EPA/630/P-03/001A
NCEA-F-0644A
February 2003
Draft Final
www.epa.gov/ncea/raf/cancer2003.htm

Draft Final Guidelines for Carcinogen Risk Assessment

Risk Assessment Forum
U.S. Environmental Protection Agency
Washington, DC

DISCLAIMER

This is a draft final document and does not, at this time, constitute U.S. Environmental Protection Agency policy. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

CONTENTS

List of Figures

1. INTRODUCTION

1.1. PURPOSE AND SCOPE OF THE GUIDELINES

These guidelines revise and replace the U.S. Environmental Protection Agency's (EPA's, or the Agency's) *Guidelines for Carcinogen Risk Assessment,* published in 51 FR 33992, September 24, 1986 (U.S. EPA, 1986a) and the 1999 draft guidelines (U.S. EPA, 1999a). They provide EPA staff and decisionmakers with guidance for developing and using risk assessments. They also provide basic information to the public about the Agency's risk assessment methods. These guidelines are used with other risk assessment guidelines that the Agency has developed, such as the *Guidelines for Mutagenicity Risk Assessment* (U.S. EPA, 1986b) and the *Guidelines for Exposure Assessment* (U.S. EPA, 1992a). Consideration of other Agency guidance documents is particularly important when procedures for evaluating specific target organ effects have been developed (e.g., assessment of thyroid follicular cell tumors, U.S. EPA, 1998a) or when there is a concern for a particular susceptible subpopulation for which the Agency has developed guidance, for example, *Guidelines for Developmental Toxicity Risk Assessment* (U.S. EPA, 1991a). These guidelines discuss hazards to children that may result from exposures during preconception and prenatal or postnatal development to sexual maturity. Similar guidelines exist for reproductive toxicant risk assessments (U.S. EPA, 1996a) and for neurotoxicity risk assessment (U.S. EPA, 1998b).

All of these guidelines should be consulted when conducting a risk assessment in order to ensure that information from studies on carcinogenesis and other health effects are considered together in the overall characterization of risk. This is particularly true in the case in which a precursor effect to tumor is also a precursor or endpoint of other health effects and is used in dose-response assessment. The overall characterization of risk is the basis for carrying out assessments of instances in which fetuses, infants, or children are at risk or disproportionately affected by economically significant Agency actions. Characterization for the latter purpose is outlined in the Agency guidance by the Office of Children's Health Protection to carry out Executive Order 13045, "Protection of Children From Environmental Health Risks and Safety Risks," issued on April 21, 1997.

The guidelines encourage both regularity in procedures to support consistency in scientific components of Agency decision making and innovation to remain up to date in scientific thinking. In balancing these goals, the Agency relies on established scientific peer review processes (U.S. EPA, 2000a). The guidelines incorporate basic principles and science policies based on evaluation of the currently available information. As more is discovered about

carcinogenesis, the need will arise to make appropriate changes in risk assessment guidance. The Agency intends to revise these guidelines when extensive changes are due. In the interim, the Agency intends to issue special reports, after appropriate peer review, to supplement and update guidance on single topics (e.g., U.S. EPA, 1991b). The consideration of new, peer-reviewed scientific understanding and data in an assessment is always consistent with the purposes of these guidelines.

These guidelines are intended as guidance only. They do not establish any substantive "rules" under the Administrative Procedure Act or any other law and will have no binding effect on EPA or any regulated entity, but instead represent a non-binding statement of policy. EPA believes that the guidelines represent a sound and up-to-date approach to cancer risk assessment, and the guidelines enhance the application of the best available science in EPA's risk assessments. However, EPA cancer risk assessments may be conducted differently than envisioned in the guidelines for many reasons, including (but not limited to) new information, new scientific understanding, or new science policy judgment. The science of risk assessment continues to develop rapidly, and specific components of the guidelines may become outdated or may otherwise require modification in individual settings. Use of the guidelines in future risk assessments will be based on decisions by EPA that approaches are suitable and appropriate in the context of those particular risk assessments. These judgments will be tested through peer review, and risk assessments will be modified to use different approaches if appropriate.

1.2. ORGANIZATION AND APPLICATION OF THE GUIDELINES

1.2.1. Organization

Publications by the Office of Science and Technology (OSTP, 1985) and the National Research Council (NRC) (NRC, 1983, 1994) provide information and general principles about risk assessment. Risk assessment uses available scientific information on the properties of an agent[1] and its effects in biological systems to provide an evaluation of the potential for harm as a consequence of environmental exposure. The 1983 and 1994 NRC documents organize risk assessment information into four areas: hazard identification, dose-response assessment, exposure assessment, and risk characterization. This structure appears in these guidelines, which additionally emphasize characterization of evidence and conclusions in each part of the assessment. In particular, the guidelines adopt the approach of the NRC's 1994 report in adding a dimension of characterization to the hazard identification step. Added to the identification of

[1] The term "agent" refers generally to any chemical substance, mixture, or physical or biological entity being assessed, unless otherwise noted (See Section 1.2.2 for a note on radiation.).

hazard is an evaluation of the conditions under which its expression is anticipated. The risk assessment questions addressed in these guidelines are:

- For hazard—Can the agent present a carcinogenic hazard to humans and, if so, under what circumstances?

- For dose response—At what levels of exposure might effects occur?

- For exposure—What are the conditions of human exposure?

- For risk—What is the character of the risk? How well do data support conclusions about the nature and extent of the risk?

The risk characterization process first summarizes findings on hazard, dose response, and exposure characterizations and then develops an integrative analysis of the whole risk case. It ends in a nontechnical risk characterization summary. The summary is a presentation for risk managers who may or may not be familiar with the scientific details of cancer assessment. It also provides information for other interested readers. The initial steps in the risk characterization process are to make building blocks in the form of characterizations of the assessments of hazard, dose response, and exposure. The individual assessments and characterizations are then integrated to arrive at risk estimates for exposure scenarios of interest. As part of the characterization process, explicit evaluations are made of the hazard and risk potential for susceptible populations, including children (U.S. EPA, 1995, 2000b).

There are two reasons for individually characterizing the hazard, dose response, and exposure assessments. One is that they are often done by different people than those who do the integrative analyses. The second is that there is very often a lapse of time between the conduct of hazard and dose-response analyses and the conduct of exposure assessment and integrative analysis. Thus, it is important to capture characterizations of assessments as the assessments are done to avoid the need to go back and reconstruct them. Finally, frequently a single hazard assessment is used by several programs for several different exposure scenarios. There may be one or several documents involved. "Integrative analysis" is a generic term. At EPA, the documents of various programs that contain integrative analyses have other names, such as the "Staff Paper," which discusses air quality criteria issues. In the following sections, the elements of these characterizations are discussed.

1.2.2. Application

The guidelines apply within the framework of policies provided by applicable EPA statutes and do not alter such policies. The guidelines cover assessment of available data. They do not imply that one kind of data or another is prerequisite for regulatory action concerning any agent. It is important to remember that when judging and considering the use of any data, the basic standard of quality, as defined by the EPA Information Quality Guidelines (U.S. EPA, 2002a), should be satisfied. It is very important that all analyses adhere to the basic standards of quality, including objectivity, utility, and integrity. Risk management applies directives in statutes, which may require consideration of potential risk or solely hazard or exposure potential, along with social, economic, technical, and other factors in decision making. Risk assessments may be used to support decisions, but in order to maintain their integrity as decision-making tools, they are not influenced by consideration of the social or economic consequences of regulatory action.

The assessment of risk from radiation sources is based on continuing examination of human data by the National Academy of Sciences/NRC in its series of numbered reports: "Biological Effects of Ionizing Radiation." Although the general principles of these guidelines apply to radiation risk assessments, their details are most focused on other kinds of agents. They do not attempt to guide the ongoing conduct of radiation risk assessment.

Not every EPA assessment has the same scope or depth. When a cancer risk assessment is influential information as defined in OMB and EPA information-quality guidelines (OMB, 2002; U.S. EPA, 2002a), EPA staff and decision makers should make sure that information-quality performance goals are satisfied. On the other hand, Agency staff often conduct screening-level assessments for priority setting or separate assessments of hazard or exposure for ranking purposes or to decide whether to invest resources in collecting data for a full assessment. Moreover, a given assessment of hazard and dose response may be used with more than one exposure assessment that may be conducted separately and at different times as the need arises in studying environmental problems in various media. The guidelines apply to these various situations in appropriate detail, given the scope and depth of the particular assessment. For example, a screening assessment may be based almost entirely on structure-activity relationships (SARs) and default options. As more data become available, assessments can replace or modify default options accordingly. These guidelines do not suggest that all of the kinds of data covered here be available for either assessment or decision making. The level of detail of an assessment is a matter of Agency management discretion regarding applicable decision making needs.

1.3. KEY FEATURES OF THE GUIDELINES

1.3.1. Use of Default Options

NRC (1994) reaffirmed the use of default options as "a reasonable way to cope with uncertainty about the choice of appropriate models or theory" (p. 104). It saw the need to treat uncertainty in a predictable way that is "scientifically defensible, consistent with the agency's statutory mission, and responsive to the needs of decision-makers" (p. 86). Accordingly, default options have a science component and a policy component.

Encouraging risk assessors to be receptive to new scientific information, NRC discussed the need for departures from default options when a "sufficient showing" is made. It called on EPA to articulate clearly its criteria for a departure so that decisions to depart from default options would be "scientifically credible and receive public acceptance" (p. 91). It was concerned that ad hoc departures would undercut the scientific credibility of a risk assessment. NRC envisioned that principles for choosing and departing from default options would balance several conflicting objectives, including "protecting the public health, ensuring scientific validity, minimizing serious errors in estimating risks, maximizing incentives for research, creating an orderly and predictable process, and fostering openness and trustworthiness" (p. 81).

NRC discussed two opposing principles for governing departures from default options. One suggested principle would evaluate a departure in terms of whether "it is scientifically plausible" and whether it "tends to protect public health in the face of scientific uncertainty" (p. 601). An opposing principle "emphasizes scientific plausibility with regard to the use of alternative models" (p.631). Reaching no consensus on a single approach, NRC recognized that developing criteria for departures is an EPA policy matter.

