4 Assessment of Toxicity

Introduction

This chapter discusses the methods used to evaluate the toxicity of a substance for the purpose of health risk assessment. Evaluation of toxicity involves two steps: hazard identification and dose-response evaluation. Hazard identification includes a description of the specific forms of toxicity (neurotoxicity, carcinogenicity, etc.) that can be caused by a chemical and an evaluation of the conditions under which these forms of toxicity might appear in exposed humans. Data used in hazard identification typically are derived from animal studies and other types of experimental work, but can also come from epidemiologic studies. Dose-response evaluation is a more complex examination of the conditions under which the toxic properties of a chemical might be evidenced in exposed people, with particular emphasis on the quantitative relationship between dose and toxic response. This step also includes study of how response can vary from one population subgroup to another.

Principles Of Toxicity Assessment

The basic principles guiding the assessment of a substance's toxicity are outlined in the *Guidelines for Carcinogen Risk Assessment* (EPA, 1987a) (currently being updated), *Chemical Carcinogens: A Review of the Science and Its Associated Principles* (OSTP, 1985), *Guidelines for Developmental Toxicity Risk Assessment* (EPA, 1991a) and have recently been summarized by the NRC (1993a). In addition, guidelines for the assessment of acute toxicity have recently been developed by NRC (1993b). The developmental-toxicity guidelines are used in this chapter to illustrate EPA's approach to health effects that involve noncancer end points. They constitute the first completed noncancer risk-assessment guidelines in a series that EPA plans to issue.

Hazard Identification

The first of the two questions typically considered in the assessment of chemical toxicity concerns the types of toxic effects that the chemical can cause. Can it damage the liver, the kidney, the lung, or the reproductive system? Can it cause birth defects, neurotoxic effects, or cancer? This type of *hazard* information is obtained principally through studies in groups of people who happen to be exposed to the chemical (epidemiologic studies) and through controlled laboratory experiments involving various animal species. Several other types of experimental data can also be used to assist in identifying the toxic hazards of a chemical.

Epidemiologic Studies

Epidemiologic studies clearly provide the most relevant kind of information for hazard identification, simply because they involve observations of human beings, not laboratory animals. That obvious and substantial advantage is offset to various degrees by the difficulties associated with obtaining and interpreting epidemiologic information. It is often not possible to identify appropriate populations for study or to obtain the necessary medical information on the health status of individuals in them. Information on the magnitude and duration of chemical exposure, especially that experienced in the distant past, is often available in only qualitative or semiquantitative form (e.g., the number of years worked at low, medium, and high exposure). Identifying other factors that might influence the health status of a population is often not possible. Epidemiologic studies are not controlled experiments. The investigator identifies an exposure situation and attempts to identify appropriate "control" groups (i.e., unexposed parallel populations), but the ease with which this can be accomplished is largely beyond the investigator's control. For those and several other

reasons, it is difficult or impossible to identify cause-effect relationships clearly with epidemiologic methods (OSTP, 1985).

It is rare that convincing causal relationships are identified with a single study. Epidemiologists usually weigh the results from several studies, ideally involving different populations and investigative methods, to determine whether there is a consistent pattern of responses among them. Some of the other factors that are often considered are the strength of the statistical association between a particular disease and exposure to the suspect chemical; whether the risk of the disease increases with increasing exposure to the suspect agent; and the degree to which other possible causative factors can be ruled out. Epidemiologists attempt to reach consensus regarding causality by weighing the evidence. Needless to say, different experts will weigh such data differently, and consensus typically is not easily achieved (IARC, 1987).

In the case of chemicals suspected of causing cancer in humans, expert groups ("working groups") are regularly convened by the International Agency for Research on Cancer (IARC) to consider and evaluate epidemiologic evidence. These groups have published their conclusions regarding the "degrees" of strength of the evidence on specific chemicals (sometimes chemical mixtures or even industrial processes when individual causative agents cannot be identified). The highest degree of evidence—sufficient evidence of carcinogenicity—is applied only when a working group agrees that the total body of evidence is convincing with respect to the issue of a cause-effect relationship.

