PART I

CANCER AND HUMAN HEALTH RISK ASSESSMENT

Introductory remarks

Lutz Edler and Christos P. Kitsos

In his comprehensive volume on environmental risk assessment, Paustenbach (2002) characterizes risk assessment as the description of the likelihood of adverse or unwanted responses to exposure and states as its goal the estimation of that likelihood after assembling and assessing all scientific information regarding toxicology, human experience, and environmental fate and exposure. It was not until the middle of the twentieth century that serious concern was expressed over the risks caused by synthetically manufactured chemicals and radiation, at least in industrialized countries (For a historical overview, from a US perspective, see the first chapter of Paustenbach's book.). Before that time infectious diseases had been the most feared threats to human health.

From the beginning of the risk assessment of chemicals in food and the environment, cancer risk assessment played the lead role in the development of quantitative methods. The very birth of cancer risk assessment was motivated and guided by the strong wish of most risk managers and regulators to determine threshold doses/exposures. It was thought that a specific chemical should possess an absolute threshold value below which an exposure should not elicit an adverse health effect in humans. Setting a regulatory exposure limit at that threshold (or, when conservatively accounting for uncertainty, below it) would suffice to exclude any risk from humans. Consequently, research concentrated on the estimation of such a hypothesized threshold levels. The NOAEL (no observed adverse effect level) appeared as *the* natural estimate of that level. Therefore, experimental data were screened, either from human studies or from animal dose-response experiments, doses were

Recent Advances in Quantitative Methods in Cancer and Human Health Risk Assessment
Edited by L. Edler and C. Kitsos

identified at which no differences compared to the unexposed control were observed, and the largest such dose was declared as the NOAEL value. For a thorough scientific discussion of the threshold concept and its consequences for risk assessment practice we refer to the work of Wout Slob (e.g. Slob, 2002). More formally, the NOAEL has been defined as the largest administered dose in a dose-response experiment which is not statistically significant at a predefined significance level of, say, 5 %. As such, it is a statistical estimate and is subject to the rules of statistical inference like any other parameter derived from dose-response data (see Part III below). However, insisting on the existence of an absolute threshold value or 'zero risk' as well as attempting to find 'no risk' levels of exposure would have led into a blind alley. Among the first quantitative methods for risk assessment which did not assume a threshold dose was the Mantel and Bryan (1961) procedure estimating the so-called virtually safe dose (VSD), an exposure or dose below which at maximum a very low incidence (above background), say, one in a million, would occur.

Fortunately, risk assessment has matured from a method deciding between the existence and non-existence of risk into a scientifically oriented methodology which comprises a number of life science disciplines, such as biology, toxicology, pharmacology, acute and chronic medicine, physiology, as well as chemistry, physical chemistry, mathematics and statistics. This maturation is by no means complete. Witness the barrage of new techniques, new data and new theories created in life sciences which call for recognition, as well as the further development of methodology for the analysis and interpretation of data. On the other hand, there is still a strong need to provide for the risk manager and the public adequate advice on the information content of risk measures such as the NOAEL, benchmark dose level, reference dose levels, etc. Statistics plays an important role in the transmission of correct information.

At the beginning of the twenty-first century, the discipline of risk assessment presents itself as scientific methodology which should guide and improve decision making in the sector of public health and environment, concerning pollution, contamination and accidental exposure to substances which may be hazardous for humans. The outcome of risk assessment is now increasingly becoming the basis of risk management decisions, and this fact is slowly being recognized by the public as well as by the media. Scientific risk assessment is the only rational approach to judging risks and developing measures to resolve their dangerous consequences to humans and their health.

The basics of current risk assessment were laid out in the USA by the National Research Council (1983). Since then, there has been overwhelming agreement that risk assessment consist of the four components

- *hazard identification*, involving all available valuable information to determine whether or not a substance is hazardous;
- *dose-response assessment*, replaced now by *hazard characterization*, using available quantitative information to estimate the response at exposure levels;
- *exposure assessment*, determining the sources, routes and amount of exposure;

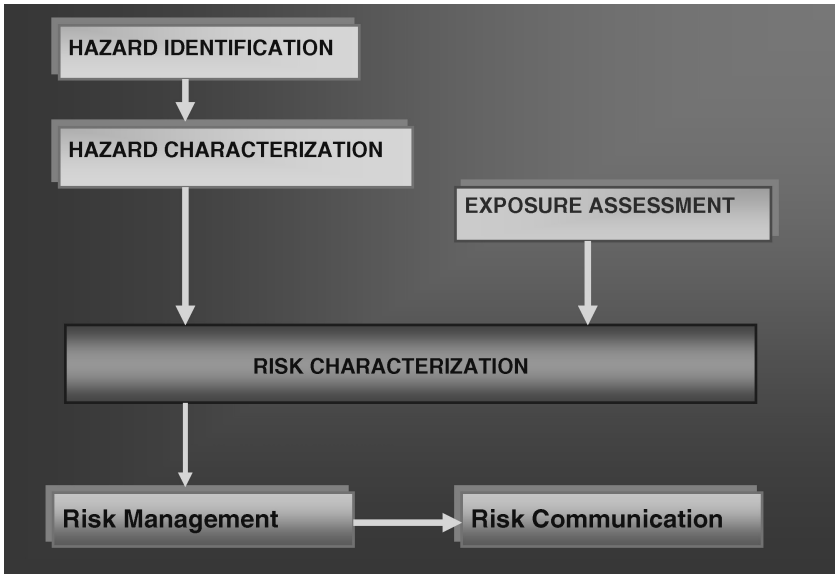

Figure I.1 Schematic view of the risk assessment paradigm.

- *risk characterization*, integrating the previous three components with a discussion of strengths and limitations; see Figure I.1 as well as Renwick *et al*. (2003).

The opening chapter of this volume, by V. J. Cogliano, provides a wealth of information and basic thoughts on the risk assessment paradigm and its present usage. That this risk assessment paradigm has not been adapted by different countries in the same way and that a large number of harmonization issues remain unresolved can be learned from an earlier article by the same author (Cogliano *et al.*, 2002).

