evidence of toxicity. Thus, the monitoring of exposure by measuring substances and their metabolites in body fluids and using **biochemical indices** of pathological change may be carried out in humans during potential exposure (see biomarkers, Chapter 4). An example is the monitoring of agricultural workers for exposure to organophosphorus compounds by measuring the degree of inhibition of **cholinesterases** in blood samples. However, acquiring such data is often difficult and is rarely complete or of a good enough standard to be more than additional to animal studies. An exception to this is the experimental administration of industrial chemicals to volunteers. But such chemicals are usually not very toxic and the exposure levels would be very low, only sufficient to determine metabolism and disposition, for example. Of particular importance is quantitative exposure data for humans which is often sadly lacking. The development of sensitive and specific biomarkers will improve the acquisition of such data.

Studying particular populations of predatory birds and measuring certain parameters, such as **eggshell thickness** and pesticide level, is an **ecotoxicological** example of testing for toxicity in the field.

Before marketing drugs are first given to a small number of human volunteers and then later to a limited number of patients (**Phase 1 clinical trials**), then to a larger number of patients (**Phase 2 clinical trials**) and then to a large number of patients (**Phase 3 clinical trials**) before being made available to the general public. Both from the clinical trials and the eventual use by the general public, adverse reactions may be detected. Phase 1 trials yield information about metabolism and disposition. Phase 2 and Phase 3 trials yield information about side effects and efficacy.

For human and veterinary medicines in the UK there is a system for reporting adverse reactions to drugs: for human medicines this is the **yellow card system**; for veterinary drugs adverse reactions of both the animal patient and the human user are reported.

Clearly, apart from accidental exposure to high levels and exceptional circumstances where unexpected toxicity occurs, human exposure will normally be to levels that cause little or no toxicity.

Data obtained from human exposure or clinical trials is analysed by epidemiological techniques (although there will be differences between clinical trials that are designed and accidental or occupational exposure). Typically, effects observed will be compared to those in control subjects with the objective of determining if there is an association between exposure to the chemical and a disease or adverse effect.

There are four types of epidemiological study:

a **Cohort studies** in which individuals exposed to the chemical of interest are followed over time *prospectively*.

b **Case-control studies** in which individuals who have been exposed and may have developed a disease are compared *retrospectively* with similar control subjects who have no disease.

c **Cross-sectional studies** in which the prevalence of a disease in an exposed group is studied.

d **Ecological studies** where the incidence of a disease in one geographical area (where there may be hazardous chemical exposure) is compared with the incidence in another area without the hazardous chemical.

For accidental, unintentional exposure the analysis is normally *retrospective* and will be a 'case-control study'. The control population will be human subjects chosen to be as similar as possible for age, sex and other parameters. This type of design would be used for studying the relationship between exposure to a volatile chemical in the workplace and lung cancer, for example. Of course there will be a prevalence of lung cancer in the controls but the intention is to discover if the prevalence is higher in those exposed to the chemical. Another example is of drugs already in use in the general population where adverse drug reactions (**ADRs**) in patients are reported by clinicians. For this type of design the data can be represented as an **odds ratio*** which is an estimate of relative risk.

The other type of design is *prospective* (known as a cohort study) and is used in clinical trials of drugs. Controls are subjects selected out of the patient population and have the disease for which the drug is prescribed. The controls receive an inactive '**placebo**'.

Epidemiological data can be analysed in various ways to give measures of effect. The data can be represented as an *odds ratio** which is the ratio of the risk of disease in an exposed group compared to a control group. The ***relative risk*** is determined as the ratio of the occurrence of the disease in the exposed to the unexposed population. The ***absolute excess risk*** is an alternative quantitative measure#.

When setting up epidemiological studies and when assessing their significance it is important to be aware of confounding factors such as bias and the need for proper controls.

For further details on epidemiology the reader is referred to the bibliography.

Although human data from epidemiological studies is useful, the majority of data on the toxicity of chemicals is gained from experimental studies in animals. The data so acquired is used for the risk assessment and safety evaluation of drugs prior to human exposure, for food additives before use and for industrial and environmental chemicals. In the case of drugs this information is essential before the drug can be administered to patients and for food additives and other chemicals it is required in order to set a No Observed Adverse Effect Level (**NOAEL**, see this chapter).

Because animal tests can be carefully controlled with the doses known exactly, the qual-

---

* Calculated as: A × B/C × D
A = no. of cases of disease in exposed population; B = no. of unexposed controls without disease; C = no. of exposed subjects without disease; D = no. of unexposed controls with disease.

\# Relative risk calculated as: A/B where A = no. of cases of disease in total exposed group per unit of population; B = no. of cases of disease in total non-exposed control group per unit of population.
Absolute excess risk calculated as: no. of cases of disease per unit of exposed population minus no. of cases of disease per unit of unexposed population.

ity of the data is generally good. The number of animals used should be enough to allow statistical significance to be demonstrated. Humane conditions and proper treatment of animals is essential for scientific as well as ethical reasons as this helps to ensure that the data is reliable and robust. The problem of extrapolation between animal species and humans always has to be considered but past data as well as theoretical considerations indicate that in the majority of cases toxic effects occurring in animals will also occur in humans.