No similar consensus-building procedure has been established regarding other forms of toxicity. Some epidemiologists disagree with IARC's cancer classification judgments in particular cases, and there seems to be even greater potential for scientific controversy regarding the strength of the epidemiologic evidence of non-cancer (e.g., reproductive, developmental, etc.) effects. There has been much less epidemiologic study of other toxic effects, in part because of lack of adequate medical documentation.

Animal Studies

When epidemiologic studies are not available or not suitable, risk assessment may be based on studies of laboratory animals. One advantage of animal studies is that they can be controlled, so establishing causation (assuming that the experiments are well conducted) is not in general difficult. Another advantage is that animals can be used to collect toxicity information on chemicals before their marketing, whereas epidemiologic data can be collected only after human exposure. Indeed, laws in many countries require that some classes of chemicals (e.g., pesticides, food additives, and drugs) be subjected to toxicity testing in animals before marketing. Other advantages of animal tests include the facts that

- The quantitative relationship between exposure (or dose) and extent of toxic response can be established.

- The animals and animal tissues can be thoroughly examined by toxicologists and pathologists, so the full range of toxic effects produced by a chemical can be identified.

- The exposure duration and routes can be designed to match those experienced by the human population of concern.

But laboratory animals are not human beings, and this obvious fact is one clear disadvantage of animal studies. Another is the relatively high cost of animal studies containing enough animals to detect an effect of interest. Thus, interpreting observations of toxicity in laboratory animals as generally applicable to humans usually requires two acts of extrapolation: interspecies extrapolation and extrapolation from high test doses to lower environmental doses. There are reasons based on both biologic principles and empirical observations to support the hypothesis that many forms of biologic responses, including toxic responses, can be extrapolated across mammalian species, including

Homo sapiens, but the scientific basis of such extrapolation is not established with sufficient rigor to allow broad and definitive generalizations to be made (NRC, 1993b).

One of the most important reasons for species differences in response to chemical exposures is that toxicity is very often a function of chemical metabolism. Differences among animal species, or even among strains of the same species, in metabolic handling of a chemical, are not uncommon and can account for toxicity differences (NRC, 1986). Because in most cases information on a chemical's metabolic profile in humans is lacking (and often unobtainable), identifying the animal species and toxic response most likely to predict the human response accurately is generally not possible. It has become customary to assume, under these circumstances, that in the absence of clear evidence that a particular toxic response is not relevant to human beings, any observation of toxicity in an animal species is potentially predictive of response in at least some humans (EPA, 1987a). This is not unreasonable, given the great variation among humans in genetic composition, prior sensitizing events, and concurrent exposures to other agents.

As in the case of epidemiologic data, IARC expert panels rank evidence of carcinogenicity from animal studies. It is generally recognized by experts that evidence of carcinogenicity is most convincing when a chemical produces excess malignancies in several species and strains of laboratory animals and in both sexes. The observation that a much higher proportion of treated animals than untreated (control) animals develops malignancies adds weight to the evidence of carcinogenicity as a result of the exposure. At the other extreme, the observation that a chemical produces only a relatively small increase in incidence of mostly benign tumors, at a single site of the body, in a single species and sex of test animal does not make a very convincing case for carcinogenicity, although any excess of tumors raises some concern.

EPA combines human and animal evidence, as shown in Table 4-1, to categorize evidence of carcinogenicity; the agency's evaluations of data on individual carcinogens generally match those of IARC. For noncancer health effects, EPA uses categories like those outlined in Table 4-2. Animal data on other forms of toxicity are generally evaluated in the same way as carcinogenicity data, although this classification looks at hazard identification (qualitative) and dose-response relationships (quantitative) together. No risk or hazard ranking schemes similar to those used for carcinogens have been adopted.

The hazard-identification step of a risk assessment generally concludes with a qualitative narrative of the types of toxic responses, if any, that can be caused by the chemical under review, the strength of the supporting evidence, and the scientific merits of the data and their value for predicting human toxicity. In addition to the epidemiologic and animal data, information on metabolism and on the behavior of the chemical in tissues and cells (i.e., on its mechanism of toxic action) might be evaluated, because clues to the reliability of interspecies extrapolation can often be found here.