There have been countless contributions on risk assessment in the scientific literature (in October 2004 PubMed returned more than 50 000 citations on 'risk assessment'). Less overwhelming is the flood of publications in the statistical literature (the CIS system of the American Statistical Association gave in its 2002 issue only about 300 citations on 'risk assessment'). The number of books on statistical and quantitative methods for risk assessment is not very large; see, for example, Travis (1989), Moolgavkar *et al.* (1999), Cox (2001), Chyczewski *et al.* (2002) and Cogliano *et al.* (1999b). For example, in the IARC monograph of Moolgavkar *et al.* (1999) 12 authors give a survey in nine chapters on the basic principles and the practice of quantitative estimation and prediction of cancer risks as used by the end of the 1990s at various regulatory agencies. This includes two chapters on epidemiological studies and two other chapters on mathematical modeling and statistics. In accordance with the editors primary scientific research

interest, biological carcinogenesis theories and data available until the mid nineties are considered with strong a focus on the two-stage clonal expansion model also known as the Moolgavkar–Venzon–Knudson model. Some of chapters below will address more recent biological concepts and their respective sources of biological data as well as additional aspects relevant for risk assessment. Other monographs have focused on standard methods of animal bioassay for carcinogenicity, discussed the use of the maximum tolerated dose, and developed dose-response models on the basis of standard models of carcinogenesis. In contrast to this, our text will concentrate primarily on human data and their modeling.

A different view on risk assessment has been taken by Vose (2000), who uses the term 'risk analysis'. Actually, he applies methods of probability theory and statistics for the assessment of the variability and uncertainty of processes. A major focus is on Monte Carlo simulation and the Bayesian approach. This work is a valuable source of parametric statistical distributions. An even broader view on risk assessment has been taken by authors who also cover economic and societal risks, e.g. the water supply system. According to Hames (1998), risk analysis combines 'concepts, tools and technologies that have been developed and practiced in such areas as design, development, system integration, prototyping, and construction of physical infrastructure; in reliability, quality control, and maintenance; and in the estimation of cost and schedule and in project management'. These authors consider systems modeling and optimization, thereby emphasizing technical systems, extreme events, uncertainty and sensitivity. Their prominent risk analysis questions are: What can go wrong? What is the likelihood that it will go wrong? What are the consequences? When doing risk assessment in the life sciences, it may be prudent not to neglect those developments of risk analysis in the engineering and, in particular, in the economic sciences.

We hope that this short introduction to a very complex field has given a global point of view on the problems to be discussed in the following chapters. We are aware that the particular order of the chapters we have chosen is subjective, but we feel it is the most convenient. The experienced reader will doubtless choose his own way of working through the book.

CHAPTER 1

Principles of Cancer Risk Assessment: The Risk Assessment Paradigm

Vincent James Cogliano

International Agency for Research on Cancer[1]

1.1 THE RISK ASSESSMENT PARADIGM

The United States National Research Council has outlined a series of steps that many agencies follow in analysing risk (National Research Council, 1983, 1994). Although this chapter is focused on cancer risk assessment, the risk assessment paradigm can apply equally to other adverse effects.

- *Hazard identification* asks whether exposure to an agent can cause an increase in the incidence of an adverse effect. This involves consideration of the epidemiologic evidence in humans, the evidence in experimental animals, and other relevant data, including studies of toxicokinetics and mechanisms.
- *Dose-response assessment* characterises the relationship between the dose of an agent and the incidence of an adverse health effect. The fundamental activity of dose-response assessment is modelling, which can range from simple linear (proportionality) models to more complex toxicokinetic and mechanism-based models. The purpose of the modelling is to extrapolate from the conditions observed in the epidemiologic and experimental studies to the conditions that are of interest in a human exposure situation. It is common for risk assessments to extrapolate from high doses to lower doses, from experimental animals to humans, from one route of exposure to another, or

[1]The views expressed in this chapter are those of the author and do not necessarily reflect the views or policies of the International Agency for Research on Cancer.

Recent Advances in Quantitative Methods in Cancer and Human Health Risk Assessment
Edited by L. Edler and C. Kitsos

from one pattern of exposure to another. There are also other extrapolations that have received less discussion.

- *Exposure assessment* determines the extent of human exposure to an agent. It identifies the *pathways* by which a population can be exposed, for example, through breathing contaminated air, eating contaminated fish, or swimming in a contaminated lake. For each pathway, an exposure assessment estimates how much of the agent the population is exposed to. This can depend on the magnitude, duration, and frequency of exposure.
- *Risk characterisation* describes the nature and magnitude of human risk, including attendant uncertainty. It integrates the hazard identification, dose-response assessment, and exposure assessment. It often emphasises a quantitative component that estimates risk by determining where an exposure estimate falls on the dose-response curve. The risk characterisation also includes a qualitative component that describes the assessment's strengths and limitations and identifies some key research needs.

The hazard identification and dose-response assessment steps are generally regarded as descriptive of the properties of a toxic agent. For example, health agencies may identify a hazard by describing a chemical agent as ''carcinogenic to humans'' or ''possibly carcinogenic to humans''. They may also estimate a dose-response curve for each agent (US Environmental Protection Agency (US EPA), 2003a). These steps are generally performed centrally, as toxic properties and dose-response curves are not thought of as varying from one location to another. In contrast to these centralised steps of a risk assessment, the exposure assessment and risk characterisation steps address a specific human population. These latter steps consider the various exposure pathways that may be involved, estimate the level of human exposure to the toxic agent, and then characterise the risk by determining where that level of exposure falls on the dose-response curve. Exposure assessment and risk characterisation often depend on specific knowledge of local populations and exposure conditions; accordingly, they are often performed at a more decentralised level.

In the future, this distinction between centralised hazard identification and dose-response assessment and decentralised exposure assessment and risk characterisation may break down as more is known about the distribution of genetic polymorphisms in different ethnic groups, and it becomes possible to estimate different dose-response curves for different segments of the population. Similarly, as more is known about the effects of one agent on the absorption, metabolism, and elimination of another agent, it may become possible to estimate different dose-response curves for different populations based on their level of exposure to other agents that modify the effects of the agent being assessed. At this time, however, because not enough is known about polymorphisms and the interactions of multiple agents to be able to estimate different dose-response curves for different populations and exposure conditions, dose-response assessment has generally estimated a single dose-response curve for each agent.