With increasing understanding becoming available, these guidelines adopt a view of default options that is consistent with EPA's mission to protect human health. Rather than viewing default options as the starting point from which departures may be justified by new scientific information, these guidelines view a critical analysis of the available information as the starting point from which a default option may be invoked if needed to address uncertainty or the absence of critical information. The primary goal of EPA actions is public health protection; accordingly, as an Agency policy, any default options used in the absence of scientific data to the contrary should be health protective (U.S. EPA, 1999b).

The basis for invoking a default option depends on the circumstances. Generally, if a gap in basic understanding exists or if agent-specific information is missing, a default option can be used. If agent-specific information is present but critical analysis reveals inadequacies, a default option can also be used. If critical analysis of agent-specific information is consistent with one

or more alternative models and with the default option, the alternative models and the default option are carried through the assessment and characterized for the risk manager. This latter case highlights the importance of extensive experimentation to support a conclusion about mode of action, including addressing the issue of whether alternative modes of action are also plausible. Section 2.5 provides a framework for critical analysis of mode of action information to address the extent to which the available information supports the hypothesized mode of action, whether alternative modes of action are also plausible, and whether there is confidence that the same inferences can be extended to populations and lifestages that are not represented among the experimental data.

Generally, these decisions strive to be "scientifically defensible, consistent with the agency's statutory mission, and responsive to the needs of decision-makers" (NRC, 1994, p. 86). Scientific defensibility would be evaluated through use of EPA's Science Advisory Board or other independent expert peer review panels to determine whether a consensus among knowledgeable scientists exists. Consistency with the Agency's statutory mission would consider whether the risk assessment overall supports EPA's mission to protect human health and safeguard the natural environment. Responsiveness to the needs of decisionmakers would take into account pragmatic considerations such as the nature of the decision; the required depth of analysis; the utility, time, and cost of generating new scientific data; and the time, personnel, and resources allotted to the risk assessment.

With a multitude of types of risk assessments and potential default options, it is neither possible nor desirable to specify step-by-step criteria for decisions to invoke a default option. A discussion of major default options appears in the Appendix. Screening-level assessments may more readily use default options, even worst-case assumptions, that would not be appropriate in a full-scale assessment. Some default options are conveniences that allow an analysis to proceed and may be easily used with minimal explanation. For example, a cross-species scaling factor is readily available as a default option if a toxicokinetic model is not used, and standard animal body weights are a further default option if actual body weights are not available.

When toxicokinetic or toxicodynamic models are developed, a quantitative uncertainty analysis would be useful for determining whether the model is sufficiently robust to support a decision. If insufficient data or understanding limit development of a robust model, an appropriate policy choice is to have a single preferred curve-fitting model for each type of data set. Many different curve-fitting models have been developed, and those that fit the observed data reasonably well may lead to several-fold differences in estimated risk at the lower end of the observed range. This presents a problem in providing assurance that risk estimates were not

obtained by choosing to present only those models that gave the most desired result. Another problem occurs when a multitude of alternatives are presented without sufficient context to make a reasoned judgment about the alternatives. This form of model uncertainty reflects primarily the availability of different computer models and not biological information about the agent being assessed or about carcinogenesis in general. In cases where curve-fitting models are used because the data are not adequate to support a toxicodynamic model, there generally would be no biological basis to choose among alternative curve-fitting models. In addition, goodness-of-fit to the experimental observations is not by itself an effective means of discriminating among models that adequately fit the data (OSTP, 1985). To provide some measure of consistency across different carcinogen assessments, EPA uses a standard curve-fitting procedure for tumor incidence data. Assessments that include a different approach should provide an adequate justification and compare their results with those from the standard procedure. Application of models to data should be conducted in an open and transparent manner.

1.3.2. Mode of Action

The use of mode of action[2] in the assessment of potential carcinogens is the main thrust of these guidelines. This area of emphasis arose because of the significant scientific breakthroughs that have developed concerning the causes of cancer induction. In the absence of mode of action information, EPA generally takes conservative (public health-protective) default positions regarding the interpretation of toxicologic and epidemiologic data: animal tumor findings are judged to be relevant to humans, and cancer risks are assumed to conform with low dose linearity. Elucidation of a mode of action for a particular cancer response in animals or humans is a data-rich determination. Significant information should be developed to ensure that a mode of action underlies the process leading to cancer at a given site.

Understanding of mode of action can be a key to identifying processes that may cause chemical exposures to differentially affect a particular population segment or lifestage. Some modes of action are anticipated to be mutagenic and are assessed with a linear approach for most, if not all, parts of the population. This is the mode of action of radiation and several other agents

[2] The term *"mode of action"* is defined as a sequence of key events and processes, starting with interaction of an agent with a cell, proceeding through operational and anatomical changes, and resulting in cancer formation. A *"key event"* is an empirically observable precursor step that is itself a necessary element of the mode of action or is a marker for such an element. Mode of action is contrasted with *"mechanism of action,"* which implies a more detailed understanding and description of events, often at the molecular level, than is meant by mode of action. The toxicokinetic processes that lead to formation or distribution of the active agent to the target tissue are considered in estimating dose but are not part of the mode of action as the term is used here. There are many examples of possible modes of carcinogenic action, such as mutagenicity, mitogenesis, inhibition of cell death, cytotoxicity with reparative cell proliferation, and immune suppression.

that are known carcinogens. Several mutagenic carcinogens are also in utero carcinogens. Other modes of action may be assessed with either linear or nonlinear[3] approaches after a rigorous analysis of available data under the guidance provided in the framework for mode of action analysis (see Section 2.5.3).

1.3.3. Weight of Evidence Narrative

The guidelines emphasize the importance of weighing all of the evidence in reaching conclusions about the human carcinogenic potential of agents. This is accomplished in a single step after assessing all of the individual lines of evidence, which is in contrast to the step-wise approach in the 1986 guidelines. Evidence considered includes tumor findings in humans and laboratory animals, an agent's chemical and physical properties, its SARs with other carcinogenic agents, and its activities in studies of carcinogenic processes. Data from human studies are generally preferred for characterizing human cancer hazard. However, all of the information discussed above could provide valuable insights into the possible mode(s) of action and likelihood of human cancer hazard and risk. The guidelines recognize the growing sophistication of research methods, particularly in their ability to reveal the modes of action of carcinogenic agents at cellular and subcellular levels as well as toxicokinetic processes.

Weighing of the evidence includes addressing not only the likelihood of human carcinogenic effects of the agent but also the conditions under which such effects may be expressed, to the extent that these are revealed in the toxicological and other biologically important features of the agent.

The weight of evidence narrative to characterize hazard summarizes the results of the hazard assessment and provides a conclusion with regard to human carcinogenic potential. The narrative explains the kinds of evidence available and how they fit together in drawing conclusions, and it points out significant issues/strengths/limitations of the data and conclusions. Because the narrative also summarizes the mode of action information, it sets the stage for the discussion of the rationale underlying a recommended approach to dose-response assessment.

---

[3]The term "*nonlinear*" is used here in a narrower sense than its usual meaning in the field of mathematical modeling. In these guidelines, the term "*nonlinear*" refers to threshold models (which show no response over a range of low doses that include zero) and some nonthreshold models (e.g., a quadractic model, which shows some response at all doses above zero). In these guidelines, a nonlinear model is one whose slope is zero at (and perhaps above) a dose of zero. A *low-dose-linear* model is one whose slope is greater than zero at a dose of zero. A low-dose-linear model approximates a straight line only at very low doses; at higher doses near the observed data, a low-dose-linear model can display curvature. The term "*low-dose-linear*" is often abbreviated "linear," although a low-dose-linear model is not linear at all doses. Use of nonlinear approaches does not imply a biological threshold dose below which the response is zero. Estimating thresholds can be problematic; for example, a response that is not statistically significant can be consistent with a small risk that falls below an experiment's power of detection.

In order to provide some measure of clarity and consistency in an otherwise free-form, narrative characterization, standard descriptors are used as part of the hazard narrative to express the conclusion regarding the weight of evidence for carcinogenic hazard potential. There are five recommended standard hazard descriptors: "*carcinogenic to humans*," "*likely to be carcinogenic to humans*," "*suggestive evidence of carcinogenic potential*," "*inadequate information to assess carcinogenic potential*," and "*not likely to be carcinogenic to humans*." Each standard descriptor may be applicable to a wide variety of data sets and weights of evidence and is presented only in the context of a weight of evidence narrative. Furthermore, more than one conclusion may be reached for an agent. For instance, using a descriptor in context, a narrative could say that an agent is *likely to be carcinogenic* by inhalation exposure and *not likely to be carcinogenic* by oral exposure.

1.3.4. Dose-response Assessment

Dose-response assessment evaluates potential risks to humans at particular exposure levels. The approach to dose-response assessment for a particular agent is based on the conclusion reached as to its potential mode(s) of action for each tumor type. Because an agent may induce multiple tumor types, the dose-response assessment includes an analysis of all tumor types, followed by an overall synthesis that includes the consistency of risk estimates across tumor types, the strength of the mode of action information of each tumor type, and the anticipated relevance of each tumor type to humans, including susceptible populations and lifestages (e.g., childhood).

Dose-response assessment for each tumor type is performed in two steps: assessment of observed data to derive a point of departure (POD),[4] followed by extrapolation to lower exposures to the extent that is necessary. Data from human studies, of sufficient quality, are generally preferred for estimating risks. When animal studies are the basis of the analysis, the estimation of a human-equivalent dose should utilize toxicokinetic data to inform cross-species dose scaling if appropriate and if adequate data are available. Otherwise, default procedures should be applied. For oral dose, based on current science, an appropriate default option is to scale daily applied doses experienced for a lifetime in proportion to body weight raised to the 3/4 power. For inhalation dose, based on current science, an appropriate default methodology estimates respiratory deposition of particles and gases and estimates internal doses of gases with

[4] A "*point of departure*" (POD) marks the beginning of extrapolation to lower doses. The POD is an estimated dose (expressed in human-equivalent terms) near the lower end of the observed range, without significant extrapolation to lower doses.

different absorption characteristics. When toxicokinetic modeling (see Section 3.1.2) is used without toxicodynamic modeling (see Section 3.2.2), the dose-response assessment develops and supports an approach for addressing toxicodynamic equivalence, perhaps by retaining some of the cross-species scaling factor (see Section 3.1.3). Guidance is also provided for adjustment of dose from adults to children (see Section 4.3.1).