Identifying the potential of a chemical to cause particular forms of toxicity in humans does not reveal whether the substance poses a risk in specific exposed populations. The latter determination requires three further analytic steps: emission characterization and exposure assessment (discussed in Chapter 3), dose-response assessment (discussed next), and risk characterization (discussed in Chapter 5).

Dose-Response Assessment

In the United States and many other countries, two forms of dose-response assessment involving extrapolation to low doses are used, depending on the nature of the toxic effect under consideration. One form is used for cancer, the other for toxic effects other than cancer.

Toxic Effects Other Than Cancer

For all types of toxic effects other than cancer, the standard procedure used by regulatory agencies for evaluating the dose-response aspects of toxicity involves identifying the highest exposure among all the available experimental studies at which no toxic effect was observed, the "no-observed-effect level" (NOEL) or "no-observed-adverse-effect level" (NOAEL). The difference between the two values is related to the definition of adverse effect. The NOAEL is the highest exposure at which there is no statistically or biologically significant increase in the frequency of an adverse effect when compared with a control group. A similar value used is the lowest-observed-adverse-effect level (LOAEL), which is the lowest exposure at which there is a significant increase in an observable effect. All are used in a similar fashion relative to the regulatory need. The NOAEL is more conservative than the LOAEL (NRC, 1986).

For example, if a chemical caused signs of liver damage in rats at a dosage of 5 mg/kg per day, but no observable effect at 1 mg/kg per day and no other study indicated adverse effects at 1 mg/kg per day or less, then 5 mg/kg per day would be the LOAEL and 1 mg/kg per day would be the NOAEL under the conditions tested in that study. For human risk assessment, the ratio of the NOAEL to the estimated human dose gives an indication of the margin of safety for the potential risk. In general, the smaller the ratio, the greater the likelihood that some people will be adversely affected by the exposure.

The uncertainty-factor approach is used to set exposure limits for a chemical when there is reason to believe that a safe exposure exists; that is, that its toxic effects are likely to be expressed in a person only if that person's exposure is above some minimum, or threshold. At exposures below the threshold, toxic effects are unlikely. The experimental NOAEL is assumed to approximate the threshold. To establish limits for human exposure, the experimental NOAEL is divided by one or more uncertainty factors, which are intended to account for the uncertainty associated with interspecies and intraspecies extrapolation and other factors. Depending on how close the experimental threshold is thought to be to the exposure of a human population, perhaps modified by the particular conditions of exposure, a larger or smaller uncertainty factor might be required to ensure adequate protection. For example, if the NOAEL is derived from high-quality data in (necessarily limited groups of) humans, even a small safety factor (10 or less) might ensure safety, provided that the NOAEL was derived under conditions of exposure similar to those in the exposed population of interest and the study is otherwise sound. If, however, the NOAEL was derived from a less similar or less reliable laboratory-animal study, a larger uncertainty factor would be required (NRC, 1986).

There is no strong scientific basis for using the same constant uncertainty factor for all situations, but there are strong precedents for the use of some values (NRC, 1986). The regulatory agencies usually require values of 10,100, or 1,000 in different situations. For example, a factor of 100 is usually applied when the NOAEL is derived from chronic toxicity studies (typically 2-year studies) that are considered to be of high quality and when the purpose is to protect members of the general population who could be exposed daily for a full lifetime (10 to account for interspecies differences and 10 to account for intraspecies differences).

Using the NOAEL/LOAEL/uncertainty-factor procedure yields an estimate of an exposure that is thought to "have a reasonable certainty of no harm." Depending on the regulatory agency involved, the resulting estimate of "safe" exposure can be termed an acceptable daily intake, or ADI (Food and Drug Administration, FDA); a reference dose, or RfD (EPA); or a permissible exposure level, or PEL (Occupational Safety and Health Administration, OSHA). For risk assessments, the dose received by humans is compared with the ADI, RfD, or PEL to determine whether a health risk is likely.