Because this chapter is focused on the assessment of potential carcinogens, the hazard identification and dose-response assessment steps are treated in greater detail.

1.2 HAZARD IDENTIFICATION

The data considered for hazard identification include human epidemiologic studies, long-term bioassays in experimental animals, and other relevant data on toxicokinetics and cancer mechanisms. Each source of data has a role in the hazard assessment. Epidemiologic studies can provide unequivocal evidence of carcinogenicity, but often are not sufficiently sensitive to identify a carcinogenic hazard except when the risk is high or involves an unusual form of cancer. For this reason, animal studies generally provide the best means of assessing potential risks to humans. To answer questions about the similarity of response between animals and humans, studies of toxicokinetics and mechanisms have been employed. Toxicokinetic studies investigate similarities and differences in absorption, distribution, metabolism, and elimination across species. Mechanistic studies can elucidate the chemical species and cellular processes involved in cancer initiation and development, including short-term tests to identify the potential for genetic toxicity.

The conclusion is often a judgement that weighs the evidence that an agent may or may not cause a specific adverse effect. For example, the International Agency for Research on Cancer uses the following terms to describe a cancer hazard:

- carcinogenic to humans (Group 1),
- probably carcinogenic to humans (Group 2A),
- possibly carcinogenic to humans (Group 2B),
- not classifiable as to its carcinogenicity to humans (Group 3),
- probably not carcinogenic to humans (Group 4).

Other agencies (for example, US EPA, 1986b, 2003a) have adopted similar descriptors for potential carcinogens.

A prominent trend in cancer hazard identification has been the greater availability and use of mechanistic information. Mechanistic studies attempt to identify the series of key precursor events that are involved in cancer development. This may permit a judgement about whether the mechanisms that cause cancer in experimental animals are likely to be operative in humans. Once the key precursor events involved in cancer development are identified, mechanistic studies also allow identification of sub-populations and life-stages that may be especially susceptible to cancer induced by the agent. In any use of mechanistic data, it is crucial to investigate whether more than one mechanism may be operating.

1.3 DOSE-RESPONSE ASSESSMENT

1.3.1 Different objectives, different data sets, different approaches

There have been two broad approaches to dose-response assessment. *Safety assessment* focuses on determining a safe dose for human exposure. This has generally involved identifying a dose that appears to be without adverse effects in the experimental studies and dividing that dose by several factors to arrive at a dose considered ''safe'' for human exposure. The other approach provides estimates of risk at low doses, by fitting dose-response models to data on the incidence of an adverse effect and using these models to estimate the incidence at lower levels of exposure. A large majority of cancer assessments have been of the latter variety, using models to estimate the cancer risk over a range of potential exposure levels.

The models used in dose-response assessment can be considered to belong to two broad classes. *Empirical models* are based on fitting a standard curve to data. (They are sometimes called *curve-fitting models*.) They are the least detailed models, describing the incidence of frankly observable adverse effects as a mathematical function of exposure to the agent. Such models are not based on specific knowledge about biological processes and mechanisms.

In contrast, *physiologically based toxicokinetic models* and *mechanism-based dose-response models* are based on specific knowledge about the biological processes and mechanisms leading from exposure to disease. (They are sometimes called *biologically based models*.) Toxicokinetic models simulate the relationship between external exposure and delivered dose at the target tissue. They can incorporate detailed information on an agent's absorption, distribution, metabolism, and elimination. Mechanism-based models simulate the relationship between cellular responses at the target tissue, precursor effects, and frankly observable adverse effects in the organism. These more detailed models require extensive knowledge of the disposition of a chemical in the body and the sequence of events leading to toxicity and disease.

Safety assessments, too, have begun to move toward increasing complexity. Formerly, dividing an apparently ''safe'' dose by a factor of 100 was believed to yield a dose fit for human exposure. This factor of 100 was later described as composed of a factor of 10 to account for differences between humans and experimental animals and another factor of 10 to account for differences between susceptible humans and the general population. More recently, these factors of 10 have in turn been described as being composed of a factor of 3 (being approximately the square root of 10) for toxicokinetic differences and another factor of 3 for toxicodynamic differences, or alternatively, a factor of 4 for toxicokinetic differences and a factor of 2.5 for toxicodynamic differences (assuming that these approximate factors can be refined to this level of precision). In addition, other factors have been used to account for differences between experimental conditions and human exposure conditions, for example, when a short-term animal study is used to infer a dose ''safe'' for lifetime human exposure, or when the experimental database lacks adequate studies of key adverse effects or studies that are pertinent to

exposure during pre-natal and post-natal development. Another recent trend in safety assessment has been to replace the various factors of 10 by *data-derived factors* that attempt to reflect more precisely the degree of difference between humans and experimental animals.

Dose-response models can be developed for different objectives. For example, a dose-response model can serve as a framework for organising and synthesising the available data and identifying research needs. In this case, a rather detailed model might be appropriate, using default assumptions and parameter values as placeholders to indicate where further research is needed. A different objective of dose-response modelling would be to provide projections of the range of risks that populations can face under different actual or hypothetical exposure conditions. In this case, a model that is well supported by data and not overly detailed might be most useful when a government agency needs to reassure a community that they do not face an unsafe exposure condition. In this case, the more speculative components of a research-oriented model may not be useful in a governmental determination of the risk that may be present.

1.3.2 Extrapolations in dose-response assessment

The fundamental objective of risk assessment, as with modelling generally, is extrapolation. Experimental data observed under specified conditions are extrapolated to human exposure conditions that may be similar in some respects and dissimilar in other respects from those in the experiments. Similarly, epidemiologic data observed in one human population exposed to certain conditions may be extrapolated to other human populations and other exposure conditions. The models developed as part of a dose-response assessment describe how these extrapolations are made and provide a basis for evaluating how well the extrapolations are supported by data. Several kinds of extrapolations are made in dose-response assessment.