Response data on effects of the agent on carcinogenic processes are analyzed (nontumor data) in addition to data on tumor incidence. If appropriate, the analyses of data on tumor incidence and on precursor effects may be combined, using precursor data to extend the dose-response curve below the tumor data. Even if combining data is not appropriate, study of the dose response for effects believed to be part of the carcinogenic process influenced by the agent may assist in evaluating the relationship of exposure and response in the range of observation and at exposure levels below the range of observation.

The first step of dose-response assessment is evaluation within the range of observation. Approaches to analysis of the range of observation of human studies are determined by the type of study and how dose and response are measured in the study. In the absence of adequate human data for dose-response analysis, animal data are generally used. If there are sufficient quantitative data and adequate understanding of the carcinogenic process, a biologically based model may be developed to relate dose and response data on an agent-specific basis. Otherwise, as a default procedure, a standard model can be used to curve-fit the data.

The POD for extrapolating the relationship to environmental exposure levels of interest when the latter are outside the range of observed data is the lower 95% confidence limit on the lowest level that can be supported by the data. Other PODs may be more appropriate for certain data sets and, as described in the guidance, may be used instead. A lower limit rather than a central estimate is appropriate for several reasons. One considers the relative consequences of overestimating or underestimating risk and the Agency's choice to use methods that are not likely to underestimate risk. Another is that use of a bound, as opposed to a central estimate, accounts for the variability (i.e., the sampling error) in the experimental data. In addition, use of the lower bound is consistent with the goal of harmonization with the current practice for assessing effects other than cancer—also based on the lower limit on dose.

The second step of dose-response assessment is extrapolation to lower dose levels, if needed. This extrapolation is based on extension of a biologically based model if supported by substantial data (see Section 3.3.2). Otherwise, default approaches can be applied that are consistent with current understanding of mode(s) of action of the agent, including approaches that assume linearity or nonlinearity of the dose-response relationship, or both. A default

approach for linearity can be to extend a straight line to zero dose/zero response (see Section 3.3.4). The linear approach is used when there is an absence of sufficient information on modes of action or the mode of action information indicates that the dose-response curve at low dose is or is expected to be linear. A default approach for nonlinearity can be to use a reference dose or a reference concentration (see Section 3.3.3).

1.3.5. Susceptible Populations and Lifestages

An important use of mode of action information is to identify susceptible populations and lifestages. It is rare to have epidemiologic studies or animal bioassays conducted in susceptible individuals. This information need can be filled by identifying the key events of the mode of action and then identifying risk factors, such as differences due to genetic polymorphisms, disease, altered organ function, lifestyle, and lifestage, that can augment these key events. To do this, the information about the key precursor events is reviewed to identify particular populations or lifestages that can be particularly susceptible to their occurrence (see Section 2.5.3.4). Any information suggesting quantitative differences between populations or lifestages is flagged for consideration in the dose-response assessment (see Section 3.5).

1.3.6. Evaluating Risks from Childhood Exposures

NRC (1994) recommended that "EPA should assess risks to infants and children whenever it appears that their risks might be greater than those of adults." Executive Order 13045 (1997) requires that "each Federal Agency shall make it a high priority to identify and assess environmental health and safety risks that may disproportionately affect children, and shall ensure that their policies, programs, and standards address disproportionate risks that result from environmental health risks or safety risks." In assessing risks to children, EPA considers both effects manifest during childhood and early-life exposures that can contribute to effects at any time later in life.

These guidelines view childhood as a sequence of lifestages rather than viewing children as a subpopulation, the distinction being that a subpopulation refers to a portion of the population, whereas a lifestage is inclusive of the entire population. Exposures that are of concern extend from conception through adolescence and also include pre-conception exposures of both parents. These guidelines use the term "childhood" in this more inclusive sense.

There are usually no studies that directly evaluate risks following early-life exposure. Epidemiologic studies of early-life exposure to environmental agents are seldom available. Standard animal bioassays generally begin dosing after the animals are several weeks old, when

many systems are mature. This could lead to an understatement of risk, because an accepted concept in the science of carcinogenesis is that young animals are usually more susceptible to the carcinogenic activity of a chemical than are mature animals (McConnell, 1992).

At this time, there is some evidence of higher cancer risks following early-life exposure. For radiation carcinogenesis, it is clear that risks for several forms of cancer are highest following childhood exposure (NRC, 1990; Miller, 1995; U.S. EPA, 1999c). These human results are supported by the few animal bioassays that include perinatal (prenatal or early postnatal) exposure. Perinatal exposure to some agents can induce higher incidences of the tumors seen in standard bioassays; some examples include vinyl chloride (Maltoni et al., 1981), diethylnitrosamine (Peto et al., 1984), benzidine, DDT, dieldrin, and safrole (Vesselinovitch et al., 1979). Moreover, perinatal exposure to some agents, including vinyl chloride (Maltoni et al., 1981) and saccharin (Cohen, 1995; Whysner and Williams, 1996), can induce different tumors that are not seen in standard bioassays. Surveys comparing perinatal carcinogenesis bioassays with standard bioassays for a limited number of chemicals (McConnell, 1992; U.S. EPA, 1996b) have concluded that

- the same tumor sites are usually observed following either perinatal or adult exposure, and

- perinatal exposure in conjunction with adult exposure usually increases the incidence of tumors or reduces the latent period before tumors are observed.

The risk attributable to early-life exposure often appears modest compared with the risk from lifetime exposure, but it can be about 10-fold higher than the risk from an exposure of similar duration occurring later in life (Ginsberg, 2003). Further research is warranted to investigate the extent to which these findings apply to specific agents, chemical classes, and modes of action or in general.

These empirical results are consistent with current understanding of the biological processes involved in carcinogenesis, which leads to a reasonable expectation that children can be more susceptible to many carcinogenic agents. Some aspects potentially leading to childhood susceptibility are:

- Differences in the capacity to metabolize and clear chemicals can result in larger or smaller internal doses of the active agent(s).

- More frequent cell division during development can result in enhanced expression of mutations due to the reduced time available for repair of DNA lesions.

- Some embryonic cells, such as brain cells, lack key DNA repair enzymes.

- More frequent cell division during development can result in clonal expansion of cells with mutations from prior unrepaired DNA damage.

- Some components of the immune system are not fully functional during development.

- Hormonal systems operate at different levels during different lifestages.

- Induction of developmental abnormalities can result in a predisposition to carcinogenic effects later in life.

To evaluate risks from early-life exposure, these guidelines emphasize the role of toxicokinetic information to estimate levels of the active agent in children and toxicodynamic information to identify whether any key events of the mode of action are of increased concern early in life. Developmental toxicity studies can provide information on critical periods of exposure for particular targets of toxicity.

An approach to assessing risks from early-life exposure is presented in Figure 1-1. In the hazard assessment, when there are mode of action data, the assessment considers whether these data have special relevance during childhood, considering the various aspects of development listed above. Examples of such data include toxicokinetics that predict a sufficiently large internal dose in children or a mode of action where a key precursor event is more likely to occur during childhood. There is no recommended default to settle the question of whether tumors arising through the hypothesized mode of action are relevant during childhood; understanding the mode of action implies that there are sufficient data (on either the specific agent or the general mode of action) to form a confident conclusion about relevance during childhood (see Section 2.5.3.4).

In the dose-response assessment, the potential for susceptibility during childhood warrants explicit consideration in each assessment. These guidelines encourage developing

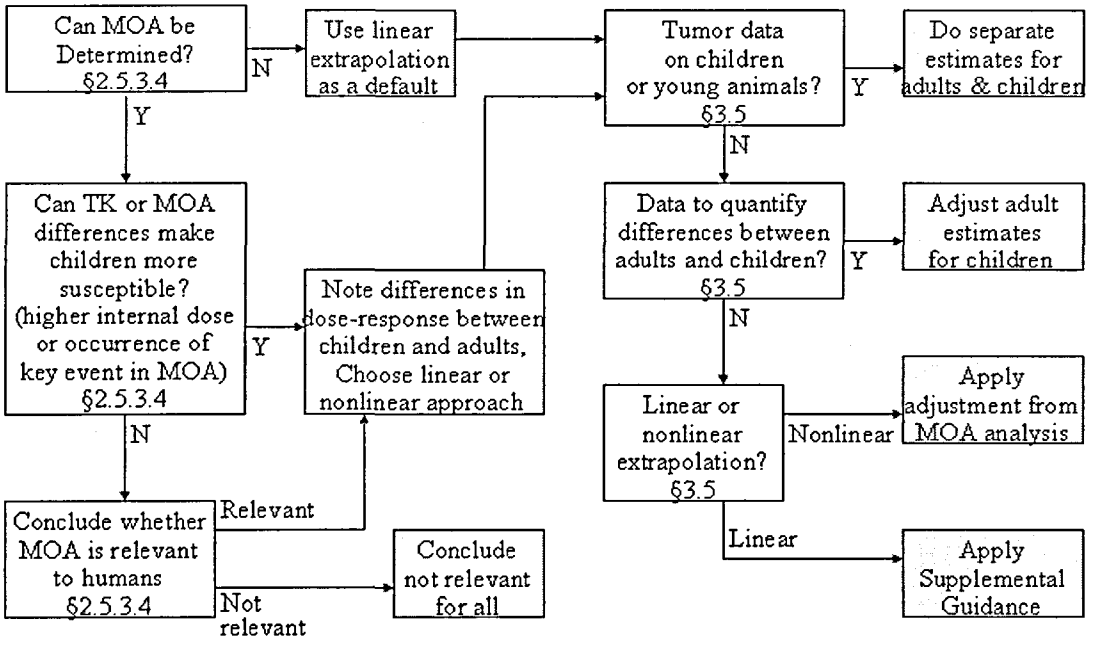
separate risk estimates for children according to a tiered approach that considers what pertinent data are available (see Section 3.5). Although childhood may be a susceptible period, exposures during childhood generally are not equivalent to exposures at other times and may be treated differently from exposures occurring later in life (see Section 3.5). In addition, adjustment of unit risk estimates can be warranted when used to estimate risks from childhood exposure (see Section 4.4).