The conduct of the animal toxicity tests required depend partly on the type of substance and its expected use and also on the regulations of the particular country. The amount of data necessary also depends on the end use of the substance. For instance, industrial chemicals produced in small quantities may only require minimal toxicity data whereas drugs to be administered to humans require extensive toxicological testing. Pesticides may have to be tested on many different types of animal and plant in the environment and examined for their persistence and behaviour in food chains. The stability of such substances in particular environments is also of importance. Consequently, ecotoxicology involves more extensive residue analysis than does drug toxicology. However, for veterinary medicines, determination of residues in food for human consumption is required.

The species selected will depend partly on the type of toxicity test, data available and also ethical and financial considerations. For example, although the old world monkey, being generally the most similar to humans, might be the desirable species to use for a particular toxicity evaluation, both cost and ethical reasons will often rule this out. The most common species used are rats and mice for reasons of size, accumulated knowledge of these species and cost. Currently, mice have the advantage in being available as **genetically modified** varieties. To show and evaluate some types of toxic effect a particular species might be required.

For veterinary drugs or environmental pollutants the target species will normally be used. Normally young adult animals of both sexes will be used. The exposure level of the chemical used will ideally span both non-toxic and maximally toxic doses.

Examples of pertinent questions which should be asked *before* any toxicity evaluation are:

1 is it a **novel compound** or has it been in use for some time?

2 is it to be **released into the environment**?

3 is it to be **added to human food**?

4 is it to be given as a **single dose** or repeatedly?

5 at what **dosage level** is it to be administered?

6 what **age group** will be exposed?

7 are **pregnant women** or women of childbearing age likely to be exposed?

Toxicity may be an *intrinsic* property of a molecule which results from interaction with a particular biological system. Consequently, a knowledge of the **physico-chemical properties** of that molecule may help the toxicologist to understand the toxicity or potential toxicity and to predict the likely disposition and metabolism. Indeed, we have seen several examples in this book of the importance of physicochemical principles in toxicology. **Structure–activity relationships** are beginning to be used in toxicology as they are in pharmacology, especially in the field of chemical mutagenesis/carcinogenesis. This initial knowledge from preliminary studies may also influence the course of the subsequent toxicity tests especially if there are similarities with other compounds of known toxicity. Hence, the

solubility, partition coefficient, melting or boiling point, vapour pressure and purity are important parameters. For example, an industrial chemical which is a very volatile liquid (i.e. with a high vapour pressure) should at least be tested for toxicity by inhalation and possibly by skin application.

As well as physico-chemical considerations there are also biological considerations and the following are the major ones:

1 the most appropriate **species** to study,
2 the **sex** of the animals used,
3 the use of **inbred** or **outbred** strains,
4 **housing**,
5 **diet**,
6 animal **health**,
7 **metabolic similarity to man**,
8 the **route of administration**,
9 **duration** of the toxicity study,
10 the **numbers** of animals used,
11 **vehicle**.

The route of administration and vehicle will depend on the expected end use or, if a drug for example, on the means of administration. The parameters to be measured may also be dependent on the particular study. For example, metabolic studies can be combined with a toxicity study and plasma levels measured as well as urinary metabolites identified and clinical chemical parameters studied. The biochemical and pathological measurements to be made will also be decided before the study is started.

Initial toxicity studies will usually be carried out to determine the *approximate range of toxic dosage*. For a drug this may already be known from pharmacological studies but for an industrial chemical, for instance, nothing may be known of its biological activity. Consequently, the initial range-finding studies may utilize dosage on a logarithmic scale or half-log scale. These initial studies are important if large numbers of animals are not to be wasted in later studies. The initial tests will also involve *observation* of the animals in order to gain insight into the possible toxic effects.

Once the approximate toxic dosage range is known then various detailed toxicity studies can be carried out. These will be followed by various other toxicity tests, usually including the following: **acute, sub-chronic (28- or 90-day), chronic (lifetime), mutagenicity, carcinogenicity, teratogenicity, reproductive studies** and ***in vitro* tests**. For some compounds there may also be other types of toxicity test such as **irritancy** and **skin sensitization** studies.

There are different requirements for drugs, food additives and contaminants, industrial chemicals, cosmetics and pesticides because of the different circumstances of exposure. Chemicals which are to be used in the environment, such as pesticides and industrial chemicals which might be accidently released into the environment, will also undergo **ecotoxicity** tests. These will include tests with invertebrates such as ***Daphnia***, **earthworms**, **fish**, **phytoplankton** and **higher plants**.

Acute toxicity tests

Acute toxicity tests are those designed to determine the effects which occur within a short period after dosing. These tests can determine a **dose–response relationship** and the **LD_{50} value** if required. The exact conduct of toxicity studies will vary depending on the compound, its eventual use and the particular regulations to be satisfied. Usually at least four dosages are used which may be in *logarithmic progression*

TABLE 12.2 *The type of information required (including toxicity) for a new chemical substance under the EU New Substances Notification Scheme*

Identity
name/trade name
formulae (empirical/structural)
composition
methods of detection/determination

Uses and Precautions
proposed uses
estimated production/importation
handling/storage/transport methods and precautions
emergency measures

Physico-chemical properties
melting point
boiling point
relative density
vapour pressure
surface tension
water solubility
fat solubility
partition coefficient (octanol/water)
flash point
flammability
explosive properties
auto-flammability
oxidizing properties

Toxicology studies
acute toxicity (oral/inhalation/cutaneous)
skin and eye irritancy
skin sensitization
subacute toxicity (28 days)
mutagenicity (bacterial and non-bacterial)

Ecotoxicological studies
toxicity to fish
toxicity to *Daphnia*
degradation data (BOD, BOD/COD)

Possibility of rendering substances harmless
for industry
for public
declaration concerning the possibility of unfavourable effects
proposed classification and labelling
proposals for any recommended precautions for safe use

Further information from Fairhurst, S., chapter 67, and Auer, C. M. and Fielder, R. J., chapter 72, in Ballantyne, Marrs and Syversen, 1999, see Bibliography.

of the compound on the development of the embryo and foetus. These may be detected as gross anatomical *abnormalities* in the newborn animal or may be more subtle effects such as changes in behaviour. The effect of the compound on the **fertility** of both male and female animals may also be determined in reproductive toxicity tests. Data from other tests may also be relevant, such as pathological evidence of testicular damage which might additionally be detected as a decrease in male fertility.