- *Extrapolation from high doses to lower doses* is often a central objective of dose-response modelling. For example, studies in experimental animals generally use high doses to determine the effects that a specific agent can induce at some dose. This is often a practical necessity when fewer than, say, 100 animals are tested (and consequently a response rate of less than 1 percent cannot be observed), because government agencies are interested in ensuring that the risk to an exposed population is much less than 1 percent. This also happens when occupational studies are used to infer the potential for risk at environmental or occupational levels that are anticipated to be lower than those observed in prior occupational settings.
- *Extrapolation from experimental animals to humans* is also a common objective of dose-response modelling. Such extrapolation is often necessary just to compare modelling results from experiments involving different animal species. Although modelling can proceed using either animal dose metrics or human dose metrics, when the ultimate objective is a statement regarding

risks to exposed humans, animal dose metrics must be converted into an *equivalent human dose*.

- *Route-to-route extrapolation* is the application of a dose-response relationship estimated from studies involving one exposure route (for example, ingestion, inhalation, or dermal exposure) to another exposure route. This form of extrapolation has both a qualitative and a quantitative component. Qualitatively, one must make a judgement about whether the effects observed following exposure by one route are pertinent to another exposure route. Such a judgement is generally warranted when effects are observed at a site distant from the site of entry and when absorption can occur by either exposure route to give an internal dose of the agent. Such a judgement is sometimes specified as a default option in the absence of adequate data to the contrary.

- *Extrapolation from one pattern of exposure to another* involves using information derived from occupational exposure patterns or experimental exposure protocols (for example, a single daily dose five days a week for 24 months) to make inferences about other human exposure patterns that are likely to be different. This common extrapolation, used in most dose-response assessments, has typically been handled by the use of default dose metrics such as average daily dose or cumulative dose.

- *Extrapolation from small samples to larger populations* generally has not been explicitly discussed in risk assessments. The experimental uncertainty inherent in using small experimental samples can be described by the confidence bounds associated with the estimates from the experimental studies. An important qualitative concern, however, with the use of small samples in experimental studies is the need to discuss the experiment's power to detect adverse effects. For example, a response that is not statistically significant does not indicate a threshold; rather, it can be consistent with a small risk that falls below the experiment's power of detection. A similar concern is that small samples generally do not include adequate numbers to make reliable inferences about safe doses for susceptible individuals.

- *Extrapolation from relatively homogeneous groups to more heterogeneous populations* also generally has not been explicitly discussed in risk assessments. It can, however, strongly influence the shape of a dose-response curve. This is particularly important in view of the fact that many experimental studies use relatively homogeneous, genetically similar animals. Similar animals may have a tendency to respond at similar dose levels, while a more heterogeneous population would not necessarily all respond at the same dose level. This may be true even when the mechanism of action suggests a threshold dose below which adverse effects would not occur, because the threshold could vary across a heterogeneous population and the population dose-response curve could become indistinguishable from those associated with non-threshold models. In the human population, genetic and lifestyle

factors contribute to increased variation in susceptibility, which spreads the dose-response curve over a wider range of doses (Lutz, 1990).

- *Extrapolation to different life-stages* also generally has not been explicitly discussed in risk assessments. It is important to recognise, however, that many chronic experimental studies begin exposing the animals only after early-life developmental stages have passed, and that this can be a susceptible period of exposure for some diseases and agents. Similarly, many epidemiologic studies involve occupational exposure, which generally miss early-life and late-in-life stages. An analysis of the studies that investigated the effects of early-life exposure indicates that is a susceptible period of exposure that can lead to a higher incidence of cancer later in life (US EPA, 2003b). This suggests that the extrapolation to life-stages that are not covered by conventional experimental bioassays and epidemiologic studies deserves more explicit consideration.
- *Extrapolation from single-variable experiments to complex exposure situations* also generally has not been explicitly discussed in risk assessments. Conventional experimental study designs rightly try to minimise exposure to other toxic agents in order that the effects of the agent being studied can be unambiguously demonstrated. Human exposure circumstances, in contrast, involve a multitude of background and other exposures that can be toxicologically significant. Finding a safe dose in an otherwise unexposed population does not mean that that dose is safe when background and other common exposures are considered (US EPA, 2001b).

For those extrapolations that generally have not been addressed in past risk assessments, this implies that a particular source of human variation or uncertainty is not being considered. For example, the use of small samples to estimate rates of metabolism may mean that differences in metabolic polymorphisms are not being reflected in those estimates. Using the results of single-chemical experiments to estimate doses in more complex human exposure situations may mean that the effect of other exposures that can alter metabolic capacity (for example, consumption of alcoholic beverages or use of certain pharmaceuticals) is not being reflected in the dose estimates. Risk assessments could be improved by considering these factors when they may be important.

1.3.3 Safety assessment

Safety assessment is an approach to dose-response assessment that does not attempt to describe dose-response curves. Rather, its focus is to estimate an exposure level where there is little concern for adverse effects. Such a dose has been called, by various health agencies, an *acceptable daily intake* (ADI), a *tolerable daily intake* (TDI), a *minimal risk level* (MRL), or a *reference dose* (RfD). These approaches have generally not been used in cancer assessments, but EPA's recent draft final cancer guidelines (US EPA, 2003a) provide some discussion on subject.

Safety assessments generally are developed through a process that selects a critical precursor effect that occurs at the lowest doses, determines a dose where this

precursor is either not observed or occurs at a specified incidence rate (for example, 5 % or 10 %), and reduces this dose by a factor that reflects the differences between study conditions and conditions of human exposure. The resulting dose is characterised as one where there is little concern for any adverse effects, but without an explicit characterisation of risk levels either above or below that dose.

1.3.3.1 Developing a safety assessment

A dose where the critical precursor effects are not observed is often called a *no observable adverse effect level* (NOAEL). Generally this is one of the doses in the experimental study. More recently, modelling has been used to derive a dose (sometimes called a *benchmark dose*) that would induce a specified observable level of adverse effects (for example, 5 % or 10 % incidence). Such a dose is generally not one of the doses used in the study, but is an interpolated dose that falls between study doses.