At this time, several limitations preclude a full assessment of children's risk. There are no generally used testing protocols to identify potential environmental causes of cancers that are unique to children, including several forms of childhood cancer and cancers that develop from parental exposures, and cases where developmental exposure may alter susceptibility to carcinogen exposure in the adult (Birnbaum and Fenton, 2003). Dose-response assessment is limited by an inability to observe how developmental exposure can modify incidence and latency and an inability to estimate the ultimate tumor response resulting from induced susceptibility to later carcinogen exposures.

1.3.7. Emphasis on Characterization

The guidelines provide greater emphasis on characterization discussions for hazard, dose response, and exposure assessment. These characterizations summarize the assessments to explain the extent and weight of evidence, major points of interpretation and rationale for their selection, and strengths and weaknesses of the evidence and the analysis and discuss alternative conclusions and uncertainties that deserve serious consideration (U.S. EPA, 2000b). They serve as starting materials for the overall risk characterization process that completes the risk assessment.

Figure 1-1. Risk assessment of childhood exposures



2. HAZARD ASSESSMENT

2.1. OVERVIEW OF HAZARD ASSESSMENT AND CHARACTERIZATION

2.1.1. Analyses of Data

The purpose of hazard assessment is to review and evaluate data pertinent to two questions: (1) whether an agent may pose a carcinogenic hazard to human beings, and (2) under what circumstances an identified hazard may be expressed (NRC, 1994). Hazard assessment is composed of analyses of a variety of data that may range from observations of tumor responses to analysis of SARs. The purpose of the assessment is not simply to assemble these separate evaluations; its purpose is to construct a total case analysis examining the biological story the data reveal as a whole about carcinogenic effects and mode of action and their implications for human hazard and dose-response evaluation. Weight of evidence conclusions come from the combined strength and coherence of inferences appropriately drawn from all of the available evidence. To the extent that data permit, hazard assessment addresses the question of mode of action as both an initial step in identifying human hazard potential and a component in considering appropriate approaches to dose-response assessment.

The topics in this chapter include analysis of tumor data, both animal and human, and analysis of other key information about properties and effects that relate to carcinogenic potential. The chapter addresses how information can be used to evaluate potential modes of action. It also provides guidance on performing a weight of evidence evaluation.

2.1.2. Presentation of Results

Presentation of the results of hazard assessment should be informed by Agency guidance as discussed in Section 2.7. The results are presented in a technical hazard characterization that serves as a support to later risk characterization. It includes

- a summary of the evaluations of hazard data,
- the rationales for its conclusions, and
- an explanation of the significant strengths or limitations of the conclusions.

Another presentation feature is the use of a weight of evidence narrative that includes both a conclusion about the weight of evidence of carcinogenic potential and a summary of the data on which the conclusion rests. This narrative is a brief summary that replaces the alphanumerical classification system used in EPA's 1986 guidelines (U.S. EPA, 1986a).

2.2. ANALYSIS OF TUMOR DATA

Evidence of carcinogenicity comes from finding tumor increases in humans or laboratory animals exposed to a given agent or from finding tumors following exposure to structural analogues to the compound under review. The significance of observed or anticipated tumor effects is evaluated in reference to all the other key data on the agent. This section contains guidance for analyzing human and animal studies to decide whether there is an association between exposure to an agent or a structural analogue and occurrence of tumors. Note that the use of the term "tumor" here is generic, meaning malignant neoplasms or a combination of malignant and corresponding benign neoplasms.

Observation of only benign neoplasia may or may not have significance. Benign tumors that are not observed to progress to malignancy are assessed on a case-by-case basis. There is a range of possibilities for their overall significance. They may deserve attention because they are serious health problems even though they are not malignant; for instance, benign tumors may be a health risk because of their effect on the function of a target tissue such as the brain. They may be significant indicators of the need for further testing of an agent if they are observed in a short-term test protocol, or such an observation may add to the overall weight of evidence if the same agent causes malignancies in a long-term study. Knowledge of the mode of action associated with a benign tumor response may aid in the interpretation of other tumor responses associated with the same agent. In other cases, observation of a benign tumor response alone may have no significant health hazard implications when other sources of evidence show no suggestion of carcinogenicity.

2.2.1. Human Data

Human data may come from epidemiologic studies or case reports. Clinical human studies, which involve intentional exposures to substances, may provide data on acute health effects, but not on cancer. The most common sources of human data for cancer risk assessment are epidemiological investigations. Epidemiology is the study of the distribution of disease in human populations and the factors that may influence that distribution. The goals of cancer epidemiology are to identify distribution of cancer risk and determine the extent to which the risk can be attributed causally to specific exposures to exogenous or endogenous factors. Epidemiologic data are extremely valuable in risk assessment because they provide direct evidence on whether a substance is likely to produce cancer in humans, thereby avoiding the problem of species-to-species inference. Thus, when available human data of high quality and adequate statistical power are available, they are generally preferable over animal data and

should be given greater weight in hazard characterization and dose-response assessment, although both are utilized.

Null results from epidemiologic studies generally do not prove the absence of carcinogenic effects because such results can arise either from being truly negative or from inadequate statistical power, inadequate design, imprecise estimates, or confounding factors. However, null results from a well-designed and well-conducted epidemiologic study that contains usable exposure data can help to define upper limits for the estimated dose of concern for human exposure if the overall weight of the evidence indicates that the agent is potentially carcinogenic in humans.

Epidemiology can also complement experimental evidence in corroborating or clarifying the carcinogenic potential of the agent in question. For example, epidemiologic studies that show elevated cancer risk for tumor sites corresponding to those at which laboratory animals experience increased tumor incidence can strengthen the weight of evidence of human carcinogenicity. On the other hand, strong nonpositive epidemiologic data in conjunction with compelling mechanistic information can lend support to a conclusion that animal responses may not be predictive of a human cancer hazard. Furthermore, biochemical or molecular epidemiology may help improve understanding of the mechanisms of human carcinogenesis.

2.2.1.1. *Types of Studies*

The major types of cancer epidemiologic designs used for examining environmental causes of cancer are analytical studies and descriptive studies. Each study type has well-known strengths and weaknesses that affect interpretation of results, as summarized below (Kelsey et al., 1986; Lilienfeld and Lilienfeld, 1979; Mausner and Kramer, 1985; Rothman, 1986).

Analytical epidemiologic studies, which include case-control and cohort designs, are generally relied on for identifying a causal association between human exposure and adverse health effects. In case-control studies, groups of individuals with (cases) and without (controls) a particular disease are identified and compared to determine differences in exposure. In cohort studies, a group of "exposed" and "nonexposed" individuals are identified and studied over time to determine differences in disease occurrence. Cohort studies can be performed either prospectively or retrospectively from historical records.

Descriptive epidemiologic studies examine symptom or disease rates among populations in relation to personal characteristics such as age, gender, race, and temporal or environmental conditions. Descriptive studies are most frequently used to generate hypotheses about exposure factors, but subsequent analytical designs are necessary to infer causality. For example, cross-

sectional designs might be used to compare the prevalence of cancer between areas near and far from a Superfund site. However, in studies where exposure and disease information applies only to the current conditions, it is not possible to infer that the exposure actually *caused* the disease. Therefore, these studies are used to identify patterns or trends in disease occurrence over time or in different geographical locations, but typical limitations in the characterization of populations in these studies make it difficult to infer the causal agent or degree of exposure.

Biochemical or molecular epidemiologic studies use biological markers of effect as indicators of disease or its precursors. The application of techniques for measuring cellular and molecular alterations due to exposure to specific environmental agents may allow conclusions to be drawn about the mechanisms of carcinogenesis. Refer to the sections on biomarkers (Section 2.4) and mode of action (Section 2.5) for more information on this topic.

Case reports describe a particular effect in an individual or group of individuals who were exposed to a substance. These reports are often anecdotal or highly selective in nature and generally are of limited use for hazard assessment. Investigative follow-up may or may not accompany such reports. For cancer, the most common types of case series are associated with occupational and childhood exposures. Case reports can be particularly valuable for identifying unique features such as an association with an uncommon tumor (e.g., vinyl chloride and angiosarcoma or diethylstilbestrol and clear-cell carcinoma of the vagina).

2.2.1.2. *Assessing the Quality of Epidemiologic Studies*

Characteristics that are generally desirable in epidemiologic studies include (1) clear articulation of study objectives or hypothesis; (2) proper selection and characterization of comparison groups (exposed and unexposed groups or case and control groups); (3) adequate characterization of exposure; (4) sufficient length of follow-up for disease occurrence; (5) valid ascertainment of the causes of cancer morbidity and mortality; (6) proper consideration of bias and confounding factors; (7) adequate sample size to detect an effect; (8) clear, well-documented, and appropriate methodology for data collection and analysis; (9) adequate response rate and methodology for handling missing data; and (10) complete and clear documentation of results. No single criterion determines the overall adequacy of a study. Practical and financial constraints may limit the ability to address all of these characteristics in a study. The risk assessor is encouraged to consider how the limitations of the available studies might influence the conclusions. The following discussions highlight the major factors included in an analysis of epidemiologic studies.

2.2.1.2.1. ***Population issues.*** When comparing cases and controls or exposed and non-exposed populations, it would be preferable for the two populations to differ only in exposure to the agent in question. Because this is seldom the case, it is important to identify sources of bias inherent in a study's design or data collection methods. Bias is a systematic error. In epidemiological studies, bias can occur in the selection of cases and controls or exposed and non-exposed populations, as well as the follow up of the groups, or the classification of disease or exposure. The size of the risks observed can be affected by noncomparability between populations of factors such as general health (McMichael, 1976), diet, lifestyle, or geographic location; differences in the way case and control individuals recall past events; differences in data collection that result in unequal ascertainment of health effects in the populations; and unequal follow-up of individuals. Both acceptance of studies for assessment and judgment of their strengths or weaknesses depend on identifying their sources of bias and the effects on study results.