The requirement for uncertainty factors stems in part from the belief that humans could be more sensitive to the toxic effects of a chemical than laboratory animals and the belief that variations in sensitivity are likely to exist within the human population (NRC, 1980a). Those beliefs are plausible, but the magnitudes of interspecies and intraspecies differences for every chemical and toxic end point are not often known. Uncertainty factors are intended to accommodate scientific uncertainty, as well as uncertainties about dose delivered, human variations in sensitivity, and other matters (Dourson and Stara, 1983).

EPA's approaches to risk assessment for chemically induced reproductive and developmental end points rely on the threshold assumption. The EPA (1987a) guidelines for health-risk assessment for suspected developmental toxicants states that, "owing primarily to a lack of understanding of the biological mechanisms underlying developmental toxicity, intra/interspecies differences in the types of developmental events, the influence of maternal effects on the dose-response curve, and whether or not a threshold exists below which no effect will be produced by an agent," many developmental toxicologists assume a threshold for most developmental effects, because "the embryo is known to have some capacity for repair of the damage or insult" and "most developmental deviations are probably multifactorial."

EPA (1988a,b) later proposed guidelines for assessing male and female reproductive risks that incorporate the threshold default assumption "usually assumed for noncarcinogenic/nonmutagenic health effects," as well as the agency's new RfD approach to deriving acceptable intakes. The RfD is obtained as described above. The total adjustment or uncertainty factor referred to in the proposed guidelines for use in obtaining an RfD from toxicity data "usually ranges" from 10 to 1,000. The adjustment incorporates (as needed) uncertainty factors ("often" 10) for "(1) situations in which the LOAEL must be used because a NOAEL was not established, (2) interspecies extrapolation, and (3) intraspecies adjustment for variable sensitivity among individuals." An additional modifying factor may be used to account for extrapolating between exposure durations (e.g., from acute to subchronic) or for NOAEL-LOAEL inadequacy due to scientific uncertainties in the available database.

EPA's 1992 revision of its guidelines for developmental-toxicity risk assessment state that "human data are preferred for risk assessment" and that the "most relevant information" is provided by good epidemiologic studies. When these data are not available, however, reproductive risk assessment and developmental-agent risk assessment, according to EPA, are based on four key assumptions:

- An agent that causes adverse developmental effects in animals will do so in humans, with sufficient exposure during development, although the types of effects might not be the same in humans as in animals.

- Any significant increase in any of the expressions of developmental toxicants (e.g., death, structural abnormalities, growth alterations, and functional deficits) indicates a likelihood that the agent is a developmental hazard.

- Although the types of effects in humans and animals might not be the same, the use of the most sensitive animal species to estimate human hazards is justified.

- A threshold is assumed in dose-response relationships on the basis of current knowledge, although some experts believe that current science does not fully support this position.

The new guidelines state that "the existence of a NOAEL in an animal study does not prove or disprove the existence or level of a biological threshold." The guidelines also address statistical deficiencies and improvements in the NOAEL-based uncertainty-factor approach (Crump, 1984; Kimmel and Gaylor, 1988; Brown and Erdreich, 1989; Chen and Kodell, 1989; Gaylor, 1989; Kodell et al., 1991a). The guidelines also discuss EPA's plans to move toward a more quantitative "benchmark dose" (BD) for risk assessment for developmental end points "when sufficient data are available"; the BD approach would be consistent with the uncertainty-factor approach now in use (EPA, 1991a). Like the NOAEL and LOAEL, the BD is based on the most sensitive developmental effect observed in the most appropriate or most sensitive mammalian species. It would be derived by modeling the data in the observed range, selecting an incidence rate at a preset low observed response (e.g., 1% or 10%), and determining the corresponding lower confidence limit on dose that would yield that level of excess response. A BD thus calculated would then be divided by uncertainty factors to derive corresponding acceptable intake (e.g., RfD) values (EPA, 1991a). Thus, the traditional uncertainty-factor approach is retained in the 1991 developmental-toxicity guidelines, as well as in the proposed BD approach. However, the new guidelines are unique, in that they emphasize both the possible effect of

interindividual variability in the interpretation of acceptable exposures and the improvements that biologically based models could bring to developmental risk assessment (EPA, 1991a):

> It has generally been assumed that there is a biological threshold for developmental toxicity; however, a threshold for a population of individuals may or may not exist because of other endogenous or exogenous factors that may increase the sensitivity of some individuals in the population. Thus, the addition of a toxicant may result in an increased risk for the population, but not necessarily for all individuals in the population. … Models that are biologically based should provide a more accurate estimation of low-dose risk to humans. … The Agency is currently supporting several major efforts to develop biologically based dose-response models for developmental toxicity risk assessment that include the consideration of threshold.