Because the study conditions associated with the NOAEL (or benchmark dose) can differ from conditions of human exposure, it would not be scientifically defensible to assume that exposure at the NOAEL would pose little concern for humans. The following are some of the ways in which study conditions can differ from the conditions of human exposure:

- *Uncertainty in extrapolating from animals to humans.* Although there are many physiological similarities that apply across mammalian species, there are also many quantitative differences to consider when using an animal NOAEL to make inferences about the safety of human doses. These differences can be toxicokinetic or toxicodynamic in nature. In the absence of specific data on the agent being assessed, a reduction factor of 10 is often used as a default to cover the differences between animals and humans.
- *Variation from average humans to susceptible humans.* Experimental or epidemiologic studies rarely target biologically susceptible individuals. Susceptible humans could be adversely affected at lower doses than a general study population (especially a healthy adult worker population); consequently, general-population NOAELs are reduced so that they would apply to more susceptible individuals. In the absence of specific data on the agent being assessed, a reduction factor of 10 is generally used as a default to cover the differences between average humans and susceptible humans.
- *Uncertainty in extrapolating from one exposure regimen to another.* Sometimes the experimental studies involve less-than-lifetime exposure. Lifetime exposure can have effects that do not appear after shorter exposure durations; consequently, a safe dose for lifetime exposure may be lower than the safe dose for less-than-lifetime exposure observed in the experimental studies. Similarly, sometimes the experimental studies use intermittent dosing, and a safe dose for continuous exposure may be lower than a safe dose for intermittent exposure observed in the experimental studies. An adjustment to the NOAEL may be appropriate when human exposure conditions are

significantly different from the experimental exposure conditions. This may be complicated in the case of short-term intermittent human exposures to temporarily high levels. For example, one agent may act through a toxic metabolite that is preferentially formed at high doses, and another agent may act through toxic metabolites that are preferentially formed at lower doses. Depending on the toxicokinetics and the mechanisms of action, some short-term high-level exposures may pose significant risks, and others may pose much smaller risks.

Other factors are sometimes used to reflect a professional judgement about scientific uncertainties not explicitly treated above, including completeness of the overall database, minimal sample size, or poor exposure characterisation. An example would be an incomplete database that does not include adequate studies of some potential adverse effects. Another example would be a metal whose absorption or retention is reduced when ingested with food. If the animal study involved exposure in the feed while human exposure involves non-food pathways (for example, a child playing in contaminated dirt), then the animal NOAEL should be reduced to compensate for the higher absorption and retention expected in humans relative to the dietary exposure conditions in the animal study.

1.3.3.2 Characterising a safety assessment

The EPA's draft final cancer guidelines (US EPA, 2003a) note that a safety assessment is only a single point on the dose-response curve, and thus cannot by itself convey all the critical information present in the data from which it is derived. These guidelines list some points that should be discussed in safety assessments where cancer is an adverse effect being assessed:

- *Nature of the response.* Is the dose-response curve based on tumours or a precursor? Does it measure incidence or mortality? Is it a measure of lifetime risk, or was the study terminated earlier?
- *Level of the response.* What level of response was used to derive the safety assessment, for example, a 1 percent cancer risk, a 5 percent cancer risk, or a 10 percent change in a precursor measure? Obviously, different adjustment factors would apply depending on the level or nature of the response.
- *Nature of the study population.* Is the dose-response curve based on humans or animals? How large is the effective sample size? Is the study group representative of the general population, of healthy adult workers, or of a susceptible group? Are both sexes represented? Did exposure occur during a susceptible life-stage?
- *Slope of the observed dose-response curve.* At the low end of the dose-response curve, how rapidly does risk decrease as dose decreases?
- *Relationship of this assessment with those for other cancers.* For example, a safety assessment derived from data on male workers would not reflect the

implications of mammary tumours in female rats or mice. It may be important to use both endpoints to provide a comprehensive picture of potential human risks.

- *Extent of the overall cancer database.* Is the database limited to particular cancers, population segments, or life-stages?

A safety assessment derived by calculating a NOAEL and dividing it by several factors is sometimes described as being based on the assumption that there exists a *threshold dose* below which an individual does not respond. This characterisation is not entirely appropriate, as experimental data are not able to estimate thresholds with much confidence. A response that is not statistically significant does not indicate a threshold; rather, it can be consistent with a small risk that falls below the experiment's power of detection. For example, if 50 animals are exposed and none develops an adverse effect, this result is completely consistent with a risk of, say, 1 in 100; consequently, a dose associated with zero incidence (in this case, 0/50) does not indicate a threshold. In addition, statistical significance is affected by study design and sample size; consequently, an NOAEL in one study may show a statistically significant adverse effect in a superior study with a different design or a larger sample size.

1.3.4 Modelling to estimate risk at low doses

The most common approach to dose-response assessment for potential carcinogens has been to fit models to the available data on a specific agent and to use those models to estimate the cancer risk in an exposed population.

The models discussed in this paper can be developed to varying levels of complexity as they consider the progression from:

(measures of dose)

1. exposure to an agent at some concentration in an external medium, to
2. internal dose, to
3. delivered dose at the target tissue or cell, to

(measures of response)

4. cellular response at the target, to
5. observable precursor effects, to
6. frankly observable adverse effects in the organism.

Models of the relationship between external exposure and either internal dose or delivered dose are discussed in Section 1.3.4.1. Models of the relationship between some measure of dose and some measure of response, at either the cellular, organ, or organism level are discussed in Section 1.3.4.2.

1.3.4.1 Dose modelling

Dose models consider the progression from exposure to an agent at some concentration in an external medium, through various intermediate measures of internal dose, to the delivered dose at the target tissue or cell.

Dose metrics

Dose-response assessment begins by determining an appropriate *dose metric* for each adverse effect that can be attributed to the agent. There are several components to a dose metric:

- *The agent.* Typically, this can be the administered agent, or it can be a metabolite or a reaction product that is produced internally.
- *The proximity to the target site.* Typically, this can be the administered dose, or it can be an internal dose such as the concentration of the agent in the blood or the concentration at some organ in the body.
- *A time component.* Typically can be the average dose, but it could also be the peak dose, the cumulative dose, or the average body burden.

The selection of an appropriate dose metric considers what is known about the agent and the mechanisms by which it produces each adverse effect.