2.2.1.2.2. ***Exposure issues.*** For epidemiologic data to be useful in determining whether there is an association between health effects and exposure to an agent, there should be adequate characterization of exposure information. In general, greater weight should be given to studies with more precise and specific exposure estimates.

Questions to address about exposure are: What can one reliably conclude about the level, duration, route, and frequency of exposure of individuals in one population as compared with another? How sensitive are study results to uncertainties in these parameters?

Actual exposure measurements are not available for many retrospective studies. Therefore, surrogates are often used to reconstruct exposure parameters. These may involve attributing exposures to job classifications in a workplace or to broader occupational or geographic groupings. Use of surrogates carries a potential for misclassification in that individuals may be placed in an incorrect exposure group. Misclassification generally leads to reduced ability of a study to detect differences between study and referent populations.

When either current or historical monitoring data are available, the exposure evaluation includes consideration of the error bounds of the monitoring and analytic methods and whether the data are from routine or accidental exposures. The potentials for misclassification and measurement errors are amenable to both qualitative and quantitative analysis. These are essential analyses for judging a study's results, because exposure estimation is the most critical part of a retrospective study.

Biological markers potentially offer excellent measures of exposure (Hulka and Margolin, 1992; Peto and Darby, 1994). Validated markers of exposure such as alkylated hemoglobin from exposure to ethylene oxide (Van Sittert et al., 1985) or urinary arsenic (Enterline et al., 1987) can greatly improve estimates of dose. Markers closely identified with effects promise to greatly increase the ability of studies to distinguish real effects from bias at low levels of relative risk between populations (Taylor et al., 1994; Biggs et al., 1993) and to resolve problems of confounding risk factors.

2.2.1.2.3. ***Confounding factors.*** In observational epidemiologic studies, it is very difficult to guarantee the control of confounding variables. A confounder is a variable that is related to both the health outcome of concern (cancer) and exposure. Common examples include age, socioeconomic status, smoking habits, and diet. For instance, if older people are more likely to be exposed to a given contaminant as well as more likely to have cancer because of their age, age is considered a confounder. Adjustment for potentially confounding factors can occur either in the design of the study (e.g., individual or group matching on critical factors) or in the statistical analysis of the results (stratification or direct or indirect adjustment). Direct adjustment in the statistical analysis may not be possible owing to the presentation of the data or because needed information was not collected during the study. In this case, indirect comparisons may be possible. For example, in the absence of data on smoking status among individuals in the study population, an examination of the possible contribution of cigarette smoking to increased lung cancer risk may be based on information from other sources, such as the American Cancer Society's longitudinal studies (Hammand, 1966; Garfinkel and Silverberg, 1991). The effectiveness of adjustments contributes to the ability to draw inferences from a study.

Different studies involving exposure to an agent may have different confounding factors. If consistent increases in cancer risk are observed across a collection of studies with different confounding factors, the inference that the agent under investigation was the etiologic factor is strengthened. It also may be the case that the agent of interest is a risk factor in conjunction with another agent. This relationship may be revealed in a collection of studies, such as in the case of asbestos exposure and smoking.

2.2.1.2.4. ***Likelihood of observing an effect.*** The power of a study – the likelihood of observing an effect if one exists – increases with sample size. If the size of the effect is expected to be very small at low doses, higher doses or longer durations of exposure may be needed to have an

appreciable likelihood of observing an effect with a given sample size. Because of the often long latency period in cancer development, the likelihood of observing an effect also depends on whether adequate time has elapsed since exposure began for effects to occur. A unique feature that can be ascribed to the effects of a particular agent (such as a tumor type that is seen only rarely in the absence of the agent) can increase sensitivity by permitting separation of bias and confounding factors from real effects. Similarly, a biomarker particular to the agent can permit these distinctions. Statistical re-analyses of data, particularly an examination of different exposure indices, can give insight into potential exposure-response relationships. These are all factors to explore in statistical analysis of the data.

2.2.1.2.5. ***Statistical considerations.*** The analysis should apply appropriate statistical methods to ascertain whether the observed association between exposure and effects would be expected by chance. A description of the method or methods used should include the reasons for their selection. Statistical analyses of the bias, confounding, and interaction are part of addressing the significance of an association and the power of a study to detect an effect.

The analysis augments examination of the results for the whole population with exploration of the results for groups with comparatively greater exposure or time since first exposure. This may support identifying an association or establishing a dose-response trend. When studies show no association, such exploration may apply to determining an upper limit on potential human risk for consideration alongside results of animal tumor effects studies.

2.2.1.2.6. ***Combining statistical evidence across studies.*** Meta-analysis is a means of integrating the results of multiple studies of similar health effects and risk factors. This technique is particularly useful when various studies yield varying degrees of risk or even conflicting associations (negative and positive). It is intended to introduce consistency and comprehensiveness into what otherwise might be a more subjective review of the literature. The value of such an analysis is dependent upon a systematic review of the literature that uses transparent criteria of inclusion and exclusion. In interpreting such analyses, it is important to consider the effects of differences in study quality, as well as the effect of publication bias. Meta-analysis may not be useful in some circumstances. These include when the relationship between exposure and disease is obvious from the individual studies; when there are only a few studies of the key health outcomes; when there is insufficient information from available studies related to disease, risk estimate, or exposure classification to insure comparability; or when there

are substantial confounding or other biases that cannot be adjusted for in the analysis (Blair et al., 1995; Greenland, 1987; Peto, 1992).

2.2.1.3. *Evidence for Causality*

Determining whether an observed association (risk) is causal rather than spurious involves consideration of a number of factors. Sir Bradford Hill developed a set of guidelines for evaluating epidemiologic associations in conjunction with the 1964 Surgeon General's Report on Smoking (Hill, 1965; Rothman, 1986; IPCS, 1999). Although these guidelines have become known as "causal criteria," it is important to note that they cannot be used as a strictly quantitative checklist. Rather, these "criteria" should be used to determine the strength of the evidence for concluding causality. The list below has been adapted from Hill's guidelines as an aid in judging causality.

(a) Consistency of the observed association. An inference of causality is strengthened when a pattern of elevated risks is observed across several independent studies. The reproducibility of findings constitutes one of the strongest arguments for causality. If there are discordant results among investigations, possible reasons such as differences in exposure, confounding factors, and the power of the study are considered.

(b) Strength of the observed association. The finding of large, precise risks increases confidence that the association is not likely due to chance, bias, or other factors. A modest risk, however, does not preclude a causal association and may reflect a lower level of exposure, an agent of lower potency, or a common disease with a high background level.

(c) Specificity of the observed association. As originally intended, this refers to increased inference of causality if one cause is associated with a single effect or disease (Hill, 1965). Based on our current understanding that many agents cause cancer at multiple sites, and many cancers have multiple causes, this is now considered one of the weaker guidelines for causality. Thus, although the presence of specificity may support causality, its absence does not exclude it.

(d) Temporal relationship of the observed association. A causal interpretation is strengthened when exposure is known to precede development of the disease. Because a latent period of up to 20 years or longer is associated with cancer development, the study should consider whether exposures occurred sufficiently long ago to produce an effect at the time the cancer is assessed. This is among the strongest criteria for an inference of causality.

(e) Biological gradient (exposure-response relationship). A clear exposure-response relationship (e.g., increasing effects associated with greater exposure) strongly suggests cause

and effect, especially when such relationships are also observed for duration of exposure (e.g., increasing effects observed following longer exposure times). Because there are many possible reasons that an epidemiologic study may fail to detect an exposure-response relationship (for example, a small range of observed exposure levels or exposure misclassification), the absence of an exposure-response relationship does not exclude a causal relationship.

(f) Biological plausibility. An inference of causality tends to be strengthened by consistency with data from experimental studies or other sources demonstrating plausible biological mechanisms. A lack of mechanistic data, however, is not a reason to reject causality.

(g) Coherence. An inference of causality may be strengthened by other lines of evidence that support a cause-and-effect interpretation of the association. Information is considered from animal bioassays, toxicokinetic studies, and short-term studies. The absence of other lines of evidence, however, is not a reason to reject causality.

(h) Experimental evidence (from human populations). Experimental evidence is seldom available from human populations and exists only when conditions of human exposure are altered to create a "natural experiment" at different levels of exposure. Strong evidence for causality can be provided when a change in exposure brings about a change in disease frequency, for example, the decrease in the risk of lung cancer that follows cessation of smoking.

(i) Analogy. SARs and information on the agent's structural analogues can provide insight into whether an association is causal.

2.2.1.4. *Assessment of Evidence of Carcinogenicity from Human Data*

All studies that are considered to be of acceptable quality, whether yielding positive or null results, or even suggesting protective carcinogenic effects, should be considered in assessing the totality of the human evidence. Conclusions about the overall evidence for carcinogenicity from available studies in humans should be summarized along with a discussion of uncertainties and gaps in knowledge. Conclusions regarding the strength of the evidence for positive or negative associations observed, as well as evidence supporting judgments of causality, should be clearly described. In assessing the human data within the overall weight of evidence, determination about the strength of the epidemiologic evidence should clearly identify the degree to which the observed associations may be explained by other factors, including bias or confounding.

2.2.2. Animal Data

Various whole-animal test systems are currently used or are under development for evaluating potential carcinogenicity. Cancer studies involving chronic exposure for most of the lifespan of an animal are generally accepted for evaluation of tumor effects (Tomatis et al., 1989; Rall, 1991; Allen et al., 1988; but see Ames and Gold, 1990). Other studies of special design are useful for observing formation of preneoplastic lesions or tumors or investigating specific modes of action. Their applicability is determined on a case-by-case basis.

2.2.2.1. *Long-term Carcinogenicity Studies*

The objective of long-term carcinogenesis bioassays is to determine the potential carcinogenic hazard and dose-response relationships of the test agent. Carcinogenicity rodent studies are designed to examine the production of tumors as well as preneoplastic lesions and other indications of chronic toxicity that may provide evidence of treatment-related effects and insights into the way the test agent produces tumors. Current standardized carcinogenicity studies in rodents test at least 50 animals per sex per dose group in each of three treatment groups and in a concurrent control group, usually for 18 to 24 months, depending on the rodent species tested (OECD, 1981; U.S. EPA, 1998c). The high dose in long-term studies is generally selected to provide the maximum ability to detect treatment-related carcinogenic effects while not compromising the outcome of the study through excessive toxicity or inducing inappropriate toxicokinetics (e.g., overwhelming absorption or detoxification mechanisms). The purpose of two or more lower doses is to provide some information on the shape of the dose-response curve. Similar protocols have been and continue to be used by many laboratories worldwide.