Mutagenicity tests determine whether the compound has potential to cause **genetic damage** and so induce a mutation in germ cells and somatic cells. Such tests indicate whether a compound may have the potential to induce cancers. Mutagenicity tests are carried out in **bacteria** and **cultured mammalian cells**

in vitro. *In vivo* assays include the **micronucleus test** and the **dominant lethal assay** (see Bibliography for details).

Carcinogenicity tests may also be required, especially if the mutagenicity tests are positive. The compound is given for the lifetime of the animal, administered either in the drinking water or diet. The appearance of tumours at post-mortem or perhaps before the animal dies are detected from histopathological studies of sections of tissues from the major organs.

Irritancy and **skin sensitization tests** may also be required, especially for industrial chemicals and pesticides. **Irritancy tests** are sometimes carried out on rabbit skin or eyes. The **skin sensitization test** is normally carried out in the guinea pig and a positive result indicates that the compound has the potential to cause contact dermatitis in humans. Some compounds may also cause **pulmonary sensitization** but there is no reliable animal model for this effect. Consideration of the toxicity data may suggest that further studies be carried out, such as an investigation to show that an effect is peculiar to a particular species and therefore not relevant to man.

Toxicity tests are normally either carried out by the company producing the compound or a **contract research laboratory** or a combination of both. The conduct of the toxicity and ecotoxicity studies should conform to certain guidelines, such as those issued by the **Organisation for Economic Cooperation and Development (OECD)**. These guidelines are often enshrined in national regulatory requirements such as those in the UK and USA. Toxicity tests also now must be carried out in compliance with a system known as **Good Laboratory Practice (GLP)**, which governs every aspect of the conduct of studies including the reporting of results. This system was introduced to ensure that toxicity tests are competently carried out and that data is not fabricated, following a notorious situation which arose in the USA.

As well as the requirements of regulatory agencies, toxicity data may also have other uses. Indeed, the data may be life saving in cases of human and animal poisoning. For example, animal studies on **cyanide toxicity** provided data which was useful in the **treatment of poisoning** with cyanide. The *absence* of any toxicity data on **methylisocyanate** probably hampered the efforts of rescue workers and clinicians at **Bhopal** in India after the massive disaster where methylisocyanate leaked from a chemical plant there. Basic studies on **paracetamol toxicity** led directly to the use of an **antidote** which has proved extremely successful and life saving. Attempts to understand the mechanisms underlying the toxicity of compounds will allow better prediction of toxicity and also better design of tests to discover toxic potential.

Testing in vitro

It has become necessary to question the use of *in vivo* safety evaluation studies because of the pressure from society to reduce the use of live animals in medical research. Consequently, there has been an increase in the exploration and use of various *in vitro* systems in toxicity testing. The current philosophy is embodied in the concept of the **three R's**: **replacement**, **reduction** and **refinement**. Thus, if possible, live animals should be *replaced* with alternatives. If this is not possible then measures should be adopted to *reduce* the numbers used. Finally, research workers should also *refine* the methods used to ensure greater animal welfare and reduction in distress and improve the quality of the data derived, if possible.

In some areas the use of *in vitro* systems has been successful. For example the use of *in vitro* tests for the detection of genotoxicity is now well established. These tests include the well known **Ames test** which relies on detecting mutations in **bacteria** (*Salmonella typhimurium*). These are useful early screens for detecting potential toxicity, in particular genotoxicity, which may lead to the production of tumours in whole animals.

Other microorganisms such as *E. coli* bacteria and **yeast** may be used. Mammalian cells are also used for tests for genotoxicity, typically **mouse lymphoma** or **Chinese hamster ovary** cell lines. Human lymphocytes can also be used for the detection of chromosomal damage. Fruit flies are sometimes used for specific tests such as the detection of sex-linked recessive lethal mutations. However, the correlation between a positive result for mutagenicity in tests such as the bacterial test and carcinogenicity in an animal is not 100 per cent. That is, known animal carcinogens are not universally mutagenic in the bacterial tests and *vice versa* some mutagenic chemicals are not carcinogenic in animals. Therefore although *in vitro* bacterial tests may be used to screen out potential genotoxic carcinogens, those compounds which are not apparently mutagenic may still have to be tested for carcinogenicity *in vivo*.

One area where *in vitro* tests have been successful is in the testing of **cosmetics**. The use of skin cells and simpler *in vitro* systems has allowed the cosmetic industry to dramatically reduce the use of *in vivo* testing of substances for irritancy, for example. However, human skin is generally more readily available than other human tissues and is also more readily utilized in *in vitro* systems.