The US EPA's Guidelines for Exposure Assessment contain useful definitions for different forms of dose (US EPA 1992a). *Exposure* is the contact of an agent with the outer boundary of an organism. *Exposure concentration* is the concentration of a chemical in its transport or carrier medium at the point of contact. *Dose* is the amount of a substance available for interaction with metabolic processes or biologically significant receptors after crossing the outer boundary of an organism. *Potential dose* is the amount ingested, inhaled, or applied to the skin. *Applied dose* is the amount of a substance presented to the absorption barrier and available for absorption (although not necessarily having yet crossed the outer boundary of the organism). *Absorbed dose* is the amount crossing a specific absorption barrier (for example, the exchange boundaries of skin, lung, and digestive tract) through uptake processes. *Internal dose* is a more general term without respect to specific absorption barriers or exchange boundaries. *Delivered dose* is the amount of the chemical available for interaction by any particular organ or cell.

The goal of dose modelling is to estimate, to the extent possible, the delivered dose of the *active agent* at the *target organ* or *target cell*. Having the knowledge about the active agent and the target site, plus the data to be able to estimate a dose metric at this level of detail, would allow one to estimate the level of exposure that is associated with the first organ or cellular response that can lead to an adverse effect. When the delivered dose cannot be determined with confidence, modelling often proceeds by using another less specific dose metric, for example, the average daily dose of the administered agent.

Standardising different exposure patterns

Complex exposure patterns are often present in the epidemiologic or experimental animal studies that are used for dose-response assessment. For example, workers can be exposed to intense but intermittent exposures during the workday. Experimental animals can receive a daily dose by gavage (that is, via a tube inserted through the mouth to the forestomach), or they can receive intermittent doses throughout the day via their food or drinking water. Doses also can vary from one time to another, or there can be brief or prolonged periods between times that exposure occurs. The resulting internal dose depends on many variables, including exposure concentration, duration of exposure, frequency of exposure, and duration of recovery periods between exposures. Dose-response models typically simplify these complex exposure patterns by estimating some *summary measure* of exposure, for example, an average daily dose.

When enough information is available, *toxicokinetic modelling* (discussed below) is the preferred approach for determining a summary measure for a complex exposure pattern. Toxicokinetic models are suited for this purpose because they generally consider an exposure profile over time and they can incorporate information about metabolism and target tissue. For example, if there is sufficient information to identify the active agent(s), a toxicokinetic model can be useful in modelling the internal concentrations of the active agent, whether it be the administered agent or a metabolite or a reaction product. When there is sufficient information to identify the target site(s) of biological interaction, a toxicokinetic model can be useful in modelling the concentration of the active agent at the target site, for example, the concentration in the blood or the liver or the brain. When there is sufficient information to identify which summary measure of dose is most relevant, for example, average concentration or peak concentration or cumulative dose, a toxicokinetic model can be useful for investigating the relationship between external exposure and the relevant internal measure of dose.

When there is not enough information to construct and fit a toxicokinetic model, or when the nature of the problem does not justify the time and effort to develop one, less detailed *default approaches* are used to estimate a summary measure of dose. These approaches generally involve computing a daily dose or concentration that is averaged over the duration of the study (US EPA, 2003a). In occupational studies, cumulative dose is often used. These practices make an implicit assumption that each unit of exposure carries an equal risk, so exposures can be averaged or summed across the duration of the study. Such approaches generally acknowledge that default approaches become less reliable as the duration of exposure becomes shorter or the intermittent doses become more intense.

Timing of exposure can also be important. When there is a susceptible life-stage, doses during the susceptible period are not equivalent to doses at other times, and they should be analysed separately (US EPA 2003a, 2003b).

Cross-species extrapolation

A milligram in a mouse is not the same as a milligram in a human. An appropriate measure of dose would consider the body size of the species, the size of the target

organ, and the rates of pharmacokinetic processes. Large cross-species differences in dose can occur as a result of dissimilar rates of metabolism or elimination.

When enough information is available, toxicokinetic modelling is the preferred approach for extrapolating dose measures across species. This is because the models include many of the pertinent variables that differ across species, for example, organ sizes, blood volume, blood flow rates, intake rates, and metabolic rates. The values of many of these variables, with the exception of metabolic pathways and metabolic rates, are species-related (that is, independent of the agent being studied) and, hence, readily available to the model. Metabolic pathways and rates for humans, however, may not be available for toxic agents, due to ethical concerns about conducting experiments on humans to obtain such data.

When there is not enough information to construct and fit a toxicokinetic model, or when the nature of the problem doses not justify the time and effort to develop one, less detailed default approaches are used to extrapolate dose measures across species. One approach is to express experimental doses in terms of milligrams of the agent per kilogram of body weight per day (mg/kg-d) and to apply this metric directly to humans without further conversion; this makes an implicit assumption that mg/kg-d is the appropriate dose metric for cross-species extrapolation. Another approach in common use is to scale doses from animals to humans on the basis of equivalence of $mg/kg^{3/4}$ per day (US EPA, 1992b). This formula is supported by analyses of the allometric variation of key physiological parameters across mammalian species. Allometric scaling may be more appropriate for some toxic endpoints than for others. Other approaches that equate exposure concentrations in food or water are alternative versions of the same approach, because daily intakes of food or water are approximately proportional to the $3/4$ power of body weight. Most cancer assessments use one of these default approaches for scaling dose across species.

Route extrapolation

Route-to-route extrapolation (that is, extrapolation between ingestion, inhalation, and dermal exposures) has both qualitative and quantitative components. The qualitative aspect involves a judgement that effects observed by one exposure route would be expected to occur if the agent were administered by another exposure route. This is likely to be the case if the agent (or the active metabolite or reaction product) can reach the systemic circulation and be transported to distal sites in the body. The goal of route extrapolation is to estimate the doses by each exposure route that result in equivalent delivered doses at the target site.

This can involve some rather complicated modelling. For example, ingested doses experience what is known as the *first-pass effect*, in which the agent is passed to the liver and is subject to metabolism before reaching the systemic circulation; this does not occur for the inhalation or dermal exposure routes. When enough information is available, toxicokinetic modelling is the preferred approach for determining equivalent doses across exposure routes. This is because the systemic circulation is at the core of a toxicokinetic model, so if the model includes components to simulate intake and differences in metabolism across exposure

routes (including the first-pass effect), it can be used to estimate equivalent doses by different exposure routes.