All available studies of tumor effects in whole animals should be considered, at least preliminarily. The analysis should discard studies judged to be wholly inadequate in protocol, conduct, or results. Criteria for the technical adequacy of animal carcinogenicity studies have been published and should be used as guidance to judge the acceptability of individual studies (NTP, 1984; OSTP, 1985). Care should be taken to include studies that provide some evidence bearing on carcinogenicity or that help interpret effects noted in other studies, even if these studies have some limitations of protocol or conduct. Such limited, but not wholly inadequate, studies can contribute as their deficiencies permit. The findings of long-term rodent bioassays should be interpreted in conjunction with results of prechronic studies along with toxicokinetic studies and other pertinent information, if available. Evaluation of tumor effects takes into consideration both biological and statistical significance of the findings (Haseman, 1984, 1985, 1990, 1995). The following sections highlight the major issues in the evaluation of long-term carcinogenicity studies.

2.2.2.1.1. ***Dosing issues.*** Among the many criteria for technical adequacy of animal carcinogenicity studies is the appropriateness of dose selection. The selection of doses for chronic bioassays is based on scientific judgments and sound toxicologic principles. Dose selection should be made on the basis of relevant toxicologic information from prechronic, mechanistic, and toxicokinetic and mechanistic studies. How well the dose selection is made is evaluated after the completion of the bioassay. A scientific rationale for dose selection should be clearly articulated (e.g., ILSI, 1997).

Interpretation of carcinogenicity study results is profoundly affected by study exposure conditions, especially by inappropriate dose selection. This is particularly important in studies that are nonpositive for carcinogenicity, because failure to reach a sufficient dose reduces the sensitivity of the studies. A lack of tumorigenic responses at exposure levels that cause significant impairment of animal survival may also not be acceptable. In addition, overt toxicity or inappropriate toxicokinetics due to excessively high doses may result in tumor effects that are secondary to the toxicity rather than directly attributable to the agent.

With regard to the appropriateness of the high dose, an adequate high dose would generally be one that produces some toxic effects without unduly affecting mortality from effects other than cancer or producing significant adverse effects on the nutrition and health of the test animals (OECD, 1981; NRC, 1993a). If the test agent does not appear to cause any specific target organ toxicity or perturbation of physiological function, an adequate high dose can be specified in terms of a percentage reduction of body weight gain over the lifespan of the animals. The high dose would generally be considered inadequate if neither toxicity nor change in weight gain is observed. On the other hand, significant increases in mortality from effects other than cancer generally indicate that an adequate high dose has been exceeded.

Other signs of treatment-related toxicity associated with an excessive high dose may include (a) significant reduction of body weight gain (e.g., greater than 10%), (b) significant increases in abnormal behavioral and clinical signs, (c) significant changes in hematology or clinical chemistry, (d) saturation of absorption and detoxification mechanisms, or (e) marked changes in organ weight, morphology, and histopathology. It should be noted that practical upper limits have been established to avoid the use of excessively high doses in long-term carcinogenicity studies of environmental chemicals (e.g., 5% of the test substance in the feed for dietary studies or 1 g/kg body weight for oral gavage studies [OECD, 1981]).

For dietary studies, weight gain reductions should be evaluated as to whether there is a palatability problem or an issue with food efficiency; certainly, the latter is a toxic manifestation. In the case of inhalation studies with respirable particles, evidence of impairment of normal

clearance of particles from the lung should be considered along with other signs of toxicity to the respiratory airways to determine whether the high exposure concentration has been appropriately selected. For dermal studies, evidence of skin irritation may indicate that an adequate high dose has been reached (U.S. EPA, 1989).

In order to obtain the most relevant information from a long-term carcinogenicity study, it is important to maximize exposure conditions to the test material. At the same time, caution is appropriate in using excessive high-dose levels that would confound the interpretation of study results to humans. The middle and lowest doses should be selected to characterize the shape of the dose-response curve as much as possible. It is important that the doses be adequately spaced so that the study can provide relevant dose-response data for assessing human hazard and risk. If the testing of potential carcinogenicity is being combined with an evaluation of noncancer chronic toxicity, the study should be designed to include one dose that does not elicit adverse effects.

There are several possible outcomes regarding the study interpretation of the significance and relevance of tumorigenic effects associated with exposure or dose levels below, at, or above an adequate high dose. The general guidance is given here; for each case, the information at hand should be evaluated and a rationale should be given for the position taken.

- *Adequate high dose.* If an adequate high dose has been used, tumor effects are judged positive or negative depending on the presence or absence of significant tumor incidence increases, respectively.

- *Excessive high dose.* If toxicity or mortality is excessive at the high dose, interpretation depends on the finding of tumors or not.

  - Studies that show tumor effects only at excessive doses may be compromised and may or may not carry weight, depending on the interpretation in the context of other study results and other lines of evidence. Results of such studies, however, are generally not considered suitable for dose-response extrapolation if it is determined that the mode(s) of action underlying the tumorigenic responses at high doses is not operative at lower doses.

  - Studies that show tumors at lower doses, even though the high dose is excessive and may be discounted, should be evaluated on their own merits.

- If a study does not show an increase in tumor incidence at a toxic high dose and appropriately spaced lower doses are used without such toxicity or tumors, the study is generally judged as negative for carcinogenicity.

- *Inadequate high dose.* Studies of inadequate sensitivity where an adequate high dose has not been reached may be used to bound the dose range where carcinogenic effects might be expected.

2.2.2.1.2. ***Statistical considerations.*** The main aim of statistical evaluation is to determine whether exposure to the test agent is associated with an increase of tumor development. Statistical analysis of a long-term study should be performed for each tumor type separately. The incidence of benign and malignant lesions of the same cell type, usually within a single tissue or organ, are considered separately and are combined when scientifically defensible (McConnell et al., 1986).

Trend tests and pairwise comparison tests are the recommended tests for determining whether chance, rather than a treatment-related effect, is a plausible explanation for an apparent increase in tumor incidence. A trend test such as the Cochran-Armitage test (Snedecor and Cochran, 1967) asks whether the results in all dose groups together increase as dose increases. A pairwise comparison test such as the Fisher exact test (Fisher, 1950) asks whether an incidence in one dose group is increased over that of the control group. By convention, for both tests a statistically significant comparison is one for which p is less than 0.05 that the increased incidence is due to chance. Significance in either kind of test is sufficient to reject the hypothesis that chance accounts for the result.

A statistically significant response may or may not be biologically significant and vice versa. The selection of a significance level is a policy choice based on a trade-off between the risks of false positives and false negatives. A result with a significance level of greater or less than 5% (the most common significance level) is examined to see if the result confirms other scientific information. When the assessment departs from a simple 5% level, this should be highlighted in the risk characterization. A two-tailed test or a one-tailed test can be used. In either case a rationale is provided.

Statistical power can affect the likelihood that a statistically significant result could reasonably be expected. This is especially important in studies or dose groups with small sample sizes or low dose rates. Reporting the statistical power can be useful for comparing and reconciling positive and negative results from different studies.

Considerations of multiple comparisons should also be taken into account. Haseman (1983) analyzes typical animal bioassays that test both sexes of two species and concludes that, because of multiple comparisons, a single tumor increase for a species-sex-site combination that is statistically significant at the 1% level for common tumors or 5% for rare tumors corresponds to a 7–8% significance level for the study as a whole. Therefore, animal bioassays presenting only one significant result that falls short of the 1% level for a common tumor should be treated with caution.

2.2.2.1.3. ***Concurrent and historical controls.*** The standard for determining statistical significance of tumor incidence comes from a comparison of tumors in dosed animals with those in concurrent control animals. Additional insights about both statistical and biological significance can come from an examination of historical control data (Tarone, 1982; Haseman, 1995). Historical control data can add to the analysis, particularly by enabling identification of uncommon tumor types or high spontaneous incidence of a tumor in a given animal strain. Identification of common or uncommon situations prompts further thought about the meaning of the response in the current study in context with other observations in animal studies and with other evidence about the carcinogenic potential of the agent. These other sources of information may reinforce or weaken the significance given to the response in the hazard assessment. Caution should be exercised in simply looking at the ranges of historical responses, because the range ignores differences in survival of animals among studies and is related to the number of studies in the database.

In analyzing results for uncommon tumors in a treated group that are not statistically significant in comparison with concurrent controls, the analyst can use the experience of historical controls to conclude that the result is in fact unlikely to be due to chance. In analyzing results for common tumors, a different set of considerations comes into play. Generally speaking, statistically significant increases in tumors should not be discounted simply because incidence rates in the treated groups are within the range of historical controls or because incidence rates in the concurrent controls are somewhat lower than average. Random assignment of animals to groups and proper statistical procedures provide assurance that statistically significant results are unlikely to be due to chance alone. However, caution should be used in interpreting results that are barely statistically significant or in which incidence rates in concurrent controls are unusually low in comparison with historical controls.

In cases where there may be reason to discount the biological relevance to humans of increases in common animal tumors, such considerations should be weighed on their own merits and clearly distinguished from statistical concerns.

When historical control data are used, the discussion should address several issues that affect comparability of historical and concurrent control data, such as genetic drift in the laboratory strains, differences in pathology examination at different times and in different laboratories (e.g., in criteria for evaluating lesions; variations in the techniques for the preparation or reading of tissue samples among laboratories), and comparability of animals from different suppliers. The most relevant historical data come from the same laboratory and the same supplier and are gathered within 2 or 3 years one way or the other of the study under review; other data should be used only with extreme caution.