Cancer

For some toxic effects, notably cancer, there are reasons to believe either that no threshold for dose-response relationships exists or that, if one does exist, it is very low and cannot be reliably identified (OSTP, 1985; NRC, 1986). This approach is taken on the basis not of human experience with chemical-induced cancer, but rather of radiation-induced cancer in humans and radiologic theory of tissue damage. Risk estimation for carcinogens therefore follows a different procedure from that for noncarcinogens: the relationship between cancer incidence and the dose of a chemical observed in an epidemiologic or experimental study is extrapolated to the lower doses at which humans (e.g., neighboring population) might be exposed (e.g., due to emissions from a plant) to predict an excess lifetime risk of cancer—that is, the added risk of cancer resulting from lifetime exposure to that chemical at a particular dose. In this procedure, there is no "safe" dose with a risk of zero (except at zero dose), although at sufficiently low doses the risk becomes very low and is generally regarded as without publichealth significance.

The procedure used by EPA is typical of those used by the other regulatory agencies. The observed relationship between lifetime daily dose and observed tumor incidence is fitted to a mathematical model to predict the incidence at low doses. Several such models are in wide use. The so-called linearized multistage model (LMS) is favored by EPA for this purpose (EPA, 1987a). FDA uses a somewhat different procedure that nevertheless yields a similar result. An important feature of the LMS is that the dose-response curve is linear at low doses, even if it displays nonlinear behavior in the region of observation.

EPA applies a statistical confidence-limit procedure to the linear multistage no-threshold model to generate what is sometimes considered an upper bound on cancer risk. Although the actual risk cannot be known, it is thought that it will not exceed the upper bound, might be lower, and could be zero. The result of a dose-response assessment for a carcinogen is a potency factor. EPA also uses the term *unit risk factor* for cancer potency. This value is the plausible upper bound on excess lifetime risk of cancer per unit of dose. In the absence of strong evidence to the contrary, it is generally assumed that such a potency factor estimated from animal data can be applied to humans to estimate an upper bound on the human cancer risk associated with lifetime exposure to a specified dosage.

The dose-response step involves considerable uncertainty, because the shape of the dose-response curve at low doses is not derived from empirical observation, but must be inferred from theories that predict the shape of the curve at the low doses anticipated for human exposure. The adoption of linear models is based largely on the science-policy choice that calls for caution in the face of scientific uncertainty. Models that yield lower risks, indeed models incorporating a threshold dose, are plausible for many carcinogens, especially chemicals that do not directly interact with DNA and produce genetic alterations. For example, some chemicals, such as chloroform, are thought to produce cancers in laboratory animals as a result of their cell-killing effects and related stimulation of cell division. However, in the absence of compelling mechanistic data to support such models, regulators are reluctant to use them, because of a fear that risk will be understated. For other substances (e.g., vinyl chloride), evidence shows that the human cancer risk at low doses could be substantially higher than would be estimated by the usual procedures from animal data.

Models that yield higher potency estimates at lower doses than the LMS model might also be plausible, but are rarely used (Bailar et al., 1988).

New Trends In Toxicity Assessment

With respect to carcinogenic agents, two types of information are beginning to influence the conduct of risk assessment.

For any given chemical, a multitude of steps can occur between intake and the occurrence of adverse effects. Those events can occur dynamically over an extended period, in some cases decades. One approach to understanding the complex interrelationships is to divide the overall scheme into two pieces, the linkages between exposure and dose and between dose and response. *Pharmacokinetics* has often been used to describe the linkage between exposure (or intake) and dose, and *pharmacodynamics* to describe the linkage between dose and response. Use of the root *pharmaco* (for drug) reflects the origin of those terms. When applied to the study and evaluation of toxic materials, the corresponding terms might more appropriately be *toxicokinetics* and *toxicodynamics*.