Apart from bacterial mutagenicity tests and other such tests using single-celled organisms and skin testing, other *in vitro* systems are still not yet widely used as alternatives to *in vivo* experiments. However, progress is being made and recently an *in vitro* alternative to the *in vivo* test for **allergenicity/sensitization** was developed. However, currently many of these tests do not stand alone and require additional data to be gathered *in vivo*. For example, although a bacterial mutagenicity test might indicate a chemical is a potential genotoxic carcinogen, actual carcinogenicity can only be demonstrated in an animal *in vivo*. A positive result in the bacterial test might be sufficient to stop development of a drug but with other compounds such as industrial chemicals which may already be in use, an indication of the actual carcinogenicity may be needed. Similarly, with a cosmetic in development, a positive result in an *in vitro* test might be sufficient to stop development but with other chemicals a more definitive answer may be needed.

One of the *in vitro* systems most used is the **isolated liver cell**. These may be primary liver cells derived from animal or human liver or alternatively cell lines, such as **HepG2** cells, derived from liver tumours.

Unfortunately there are a number of problems with many of the *in vitro* systems currently in use which makes the use of such systems for prediction and risk assessment difficult. Thus, primary cells may show poor viability in medium to long-term experiments and this may limit their usefulness to short-term exposure. There are also major biochemical changes which occur with time in primary cells, starting from almost the moment of preparation of the tissue. Changes, such as in the level and proportions of isozymes of **cytochrome P450** which occur over the first 24 hours after isolation for instance will influence the toxicity of chemicals in those cells if metabolic activation is a factor.

An alternative *in vitro* system is the use of cell lines, immortal cells which will continue to grow and can be frozen and used when

needed. These cells are not, however, the same as those in normal tissue and are often derived from tumours.

When comparisons have been made with *in vivo* data, in many cases the *in vitro* system reacts differently to the tissue in the animal *in vivo*. This difference may be qualitative or quantitative. Therefore, although *in vitro* systems are used and are especially useful for mechanistic studies, the data generated from them has to be viewed with caution. This is particularly the case if the data is being used as part of a risk assessment. Such *in vitro* data may underestimate the toxicity *in vivo*.

Thus, it is not yet possible to replace all animal experiments with *in vitro* systems even though considerable progress has been made. *In vitro* systems are particularly useful, however, for screening out toxic compounds which might otherwise be developed, for mechanistic studies and for comparing different compounds within a group of analogues for example.

Risk assessment and interpretation of toxicological data

At least *65 000 chemicals* are currently produced in the USA with 500–1000 new chemicals added each year. In the past, perhaps chemicals were too readily produced and used without due care and attention. **Rachel Carson** in her book, *Silent Spring* showed the risks of such actions. The general public is now very suspicious of all chemicals and there is perhaps an exaggerated fear of poisoning from chemicals in the environment and a belief that all chemicals are hazardous. Regulation has been introduced in many countries in response to this public fear and pressure. Clearly regulation is necessary, but where possible guidelines should be issued rather than strict rules for the assessment of every case in the same way. A major problem with toxicological data is the assessment of **hazards** and the subsequent calculation of **risks** and estimation of **risk versus benefit**.

RISK ASSESSMENT

Risk is a mathematical concept which refers to the likelihood of undesirable effects resulting from exposure to a chemical.

Risk may be defined as the probability that a hazard will cause an adverse effect under specific exposure conditions.

Risk may also be defined in the following way: Risk = hazard × exposure.

Hazard may be defined as the intrinsic capability of a substance to cause an adverse effect.

Conversely, safety may be defined as 'the practical certainty that adverse effects will not occur when the substance is used in the manner and quantity proposed for its use'.

As exposure increases so does the probability of harm and therefore a reduction in exposure reduces the risk.

Risk assessment would be carried out on chemicals for the following reasons:

a the chemical is likely to be a **hazard** to humans in the environment;

b the likelihood of **persistence** of the chemical in the environment and bioaccumulation;

c the likelihood that **sensitive human** and **ecological** populations may be exposed to significant levels;

d indication of hazard to **human** health;

e likelihood of **exposure** via use or production.

Risk assessment is the process whereby hazard, exposure and risk are determined. An underlying concept in risk assessment relies on the statement by Paracelsus (see Chapter 1), and so for many, although not all chemicals, there will be a dose–effect relationship. Therefore the corrollary is that there should be a safe dose. Consequently, it should be possible to determine a level of exposure that is without appreciable risk to human health or the ecosystem. Risk assessment is a scientific process. The next stages are risk benefit analysis and risk management that require a different type of approach.

Risk management is a process of considering alternative policies and choosing the most appropriate course of regulatory action based on the results of risk assessment and social, economic and political considerations.

Risk assessment is the process whereby the nature and magnitude of the risk is determined. It requires four steps:

i **Hazard identification**. This is the evaluation of the toxic effects of the chemical in question.

ii Demonstration of a **dose–response** or dose–effect relationship. Evaluation of the causal relationship between the exposure to the hazard and an adverse effect in individuals or populations, respectively.

iii **Exposure assessment**. Determination of the level, frequency and duration of exposure of humans to the hazardous substance.

iv **Risk characterization**. Estimation of the incidence of adverse effects under the various conditions of human exposure.

Considering each of these in turn:

Hazard identification

This is the evaluation of the *potential* of a chemical to cause toxicity. The data used is normally derived from:

a **human epidemiology**

b **animal toxicity studies**

c *in vivo* and *in vitro* **mechanistic** or other **studies**.