2.2.2.1.4. ***Assessment of evidence of carcinogenicity from long-term animal studies.*** In general, observation of tumor effects under different circumstances lends support to the significance of the findings for animal carcinogenicity. Significance is a function of the number of factors present and, for a factor such as malignancy, the severity of the observed pathology. The following observations add significance to the tumor findings:

- uncommon tumor types;
- tumors at multiple sites;
- tumors by more than one route of administration;
- tumors in multiple species, strains, or both sexes;
- progression of lesions from preneoplastic to benign to malignant;
- reduced latency of neoplastic lesions;
- metastases;
- unusual magnitude of tumor response;
- proportion of malignant tumors; and
- dose-related increases.

Tumor findings in animals generally indicate that an agent may produce such effects in humans. Moreover, the absence of tumor findings in well-conducted, long-term animal studies in at least two species provides reasonable assurance that an agent may not be a carcinogenic concern for humans. Each of these assumptions may be adopted, when appropriate, after evaluation of tumor data and other key evidence.

2.2.2.1.5. ***Site concordance.*** Site concordance of tumor effects between animals and humans should be considered in each case. Thus far, there is evidence that growth control mechanisms at the level of the cell are homologous among mammals, but there is no evidence that these mechanisms are site concordant. Moreover, agents observed to produce tumors in both humans and animals have produced tumors either at the same site (e.g., vinyl chloride) or different sites (e.g., benzene) (NRC, 1994). Hence, site concordance is not assumed a priori. On the other hand, certain processes with consequences for particular tissue sites (e.g., disruption of thyroid function) may lead to an anticipation of site concordance.

2.2.2.2. ***Perinatal Carcinogenicity Studies***

The objective of perinatal carcinogenesis studies is to determine the carcinogenic potential and dose-response relationships of the test agent in the developing organism. Some investigators have hypothesized that the age of initial exposure to a chemical carcinogen may influence the carcinogenic response (Vesselinovitch et al., 1979; Rice, 1979; McConnell, 1992). Current standardized long-term carcinogenesis bioassays generally begin dosing animals at 6–8 weeks of age and continue dosing for the lifespan of the animal (18–24 months). This protocol has been modified in some cases to investigate the potential of the test agent to induce transplacental carcinogenesis or to investigate the potential differences following perinatal and adult exposures, but currently there is not a standardized protocol for testing agents for carcinogenic effects following prenatal or early postnatal exposure.

Several cancer bioassay studies have compared adult and perinatal exposures (see McConnell, 1992; U.S. EPA, 1996b). A review of these studies reveals that perinatal exposure rarely identifies carcinogens that are not found in standard animal bioassays. Exposure that is perinatal can increase the incidence of a given type of tumor. The increase may reflect an increased length of exposure and a higher dose for the developing organism relative to the adult or an increase in susceptibility in some cases. Additionally, exposure that is perinatal through adulthood sometimes reduces the latency period for tumors to develop in the growing organism (U.S. EPA, 1996b). EPA evaluates the usefulness of perinatal studies on an agent-by-agent basis (for example, U.S. EPA, 1997a,b).

Perinatal study data analysis generally follows the principles discussed above for evaluating other long-term carcinogenicity studies. When differences in responses between perinatal animals and adult animals suggest an increased susceptibility of perinatal or postnatal animals, such as the ones below, a separate evaluation of the response should be prepared:

- a difference in dose-response relationship,
- the presence of different tumor types,
- an earlier onset of tumors, or
- an increase in the incidence of tumors.

2.2.2.3. *Other Studies*

Intermediate-term studies often use protocols that screen for carcinogenic or preneoplastic effects, sometimes in a single tissue. Some protocols involve the development of various proliferative lesions, such as foci of alteration in the liver (Goldsworthy et al., 1986). Others use tumor endpoints, such as the induction of lung adenomas in the sensitive strain A mouse (Maronpot et al., 1986) or tumor induction in initiation-promotion studies using various organs such as the bladder, intestine, liver, lung, mammary gland, and thyroid (Ito et al., 1992). In these tests, the selected tissue rather than the whole animal is, in a sense, the test system. Important information concerning the steps in the carcinogenic process and mode of action can be obtained from "start/stop" experiments. In these protocols, an agent is given for a period of time to induce particular lesions or effects and then stopped in order to evaluate the progression or reversibility of processes (Todd, 1986; Marsman and Popp, 1994).

Assays in genetically engineered rodents may provide insight into the chemical and gene interactions involved in carcinogenesis (Tennant et al., 1995). These mechanistically based approaches involve activated oncogenes that are introduced (transgenic) or tumor suppressor genes that are deleted (knocked out). If appropriate genes are selected, not only may these systems provide information on mechanisms, but the rodents typically show tumor development earlier than in the standard bioassay. Transgenic mutagenesis assays also represent a mechanistic approach for assessing the mutagenic properties of agents as well as developing quantitative linkages between exposure, internal dose, and mutation related to tumor induction (Morrison and Ashby, 1994; Sisk et al., 1994; Hayward et al., 1995).

The support that these studies give to a determination of carcinogenicity rests on their contribution to the consistency of other evidence about an agent. For instance, benzoyl peroxide has promoter activity on the skin, but the overall evidence may be less supportive (Kraus et al., 1995). These studies also may contribute information about mode of action. It is important to recognize the limitations of these experimental protocols, such as short duration, limited histology, lack of complete development of tumors, or experimental manipulation of the carcinogenic process, that may limit their contribution to the overall assessment. Generally, their results are appropriate as aids in the interpretation of other toxicological evidence (e.g., rodent

chronic bioassays), especially regarding potential modes of action. On the basis of currently available information, it is unlikely that any of these assays, which are conducted for 6 months with 15 animals per group, will replace all chronic bioassays for hazard identification (Spalding et al., 2000; Guzelian et al., 2000; ILSI, 2001).

2.2.3. Structural Analogue Data

For some chemical classes, there is significant available information, largely from rodent bioassays, on the carcinogenicity of analogues. Analogue effects are instructive in investigating carcinogenic potential of an agent as well as in identifying potential target organs, exposures associated with effects, and potential functional class effects or modes of action. All appropriate studies should be included and analyzed, whether indicative of a positive effect or not. Evaluation includes tests in various animal species, strains, and sexes; with different routes of administration; and at various doses, as data are available. Confidence in conclusions is a function of how similar the analogues are to the agent under review in structure, metabolism, and biological activity. It is important to consider this confidence to ensure a balanced position.

2.3. ANALYSIS OF OTHER KEY DATA

The physical, chemical, and structural properties of an agent, as well as data on endpoints that are thought to be critical elements of the carcinogenic process, provide valuable insights into the likelihood of human cancer risk. The following sections provide guidance for analyses of these data.

2.3.1. Physicochemical Properties

Physicochemical properties affect an agent's absorption, tissue distribution (bioavailability), biotransformation, and degradation in the body and are important determinants of hazard potential (and dose-response analysis). Properties that should be analyzed include, but are not limited to, molecular weight, size, and shape; valence state; physical state (gas, liquid, solid); water or lipid solubility, which can influence retention and tissue distribution; and potential for chemical degradation or stabilization in the body.

An agent's potential for chemical reaction with cellular components, particularly with DNA and proteins, is also important. The agent's molecular size and shape, electrophilicity, and charge distribution are considered in order to decide whether they would facilitate such reactions.

2.3.2. Structure-Activity Relationships

SAR analyses and models can be used to predict molecular properties, surrogate biological endpoints, and carcinogenicity. Overall, these analyses provide valuable initial information on agents, they may strengthen or weaken concern, and they are part of the weight of evidence.

Currently, SAR analysis is most useful for chemicals and metabolites that are believed to initiate carcinogenesis through covalent interaction with DNA (i.e., DNA-reactive, mutagenic, electrophilic, or proelectrophilic chemicals) (Ashby and Tennant, 1991). For organic chemicals, the predictive capability of SAR analysis combined with other toxicity information has been demonstrated (Ashby and Tennant, 1994). The following parameters are useful in comparing an agent to its structural analogues and congeners that produce tumors and affect related biological processes such as receptor binding and activation, mutagenicity, and general toxicity (Woo and Arcos, 1989):

- nature and reactivity of the electrophilic moiety or moieties present;
- potential to form electrophilic reactive intermediate(s) through chemical, photochemical, or metabolic activation;
- contribution of the carrier molecule to which the electrophilic moiety(ies) is attached;
- physicochemical properties (e.g., physical state, solubility, octanol/water partition coefficient, half-life in aqueous solution);
- structural and substructural features (e.g., electronic, stearic, molecular geometric);
- metabolic pattern (e.g., metabolic pathways and activation and detoxification ratio); and
- possible exposure route(s) of the agent.

Suitable SAR analysis of non-DNA-reactive chemicals and of DNA-reactive chemicals that do not appear to bind covalently to DNA should be based on knowledge or postulation of

the probable mode(s) of action of closely related carcinogenic structural analogues (e.g., receptor mediated, cytotoxicity related). Examination of the physicochemical and biochemical properties of the agent may then provide the rest of the information needed in order to make an assessment of the likelihood of the agent's activity by that mode of action.

2.3.3. Comparative Metabolism and Toxicokinetics

Studies of the absorption, distribution, biotransformation, and excretion of agents permit comparisons among species to assist in determining the implications of animal responses for human hazard assessment, supporting identification of active metabolites, identifying changes in distribution and metabolic pathway or pathways over a dose range, and making comparisons among different routes of exposure.

If extensive data are available (e.g., blood/tissue partition coefficients and pertinent physiological parameters of the species of interest), physiologically based toxicokinetic models can be constructed to assist in a determination of tissue dosimetry, species-to-species extrapolation of dose, and route-to-route extrapolation (Connolly and Andersen, 1991; see Section 3.1.2). If it is not contrary to available data, it may be assumed as a default that toxicokinetic and metabolic processes are qualitatively comparable among species. Discussion of appropriate defaults regarding quantitative comparison and their modifications appears in Chapter 3.

The *qualitative* question of whether an agent is absorbed by a particular route of exposure is important for weight of evidence classification, discussed in Section 2.6. Decisions about whether route of exposure is a limiting factor on expression of any hazard, in that absorption does not occur by a route, are generally based on studies in which effects of the agent or its structural analogues have been observed by different routes, on physical-chemical properties, or on toxicokinetics studies.