When this is not the case, less detailed default approaches are sometimes used. Differences in intake rates (considering the fraction absorbed by each exposure route) can be used to estimate a rough equivalence across exposure routes. There are no generally applicable default methods for accounting for differences like the first-pass effect.

Toxicokinetic modelling

Toxicokinetic models simulate the relationship between applied dose and internal dose. They do this by tracing the flow of an administered chemical or its metabolites through the blood to different organs of the body (called *compartments*). Toxicokinetic models can range in level of detail from the simple to the complex. The least detailed *one-compartment model* is based on a simple relationship between intake, retention, and elimination from ''the body'' considered as a homogeneous whole. *Multi-compartment models* divide the body into several discrete components, for example, blood, liver, fatty tissue, and other tissue. More elaborate models simulate metabolism occurring in one or more compartments and trace the flow of both the administered chemical and its metabolites through the body.

Because toxicokinetic models are based on the flow of chemicals through the blood to different body compartments, they require information on the size of each compartment (known as the *volume of distribution*) and the rate of blood flow between compartments. These parameters, which generally do not depend on the chemicals being modelled, allow generic models to be developed for different animal strains or human populations. Chemical-specific rates of absorption and elimination are needed to adapt these generic models so that they can be used to describe the behaviour of a specific chemical in the body of a specific animal strain or human population. When metabolism occurs, chemical-specific descriptions of metabolism in different compartments are needed for an accurate model of the flow of the chemical and its metabolites through the body. The descriptions of metabolism are often complex; for example, there may be a maximal rate of metabolism that decreases as concentration increases.

Toxicokinetic models can improve dose-response assessment by revealing and describing complex relationships between applied and internal dose. Some dose-response relationships that are nonlinear at higher doses (for example, haemangiosarcomas induced in rats exposed to vinyl chloride) can be attributed to nonlinear toxicokinetics. These can involve, for example, saturation or induction of enzymatic processes at high doses.

Good modelling practice dictates that a discussion of confidence should accompany the presentation of modelling results. This would include consideration of model validation and sensitivity analysis, stressing the predictive performance of the model. Quantitative uncertainty analysis is important for evaluating the performance of a model. The uncertainty analysis covers questions of model

uncertainty (is the model based on the appropriate dose metrics?) and parameter uncertainty (do the data support unbiased and stable estimates of the model parameters?); see Section 1.3.5.

When toxicokinetic modelling is used to make extrapolations across species (for example, from experimental animals to human populations), a key issue to consider is whether toxicity should be expected in the same tissues in different species. The objective of cross-species toxicokinetic modelling is to use differences in organ size and blood flow and metabolic rates between animals and humans to predict differences in tissue concentrations between animals and humans. This assumes that the same target sites are expected in animals and humans, an assumption that does not hold true in all cases.

1.3.4.2 Dose-response modelling

Empirical dose-response modelling

Empirical models generally have been used to fit tumour incidence data. These models express tumour incidence as an increasing function of dose. The functional form of the model can be a relatively simple mathematical expression, such as a polynomial model, or it can be based on a standard statistical distribution. These models are often sufficiently flexible to fit a wide variety of dose-response data sets and may be either linear or nonlinear at lower doses. When there is a choice between standard statistical models, there is often a preferred model (for example, the US EPA's linearised multistage model) to provide some consistency across analyses and because different empirical models can give different results at low doses with no scientific basis to choose among them (this is because empirical models are generally not grounded in the underlying biology leading to the disease). Goodness of fit to the experimental observations is not, by itself, an effective means of discriminating among models that adequately fit the data (US EPA, 1986b).

When the observed tumour incidences do not increase with dose, it is often because there is a competing mechanism of toxicity at the high doses. In this case, the high dose is dropped from the analysis and an attempt is made to fit the remaining data with the model. This approach is used because there is more value in a model that describes low-dose behaviour than that at high doses.

Sometimes the observed dose-response data are highly nonlinear, and it is difficult to fit them with any standard mathematical or statistical model. This can happen if the toxicokinetics of the agent are highly nonlinear. In this case, the analysis can be improved by using a toxicokinetic model to explain the nonlinearity that is attributable to dose. Similarly, when there are large survival differences across exposure groups and time-to-tumour data are available, the analysis can be improved by using an empirical time-to-tumour model.

Empirical dose-response models also can be fitted to data on a tumour precursor. Quantitative data on precursors can be used in conjunction with, or in lieu of, data on tumour incidence to extend the dose-response curve to lower doses. This generally implies that the relationship between tumours and key tumour precursors is known (see below).

Mechanism-based dose-response modelling

Mechanism-based toxicodynamic models simulate the relationship between responses at the target tissue or cell and frankly observable adverse effects in the organism. Toxicodynamic modelling is potentially the most comprehensive way to account for the biological processes involved in the development of adverse effects. These models reflect the sequence of key precursor events that lead to the adverse effect. Toxicodynamic modelling can be used when there are sufficient data to ascertain the agent's mechanism(s) of action and quantitatively estimate model parameters that represent rates and other quantities associated with the key precursor events of the mechanism.

As mechanisms of carcinogenesis are better understood, mechanism-based dose-response models can become more complex. For example, the Armitage–Doll multistage model (Armitage and Doll, 1954) can be considered to be a mechanism-based model: the hypothesis at the time was that cancer was the result of a sequence of finite number of mutations; the rate of each mutation could be parameterised as the sum of a fixed background rate plus a term that was proportional to the dose of the carcinogen; and the functional form of the model is a mathematical description of the probability of each mutation occurring in sequence. The Moolgavkar–Venzon–Knudson two-stage clonal-expansion model reflects a hypothesis that cancer is the result of two mutations, while cells altered by the first mutation can undergo clonal expansion to present a larger number of target cells that are susceptible to alteration by the second mutation (Moolgavkar and Knudson, 1981).

The events considered in these dose-response models are cellular events: for example, mutations, cell division, and cell death. The model, therefore, will include parameters that describe the rates of these events: rates of the first and second mutations, the rate of cell division for cells altered by the first mutation, and the rate of death of such cells. Numerous variations of the same basic model can be obtained by different choices of how to make these parameters depend on dose. For example, the mutation rates can be linear functions of dose, or the mutation rates can be constant independent of dose, and the cell division rate can be an increasing function of dose.