A chemical may constitute a number of hazards of different severity. However, the *primary hazard* will be the one used for the subsequent stages of the risk assessment process. For example, a chemical may cause reversible liver toxicity at high doses but cause tumours in the skin at lower doses. The carcinogenicity is clearly the hazard of concern.

Although human data is ideal, reliable data from humans is not often available and must be supplemented with other data in order to define a dose–response relationship. However, epidemiological data may at least indicate that a **causal relationship** exists between exposure to the chemical and an effect in humans. Therefore, in practice, animal toxicity data is normally required. This will be generated by toxicity studies that are controlled and that generate histopathological, clinical chemical and biochemical data (see Chapter 4 and this chapter). Of course, the differences between humans and other species must always be recognized and taken into account (see below). It may be possible to use *in vitro* data both from human cells and tissues as well as those from other animals to supplement the epidemiological and animal *in vivo* toxicity data. However, at present such data cannot *replace* experimental animal or human epidemiological data. The predictive use of structure activity relationships is also possible and an approach which is becoming increasingly important.

Dose–response assessment

This stage quantitates the hazards already identified and estimates the relationship between the dose and the adverse effect in humans. However, this requires extrapolation from possibly high, experimental doses used in animals to levels likely to be encountered by humans.

The extrapolation from high to low doses will depend on the type of primary toxic effect. If this is a carcinogenic effect then a **threshold** normally cannot be assumed and a mathematical model is used to estimate the risk at low doses. If the primary toxic effect is non-carcinogenic then it will normally be assumed that a threshold exists.

Risk assessment of carcinogens is a two-step process involving, firstly, a qualitative assessment of the data from the hazard identification stage (see above) and, secondly, a quantitation of the risk for definitive or probable human carcinogens.

The first stage uses either the EPA or IARC classification system which are very similar. The IARC system is shown below.

IARC classification of chemicals in relation to carcinogenicity:

Group 1 *The agent* ***is carcinogenic*** *to humans.* This category is used when there is sufficient evidence of carcinogenicity in humans (e.g. aflatoxin, benzene, arsenic, tobacco smoke).

Group 2A *The agent is* ***probably carcinogenic*** *in humans.* This category is used when there is limited evidence of carcinogenicity in humans but convincing evidence of carcinogenicity in experimental animals (e.g. acrylonitrile, cadmium, benzo[a]pyrene).

Group 2B *The agent is* ***possibly carcinogenic*** *in humans.* This category is used when there is only limited evidence of carcinogenicity in humans and less than convincing evidence of carcinogenicity in experimental animals (e.g. carbon tetrachloride, urethane, hexachlorobenzene).

Group 3 *The agent is* ***not classifiable*** *as to its* ***carcinogenicity***. This is used when the evidence for carcinogenicity of the agent in humans and experimental animals is inadequate (e.g. aniline, dieldrin, maneb).

Group 4 *The agent is* ***probably not carcinogenic*** *in humans.* This is used when the agent has not been found to induce cancer in either experimental animals or humans despite thorough testing (e.g. Caprolactam).

For group 1 the data should show a causal relationship between exposure and cancer in humans. For chemicals classified as group 1 or 2A the second stage is a quantitative risk assessment. The classification may change for a chemical when more information becomes available.

There are several models that can be used and these range from ultraconservative to least conservative:

a the **one-hit model**. This is ultra-conservative as it assumes that cancer involves only one stage and a single molecular event is sufficient to induce a cellular transformation.

b The **linearized Multistage Model** (used by the EPA). This determines the cancer slope factor that can be used to predict cancer risk at a specific dose. It assumes a linear extrapolation to a zero dose threshold (see Figure 1.7). This factor is an estimate (expressed in mg/kg/day) of the probability that an individual will develop cancer if exposed to the chemical for 70 years.

c The **multi-hit model**, which assumes several interactions are necessary for transformation of a normal to a cancerous cell.

d **Probit model**. This assumes a log normal distribution for tolerance in the exposed population.

Another model that is increasingly being used is the **physiologically based pharmacokinetic model**. This utilizes data on the absorption, distribution, metabolism, tissue sequestration, kinetics, elimination and mechanism to determine the target dose used for the extrapolation but it requires extensive data.

The cancer risk values that these models generate are of course very different. For example for the chemical **chlordane**, the lifetime risk for one cancer death in one million people ranges from 0.03 $\mu g l^{-1}$ of drinking water for the one-hit model, 0.07 $\mu g l^{-1}$ from the linearized multi-stage model to 50 $\mu g l^{-1}$ for the probit model.