Exploration of the use of pharmacokinetic data is especially vigorous. Risk assessors are seeking to understand the quantitative relationships between chemical exposures and target-site doses over a wide range of doses. Because the target-site dose is the ultimate determinant of risk, any nonlinearity in the relationship between administered dose and target-site dose or any quantitative differences in the ratio of the two quantities between humans and test animals could greatly influence the outcome of a risk assessment (which now generally relies on an assumed proportional relationship between administered and target doses). The problem of obtaining adequate pharmacokinetic data in humans is being attacked by the construction of physiologically based pharmacokinetic (PBPK) models, whose forms depend on the physiology of humans and test animals, solubilities of chemicals in various tissues, and relative rates of metabolism (NRC, 1989). Several relatively successful attempts at predicting tissue dose in humans and other species have been made with PBPK modeling, and greater uses of this tool are being encouraged by the regulatory community (NRC, 1987).

A second major trend in risk assessment stems from investigations indicating that some chemicals that increase tumor incidence might do so only indirectly, either by causing first cell-killing and then compensatory cell proliferation or by increasing rates of cell proliferation through mitogenesis. In either case, increasing cell proliferation rates puts cells at increased risk of carcinogenesis from spontaneous mutation. Until a dose of such a carcinogen sufficient to cause the necessary toxicity or intracellular response is reached, no significant risk of cancer can exist. Such carcinogens, or their metabolites, show little or no propensity to damage genes (they are nongenotoxic).

Tables

TABLE 4-1 Categorization of Evidence of Carcinogenicity

Group		Criteria for Classification
A	Human carcinogen	Sufficient evidence from epidemiologic studies
B	Probable human carcinogen (two subgroups)	Limited evidence from epidemiologic studies and sufficient evidence from animal studies (B1); *or* inadequate evidence from epidemiologic studies (or no data) and sufficient evidence from animal studies (B2)
C	Possible human carcinogen	Limited evidence from animal studies and no human data
D	Not classifiable as to human carcinogenicity	Inadequate human and animal data or no data
E	Evidence of noncarcinogenicity in humans	No evidence of carcinogenicity from adequate human and animal studies

SOURCE: Adapted from EPA, 1987a.

TABLE 4-2   Weight-of-Evidence Classification Methods for Noncancer Health Effects

Sufficient Evidence
The sufficient-evidence category includes data that collectively provide enough information to judge whether a human developmental hazard could exist within the context of dose, duration, timing, and route of exposure. This category includes both human and experimental-animal evidence.
Sufficient Human Evidence: This category includes data from epidemiologic studies (e.g., case-control and cohort studies) that provide convincing evidence for the scientific community to judge that a causal relationship is or is not supported. A case series in conjunction with strong supporting evidence may also be used. Supporting animal data might or might not be available.
Sufficient Experimental Animal Evidence or Limited Human Data: This category includes data from experimental-animal studies or limited human data that provide convincing evidence for the scientific community to judge whether the potential for developmental toxicity exists. The minimal evidence necessary to judge that a potential hazard exists generally would be data demonstrating an adverse developmental effect in a single appropriate, well-conducted study in a single experimental-animal species. The minimal evidence needed to judge that a potential hazard does not exist would include data from appropriate, well-conducted laboratory-animal studies in several species (at least two) that evaluated a variety of the potential manifestations of developmental toxicity and showed no developmental effects at doses that were minimally toxic to adults.
Insufficient Evidence
This category includes situations for which there is less than the minimal sufficient evidence necessary for assessing the potential for developmental toxicity, such as when no data are available on developmental toxicity, when the available data are from studies in animals or humans that have a limited design (e.g., small numbers, inappropriate dose selection or exposure information, or other uncontrolled factors), when the data are from a single species reported to have no adverse developmental effects, or when the data are limited to information on structure/activity relationships, short-term tests. pharmacokinetics, or metabolic precursors.

SOURCE: EPA, 1987a.

Bookshelf ID: NBK208246