The extent to which the parameterisation is informed by laboratory data is a major factor in how well a model can be expected to perform below the high-dose range where there are data. It is possible for different models to provide equivalent fits to the observed data but to diverge substantially in their projections at lower doses. When model parameters are estimated from tumour incidence data, it is often the case that different combinations of parameter estimates can yield similar results in the observed range. For this reason, the critical parameters should ideally be estimated from laboratory studies and not by curve-fitting to tumour incidence data (Portier, 1987). This approach reduces model uncertainty (see Section 1.3.5) and ensures that the model does not give answers that are biologically unrealistic. This approach also provides a robustness of results, where the results are not likely to change substantially if fitted to slightly different data (US EPA, 2003a).

Mechanism-based modelling can provide insight into the relationship between tumours and key precursor events. To illustrate, a model that includes cell

proliferation can be used to explore the relationship between changes in the rate of cell proliferation and the incidence of tumours – for example, what level of change in the cell proliferation rate is associated with a 10 percent increase in tumour incidence (Gaylor and Zheng, 1996). In this way, mechanism-based modelling can be used to select an appropriate precursor response level and characterise its relationship to tumour incidence.

1.3.4.3 Extrapolation below the range where data support dose-response modelling

Below the range of doses where data have been collected, alternative models that fit the data equally well in the observed range will begin to diverge. This form of *model uncertainty* (see Section 1.3.5) often applies at exposure levels that are of interest to decision-makers. In this case, health agencies often use a default procedure to provide an upper bound on the risk at lower doses where precise risk estimates may not be possible. One common approach is to use *linear extrapolation* to lower doses. Linear extrapolation has been justified in a variety of circumstances: as the most appropriate model when an agent is genotoxic; when human exposure or body burden is high and near the doses where key precursor events in the carcinogenic process can occur; or as a default approach that is not likely to underestimate risk at lower doses (US EPA, 2003a).

With linear extrapolation, a straight line is drawn from the modelled data to the origin (zero dose, zero risk, corrected for background). This implies a proportional (that is, linear) relationship between risk and dose at lower doses. (Note that this linear relationship is assumed only at lower doses; at higher doses, the dose-response curve generally does not have a linear shape.) The slope of the dose-response curve at lower doses (more commonly, an upper bound on this slope) is known as the *slope factor* and is used as an upper-bound estimate of risk per increment of dose at low doses. It is used to calculate an upper bound on risk over a range of exposure levels (US EPA, 2003a).

A different approach (one that provides less information) is not to estimate risk below the dose range supported by the available data. In this case, a *safety assessment* is often used to estimate a dose that might be safe for human exposure. This is the approach generally used for adverse effects other than cancer, but it is used in some cancer assessments. Although these safety assessments sometimes are described as being based on the assumption of biological thresholds, it is difficult to empirically distinguish a true threshold from a dose-response curve that is simply nonlinear at low doses (US EPA, 2003a).

1.3.5 Uncertainty and human variation

Model uncertainty refers to a lack of knowledge needed to determine whether the scientific theory on which the model is based is correct. There could be several alternative scientific theories, each giving rise to a different model, with no scientific basis for choosing among them. In risk assessment, model uncertainty is reflected in alternative choices for model structure, dose metrics, and extrapolation

approaches. Other sources of model uncertainty concern whether surrogate data are appropriate, for example, using data on adults to make inferences about children. The full extent of model uncertainty cannot be quantified; a partial characterisation can be obtained by comparing the results of alternative models. Model uncertainty can be expressed through comparison of separate analyses from each model (US EPA, 2003a).

Some aspects of model uncertainty that should be addressed in an assessment include the use of animal models as a surrogate for humans, the consequences of using linear or nonlinear extrapolation to estimate risks at lower doses, and the use of a study sample to make inferences about the human population, including susceptible population segments and life-stages (US EPA, 2003a).

Toxicokinetic and toxicodynamic models are generally premised on *site concordance* across species, modelling, for example, the relationship between administered dose and liver tissue concentrations to predict increased incidences of liver cancer. Site concordance, however, does not hold true in general, as there are numerous examples of an agent causing different cancers in different species (US EPA, 2003a).

Parameter uncertainty refers to a lack of knowledge about the values of a model's parameters. This leads to a distribution of values for each parameter. Common sources of parameter uncertainty include random measurement errors, systematic measurement errors, use of surrogate data instead of direct measurements, misclassification of exposure status, random sampling errors, and use of an unrepresentative sample. Most types of parameter uncertainty can be quantified by statistical analysis (US EPA, 2003a).

Human variation refers to person-to-person differences in biological susceptibility or in exposure. Although both human variation and uncertainty can be characterised as ranges or distributions, they are fundamentally different concepts. Uncertainty can be reduced by further research that supports a model or improves a parameter estimate, but human variation is a reality that can be better characterised, but not reduced, by further research. Fields other than risk assessment use ''variation'' or ''variability'' to mean dispersion about a central value, including measurement errors and other random errors that risk assessors address as uncertainty (US EPA, 2003a).

Probabilistic risk assessment has been used in exposure assessment to estimate human variation and uncertainty in lifetime average daily dose. Probabilistic methods can be used in this exposure assessment context because the pertinent variables (for example, concentration, intake rate, exposure duration, and body weight) have been identified, their distributions can be observed, and the formula for combining the variables to estimate the lifetime average daily dose is well defined (US EPA, 1992a). Similarly, probabilistic methods can be applied in dose-response assessment when there is an understanding of the important parameters and their relationships, such as identification of the key determinants of human variation (for example, metabolic polymorphisms, hormone levels, and cell replication rates), observation of the distributions of these variables, and valid models for combining these variables. Specification of joint probability distributions is

appropriate when the variables are not independent of each other. With appropriate data, formal approaches to probabilistic risk assessment can be applied to provide insight into the overall extent and dominant sources of human variation and uncertainty. On the other hand, there is little value in using probabilistic analyses to combine results of incompatible models or to bypass default options with alternatives that have no greater validity (National Research Council, 1994). It is important to note that analyses that omit or underestimate some principal sources of variation or uncertainty could provide a misleadingly narrow description of the true extent of variation and uncertainty and give decision-makers a false sense of confidence in estimates of risk (US EPA, 2003a).