The results from **animal carcinogenicity testing** studies are particularly hard to assess as it is necessary but difficult to show an *increased frequency* of tumours in a small population such as those used in animal cancer studies, in which there may already be a significant incidence of some types of tumours. There is a practical, statistical limit which determines the incidence or frequency of occurrence of a cancer which can be detected. For example, using 1000 animals it is necessary for more than five animals to be affected by cancer for the effect to be detected at the 99 per cent confidence level; but an incidence of five cases in 1000 test animals if extrapolated to man would translate into over *1 million cases* of cancer in a population the size of that of the US. To use even larger numbers of animals would be impractical, extremely expensive, and challenged on ethical (animal rights) grounds. So assessing cancer risk from carcinogenicity studies is very difficult and those conducting and assessing the tests tend to err on the side of caution. One way around the dilemma of low incidence is to *increase the doses* used in the animal tests on the assumption that the dose-response is linear and so extrapolation backwards is possible. This has given rise to various models but estimates from these models vary; the precision of the mathematical model is largely irrelevant if the quality of the original toxicological data is poor. There may be large margins of error and uncertainty. Unfortunately the public may take the exposure limits and similar data issued at face value or alternately disbelieve them completely. Consequently, doses close to the **Maximum Tolerated Dose (MTD)** are used in carcinogenicity testing despite the problems of **dose-dependent metabolism**, **dose-dependent kinetics**, and the possibility of other pathological effects influencing the carcinogenicity. This approach is contentious, however, as carcinogens may show dose-dependent metabolism and with weak or equivocal carcinogens such as **saccharin** (see Chapter 7) and especially non-genotoxic carcinogens this may be crucial to the interpretation of the carcinogenicity data. That is, large doses of a compound may be metabolized in a quantitatively or qualitatively different manner to that of the expected dose or exposure level. Consequently, a compound may only be carcinogenic under those extreme dosing conditions. For example, the industrial chemical **hydrazine** is a *weak* carcinogen after high exposure or dose levels. It also causes **DNA methylation**, a *possibly* mutagenic event which might lead to cancer but this methylation only occurs after *large*, hepatotoxic doses. The implications of this are that the acute toxic effect is in some way involved in the DNA methylation and that also the acute effect is necessary for the development of the cancer. For non-carcinogens where the dose response shows a threshold, a dose can be determined at which there is no adverse effect, the **No Adverse Effect Level (NOAEL)** (see Figure 1.7). The effect will be one that is likely to occur in humans and that is the most sensitive toxic

effect observed. If a NOAEL cannot be determined (if the data is insufficiently robust) then the **Lowest Adverse Effect Level (LOAEL)** is determined.

Exposure assessment

Exposure to a chemical converts it from being a hazard into a risk. Thus determination of exposure is crucial to the whole process of risk assessment. This involves evaluation of the source of the exposure, the routes by which humans are exposed and the level of exposure.

Of course in some situations of exposure to chemicals, such as around waste disposal areas or chemical factories, exposure is to a mixture of possibly many different chemicals. These may interact in a variety of ways (e.g. additivity, synergism, antagonism, potentiation, see Chapter 1). Exposure may be by more than one route (inhalation, skin contact, ingestion) and different types of organism may be exposed (human, animal, adult, infant). Therefore the real life situation of exposure to chemicals in the workplace or environment can be immensely complex when these factors are taken into account. Risk assessment requires a consideration of these.

Actual exposure levels may not always be known and therefore models may have to be used that utilize knowledge of air dispersion or ground water movements.

The **physico-chemical characteristics** of the chemical in question (i.e. lipid solubility, water solubility, vapour pressure, etc.) also will be important information.

However, the risk assessment process is more reliable if there is an indication of actual exposures for both the experimental animals and humans that have provided the data on which it is based. The exposure assessment may use **biomarkers** to improve the process (see below).

Risk characterization

The final stage involves integration of the results of the preceding stages to get a probability of the occurrence of the adverse effect in humans exposed to the chemical. The biological, statistical and other uncertainties will have to be taken into account.

For carcinogens the risk is expressed in terms of increased risk of developing a cancer (e.g. 1 in 10^6). This is calculated from the **cancer slope factor** and the 70-year average daily intake in mg/kg/day.

From the NOAEL (or LOAEL if there is no reliable NOAEL) various parameters can be determined.

For food additives this is normally the **Acceptable Daily Intake (ADI)** (or the **Reference Dose**, **Rfd**, used by the Environmental Protection Agency in the USA). The ADI is the amount of chemical to which a person can be exposed for a lifetime without suffering harmful effects. The determination of these intake values requires the use of a safety or uncertainty factor. The RfD includes an additional safety factor (modifying factor, see below). For food contaminants the parameter is the **Tolerable Daily Intake (TDI)**. The TDI is an estimate of the daily intake of the chemical that can occur over a lifetime without appreciable health risk. Daily food consumption for a particular type of food will be used for this calculation.

Food may also contain veterinary drug residues and the pesticide residues for which ADIs may be calculated.

Chemicals in water and air also have to be assessed for risk and guidelines set where appropriate. Thus there are air quality/pollution guidelines set by the World Health Organisation (WHO). Air pollutants may have acute irritant effects or chronic effects or both. The guidance values give levels combined with exposure times at which no adverse effects would be

expected. The guidance values are determined from the NOAEL (or LOAEL).

Similarly there are drinking water guidance values for a number of chemicals. For drinking water contaminants as with food contaminants a Tolerable Daily Intake can be established from the NOAEL and appropriate safety factors. The guidance value is determined from the TDI and known daily intake of water by a standard adult of 60 kg weight drinking the water for 70 years. As with air pollutants, carcinogenic, non-threshold chemicals will be considered differently to non-carcinogenic chemicals where there is considered to be no threshold.

In the case of carcinogens a **Virtually Safe Dose (VSD)** may be determined.

The modifying or safety factors are as follows:

10× for **human variability (intra species)**;
10× for **extrapolation** from animals to humans (**interspecies variability**);
10× **if less than chronic doses** have been used;
10× **if** the **LOAEL** rather than the NOAEL is used;
0.1–10× modifying factor. This is only used for determination of the RfD (EPA).

These **uncertainty factors** are combined and divided into the NOAEL (or LOAEL) to give the ADI (or RfD) or TDI. The modifying factor allows for judgement on the quality of the scientific data.

Thus:

$$\textbf{TDI} = \textit{NOAEL}/\textit{Uncertainty factor}(s)$$

$$\textbf{ADI} = \textit{NOAEL}/\textit{Uncertainty factor}(s)$$

Often an uncertainty factor of 100 is applied to account for human variability and for differences between humans and the animals used in the toxicity studies.

This approach can be applied to both chronic and shorter term (e.g. developmental) toxicity and similar methods may be used to derive permissible exposure levels for acute and short-term exposure. Clearly the toxicity data used would be derived from studies of appropriate length. For **occupational exposure** to chemicals as opposed to environmental exposure other parameters such as **Threshold Limit Values (or Maximum Exposure Limits)** are determined in a similar way and are based on exposure for an eight-hour working day.

Doses are normally either expressed on a body weight or body surface area basis and are then extrapolated to a different species. This assumes similar sensitivity per unit body weight or surface area. Thus in the risk assessment process for non-carcinogens the actual exposure level is compared with the ADI or other equivalent parameter for example. Exposure to multiple chemicals will be assumed to be additive.

Extrapolation *between species* is also a problem in risk assessment and the interpretation of toxicological data. For example, one question that arises is 'which species is the extrapolation to be made from, the most sensitive or the one which in terms of response or disposition of the compound is the most similar to man?' The species or strain used in a particular carcinogenicity study may have a high natural incidence of a specific type or types of tumour. The assessment of the significance of an increase in the incidence of this tumour and its relevance to man can pose particular problems. Therefore, risk assessment from carcinogenicity is fraught with difficulties, possibly more than any other type of toxic effect.

For acute toxic effects the dose response is often clear cut and allows a **NOAEL** to be estimated. However, the biology of the toxicity study must always be taken into account and a too exaggerated reliance on statistics must be avoided. Because of the problems of interspe-

cies extrapolation and interpretation of low incidences of tumours, risk assessment may give rise to widely disparate quantitative values. For example, for **saccharin** the expected number of **bladder cancer** cases in the USA over a 70-year period due to daily exposure to 120 mg was *estimated* as between 0.22 and 1.144×10^6! Therefore, in the risk assessment of a particular compound other factors become important such as the *likely* and *reasonable* human exposure but in the USA the strict rules of the **Delaney clause** make this difficult (see Glossary for definition of Delaney Clause).

THE USE OF BIOMARKERS IN RISK ASSESSMENT

Biomarkers are used at several stages in the risk assessment process. Biomarkers of exposure are important in risk assessment as an indication of the ***internal* dose** is necessary for the proper description of the dose–response relationship. Similarly, biomarkers of response are necessary for determination of the NOAEL and the dose–response relationship. Biomarkers of susceptibility may be important for determining specially sensitive groups for estimating an uncertainty factor. Biomarkers allow the crucial link between the response and exposure to be established.

The incidence of a toxic effect may be measured under precise laboratory conditions but extrapolation to a real life situation to give an estimate of risk involves many *assumptions* and gives rise to *uncertainties*. The risk assessor has to decide which are plausible answers to questions when in reality there are either no scientific answers or these answers are obscure.

For a new chemical substance human data is not available and toxic effects in man cannot be verified by direct experiment and so extrapolation from the results of animal studies is essential. Of course the objective is to have as large a margin of safety as possible. However when there is conflicting data does one use the single positive result or the **'weight' of all the data**? Inflated estimates of exposure may occur. **Epidemiology** may be useful for compounds that have been used for some time. Indeed, many compounds have never undergone a full range of toxicity tests (an estimated 70 per cent in the USA) and it would clearly be an enormous task to test all such compounds. Consequently, a reliance on epidemiology is unavoidable.

Conclusions

As yet, toxicologists only partially understand the mechanisms underlying relatively few toxic effects of chemicals. Consequently the assessment of risk to man will remain difficult and uncertain. The limitations need to be borne in mind by the public, by industrialists, economists and regulatory officials, but also by toxicologists themselves.

Perhaps the public expects too much from scientists in general and toxicologists in particular. Toxicology *cannot* provide all of the answers the public often demands as they are beyond current science. The public may demand **absolute safety** but this is an impossible dream. One of the duties of the toxicologist is to make sure the limitations are *understood*.

Perhaps the real crux of the problem of interpretation of toxicological data in the light of increasing and widespread exposure of humans to chemicals is the assessment of **risk versus benefit**. Although the public may not always be aware of the fact that chemicals confer benefits on society, and that there is a greater or lesser risk attached to their use, the benefits may be hard to quantify and compare with the risk. However, just as we take a quantifiable

risk when we drive a car because its use is convenient and maybe essential, then we should apply similar principles to the chemicals we use. Unfortunately the risks and benefits may not always be equally shared, with one section of society reaping financial benefits while another risks the adverse effects.

Summary and learning objectives

Toxicity testing of chemicals is a ***legal*** requirement if humans or animals in the environment are likely to be exposed. This toxicity may be determined from epidemiology studies and clinical trials, *in vivo* studies in animals and studies *in vitro* but is mostly from animals.

Epidemiology (cohort, case control, cross sectional or ecological studies) may indicate relative or absolute risk. *In vivo* tests are carried out but questions must be asked (e.g. dosage size and frequency and physico-chemical properties, novelty) and biological considerations (e.g. species and sex of animal) addressed beforehand. The nature of the test will depend on the type of chemical, its use and the particular chemical. General tests used are ***acute*** (1 dose), ***sub-chronic*** (repeated, 28 or 90 days) and ***chronic*** (at least 12 months in rodents). More specific tests include those for ***reproductive toxicity*** (effects on male or female reproductive system), ***teratogenicity*** (effects on the embryo *in utero*) and ***carcinogenicity*** (ability to cause tumours) and ***ecotoxicity*** (e.g. effects on *Daphnia* and earthworms). For some chemicals (e.g. industrial chemicals) only acute tests may be needed for classification (e.g. non-toxic/very toxic). ***Acute tests*** will help define a dose–response relationship. ***Sub-chronic*** and ***chronic studies*** indicate target organ(s) toxicity, other pathological effects, blood level and no observed effect level. Other specific *in vivo* studies will be carried out if necessary. Toxicity testing *in vivo* should consider the three Rs: ***replacement***, ***reduction*** and ***refinement***. Replacement means the use of ***in vitro test*** systems including those for mutagenicity involving bacteria (e.g. *Salmonella* in the ***Ames*** test), mammalian cells (mouse lymphoma, human lymphocytes) or insects (fruit flies). Testing for cytotoxicity may utilize mammalian-derived cells, mostly for screening out chemicals prior to *in vivo* evaluation or for evaluation of skin toxicity or allergenicity. There are significant limitations to *in vitro* tests (e.g. loss of enzyme activity). Reduction means using the minimum animals necessary and refinement means devising methods to gain the most information while causing the least distress.

Risks from chemical exposure must be assessed in relation to benefit.

Risk is the probability that an adverse effect will occur under specific exposure conditions.

Hazard is the capability of a substance to cause an adverse effect. ***Risk assessment*** is the process whereby exposure, hazard and risk are determined.

The ***hazard*** needs to be identified from human epidemiology, animal toxicity studies or *in vitro* studies. ***Dose–response relationships*** will also be determined from this information. For most chemicals a NOAEL or LOAEL can be determined. ***Exposure assessment*** and other aspects of risk assessment includes use of biomarkers of exposure, response and susceptibility and physico-chemical characteristics are important pieces of information.

Risk characterization involves integrating all the information and calculating parameters such as acceptable daily intake (ADI), tolerable daily intake (TDI) or threshold limit value (TLV) using the NOAEL and a safety or uncertainty factor. This is typically 100 (10 for species extrapolation, 10 for human variability). For carcinogens different models will be used to those

exhibiting a threshold for effect (e.g. one-hit, multi-hit). Carcinogenicity testing requires lifetime studies *in vivo* in large numbers of animals often including the maximum tolerated dose.

Questions

Q1. Indicate which of the following are true or false:

a acute toxicity studies are primarily for the determination of mutagenicity
b sub-chronic toxicity tests are for the measurement of dose response
c ecotoxicity studies may utilize tests in *Daphnia*
d teratogenicity tests are part of reproductive toxicity studies.

Q2. Which of the following are important in risk assessment:

a exposure level or dose
b hazard
c NOAEL
d benefit
e ADI
f cost
g TLV.

SHORT ANSWER QUESTIONS

Q3. Give four of the seven questions which should be asked before a toxicity study is carried out.

Q4. List the four types of epidemiological study.

Q5. Define the three Rs in relation to toxicity testing.

Q6. Define risk and hazard.

Bibliography

Alternatives in Animal Alternatives for Safety and Efficacy Testing, H. Salem and, S. A. Katz (Eds), revised edition (1997), Taylor & Francis. Covers the alternatives to animals for testing, minimizing numbers and discomfort, maximizing the information received.

AUER, C. M. and FIELDER, R. J. (1999) chapter 72, in *General and Applied Toxicology*, Ballantyne, B., Marrs, T. and Syversen, T. L. M. (Eds), 2nd edition, Basingstoke: Macmillan.

Biomarkers and Risk Assessment. Environmental Health Criteria 155, WHO, Geneva, 1993.

BROWN, V. K. (1988) *Acute and Sub-acute Toxicology*, London: Edward Arnold.

Chemical Risk Assessment, Training Module No. 3 UNEP/IPCS 1999, Geneva: WHO. An accessible source of information and teaching aid which can be obtained free from the IPCS, Geneva (ipcsmail@who.ch)

ECOBICHON, D. J. (1997) *The Basis of Toxicity Testing (Pharmacology and Toxicology)*, 2nd edition, CRC Press.

FAIRHURST, S. (1999) chapter 67, in *General and Applied Toxicology*, Ballantyne, B., Marrs, T. and Syversen, T. L. M. (Eds) 2nd edition, Basingstoke: Macmillan.

General and Applied Toxicology, Ballantyne, B., Marrs, T. and Syversen, T. L. M. (Eds), 2nd edition, 1999, Basingstoke: Macmillan. Various chapters in Part 8.

GRIFFIN, J. P. (1985) Predictive value of animal toxicity studies, *ATLA*, **12**, 163.

Hazardous chemicals in human and environmental health (2000) IPCS, Geneva: WHO. A simple but informative concise book, which can be obtained free from the IPCS, Geneva (ipcsmail@who.ch)

HEUVEL, M. J. Van Den, DAYAN, A. D. and SHILLAKER, R. O. (1987) Evaluation of the BTS approach to the testing of substances and preparations for their acute toxicity, *Human Toxicology*, **6**, 279.

HOMBURGER, F. (Ed.) (1983–) *Safety Evaluation and Regulation of Chemicals*, 3 vols, Basel: Karger.