Adequate metabolism and toxicokinetic data can be applied toward the following as data permit. Confidence in conclusions is enhanced when in vivo data are available.

- *Identifying metabolites and reactive intermediates of metabolism and determining whether one or more of these intermediates is likely to be responsible for the observed effects.* This information on the reactive intermediates appropriately focuses SAR analysis, analysis of potential modes of action, and estimation of internal dose in dose-response assessment (D'Souza et al., 1987; Krewski et al., 1987).

- *Identifying and comparing the relative activities of metabolic pathways in animals and humans and at different ages.* This analysis can provide insights for extrapolating results of animal studies to humans.

- *Describing anticipated distribution within the body and possibly identifying target organs.* Use of water solubility, molecular weight, and structure analysis can support qualitative inferences about anticipated distribution and excretion. In addition, describing whether the agent or metabolite of concern will be excreted rapidly or slowly or whether it will be stored in a particular tissue or tissues to be mobilized later can identify issues in comparing species and formulating dose-response assessment approaches.

- *Identifying changes in toxicokinetics and metabolic pathways with increases in dose.* These changes may result in important differences between high and low dose levels in disposition of the agent or its generation of active forms. These studies play an important role in providing a rationale for dose selection in carcinogenicity studies.

- *Identifying and comparing metabolic process differences by age, sex, or other characteristic so that susceptible subpopulations can be recognized.* For example, metabolic capacity with respect to P450 enzymes in newborn children is extremely limited compared to that in adults, so that a carcinogenic metabolite formed through P450 activity will have limited effect in the young, whereas a carcinogenic agent deactivated through P450 activity will result in increased susceptibility of this lifestage (Cresteil, 1998). A variety of changes in toxicokinetics and physiology occur from the fetal stage to post-weaning to young child. Any of these changes may make a difference for risk (Renwick, 1998).

- *Determining bioavailability via different routes of exposure by analyzing uptake processes under various exposure conditions.* This analysis supports identification of hazards for untested routes. In addition, use of physicochemical data (e.g., octanol-water partition coefficient information) can support an inference about the likelihood of dermal absorption (Flynn, 1990).

Attempts should be made in all of these areas to clarify and describe as much as possible the variability to be expected because of differences in species, sex, age, and route of exposure. The analysis takes into account the presence of subpopulations of individuals who are particularly vulnerable to the effects of an agent because of toxicokinetic or metabolic differences (genetically or environmentally determined) (Bois et al., 1995) and is a special emphasis for assessment of risks to children.

2.3.4. Toxicological and Clinical Findings

Toxicological findings in experimental animals and clinical observations in humans are important resources for the cancer hazard assessment. Such findings provide information on physiological effects and effects on enzymes, hormones, and other important macromolecules as well as on target organs for toxicity. Given that the cancer process represents defects in terminal differentiation, growth control, and cell death, developmental studies of agents may provide an understanding of the activity of an agent that carries over to cancer assessment. Toxicity studies in animals by different routes of administration support comparison of absorption and metabolism by those routes. Data on human variability in standard clinical tests may provide insight into the range of human susceptibility and the common mechanisms of agents that affect the tested parameters.

2.3.5. Events Relevant to Mode of Carcinogenic Action

Knowledge of the biochemical and biological changes that precede tumor development (which include but are not limited to mutagenesis, increased cell proliferation, inhibition of programmed cell death, and receptor activation) may provide important insight for determining whether a cancer hazard exists and may help inform the dose-response relationship below the range of observable tumor response. Because cancer results from a series of genetic alterations in the genes that control cell growth, division, and differentiation (Vogelstein et al., 1988; Hanahan and Weinberg, 2000; Kinzler and Vogelstein, 2002), the ability of an agent to affect genotype (and hence gene products) or gene expression is of obvious importance in evaluating its influence on the carcinogenic process. Initial and key questions to examine are: Does the agent (or its metabolite) interact directly with DNA, leading to mutations that bring about changes in gene products or gene expression? Does the agent bring about effects on gene expression via other nondirect DNA interaction processes?

Furthermore, carcinogenesis involves a complex series and interplay of events that alter the signals a cell receives from its extracellular environment, thereby promoting uncontrolled

growth. Many, but not all, mutagens are carcinogens, and some, but not all, agents that induce cell proliferation lead to tumor development. Thus, understanding the range of key steps in the carcinogenic process upon which an agent might act is essential for evaluating its mode of action. Endpoints that provide insight into an agent's ability to alter gene products and gene expression, together with other features of an agent's potential mode of carcinogenic action, are discussed below.

2.3.5.1. *Direct DNA-Reactive Effects*

It is well known that many carcinogens are electrophiles that interact with DNA, resulting in DNA adducts and breakage (referred to in these guidelines as direct DNA effects). Usually during the process of DNA replication, these DNA lesions can be converted into mutations and chromosomal alterations, which then may initiate and otherwise contribute to the carcinogenic process (Shelby and Zeiger, 1990; Tinwell and Ashby, 1991; IARC, 1999). Thus, studies of mutations and other genetic lesions continue to be predictive in the assessment of potential human cancer hazard and in the understanding of an agent's mode of carcinogenic action.

EPA has published testing guidelines for detecting the ability of an agent to damage DNA and produce mutations and chromosomal alterations. Briefly, standard tests for gene mutations in bacteria and mammalian cells in vitro and in vivo and for structural chromosomal aberrations in vitro and in vivo are important examples of relevant methods. New molecular approaches such as mouse mutations and cancer transgenic models, are providing a means to examine mutation at tissue sites where the tumor response is observed (Heddle and Swiger, 1996; Tennant et al., 1999). Additionally, continued improvements in fluorescent-based chromosome staining methods (FISH, fluorescent in situ hybridization) will allow the detection of specific chromosomal abnormalities in relevant target tissues (Tucker and Preston, 1998).

Endpoints indicative of DNA damage but not measures of mutation per se, such as DNA adducts or strand breakage, can be detected in relevant target tissues and thus contribute to evaluating an agent's mutagenic potential. Evidence of chemical-specific DNA adducts (e.g., reactions at oxygen sites in DNA bases or with ring nitrogens of guanine and adenine) provides information on a mutagen's ability to directly interact with DNA (La and Swenberg, 1996). It should be noted that an increase in DNA binding shown with a radioactive label incorporated in the chemical (e.g., C^{14}) may reflect a direct DNA-reactive mechanism, but this needs to be examined, because the label may reflect reuse of C^{14} in the synthesis of DNA rather than binding. Some planar molecules (e.g., 9-aminoacridine) intercalate between base pairs of DNA,

which results in a physical distortion in DNA that may lead to mutations when DNA replicates. As discussed below, some carcinogens do not interact directly with DNA, but they can produce increases in endogenous levels of DNA adducts (e.g., 8-hydroxyguanine) by indirect mechanisms.

2.3.5.2. *Indirect DNA Effects or Other Effects on Genes/Gene Expression*

Although some carcinogens may result in an elevation of mutations or cytogenetic anomalies, as detected in standard assays, they may do so by indirect mechanisms. These effects may be brought about by chemical-cell interactions rather than by the chemical (or its metabolite) directly interacting with DNA. An increase in mutations might be due to cytotoxic exposures causing regenerative proliferation or to mitogenic influences (Cohen and Ellwein, 1990). Increased cell division may elevate mutation by clonal expansion of initiated cells or by increasing the number of genetic errors by rapid cell division and reduced time for DNA repair. Some agents might result in an elevation of mutations by interfering with the enzymes involved in DNA repair and recombination (Barrett and Lee, 1992). Damage to certain critical DNA repair genes or other genes (e.g., the p53 gene) may result in genomic instability, which predisposes cells to further genetic alterations and increases the probability of neoplastic progression (Harris and Hollstein, 1993; Levine et al., 1994; Rouse and Jackson, 2002). Likewise, DNA repair processes may be saturated at certain doses of a chemical, leading to an elevation of genetic alterations.

The initiation of programmed cell death (apoptosis) can potentially be blocked by an agent, thereby permitting replication of cells carrying genetic errors that would normally be removed from the proliferative pool. For example, peroxisome proliferators can suppress apoptotic pathways (Shulte-Hermann et al., 1993; Bayly et al., 1994) that could enhance the carcinogenic process. At certain doses an agent may also generate reactive oxygen species that produce oxidative damage to DNA and other macromolecules (Chang et al. 1988; Kehrer, 1993; Clayson et al., 1994). The role of cellular alterations that are attributable to oxidative damage in tumorigenesis (e.g., 8-hydroxyguanine) is currently unclear.

Several carcinogens have been shown to induce aneuploidy (the loss or gain of chromosomes) (Barrett, 1992; Gibson et al., 1995). Aneuploidy can result in the loss of heterozygosity or genomic instability (Cavenee et al., 1986; Fearon and Vogelstein, 1990). Agents that cause aneuploidy typically interfere with the normal process of chromosome segregation by interacting with non-DNA targets such as the proteins needed for chromosome segregation and chromosome movement. All tumors (with the possible exception of some

leukemias and lymphomas) are aneuploid, but whether this chromosome imbalance is the cause or the effect of tumorigenesis is not clear. Thus, it is important to understand whether the agent induces aneuploidy as a key early event in the carcinogenic process or is necessary for tumor progression.

It is possible for an agent to alter gene expression by transcriptional, translational, or post-translational modifications. For example, perturbation of DNA methylation patterns may cause effects that contribute to carcinogenesis (Jones, 1986; Holliday, 1987; Goodman and Counts, 1993; Chuang et al., 1996; Baylin and Bestor, 2002). Overexpression of genes by DNA amplification has been observed in certain tumors (Vainio et al., 1992). Gene amplification may result from disproportionate DNA replication. Other mechanisms of altering gene expression may involve cellular reprogramming through hormonal or receptor-mediated mechanisms (Barrett, 1992; Ashby et al., 1994).

Both cell proliferation and programmed cell death are mandatory for the maintenance of homeostasis in normal tissues, and alterations in the level or rate of either are important elements of the carcinogenic process. The balance between the two directly affects the survival and growth of initiated cells as well as preneoplastic and tumor cell populations (i.e., increase in cell proliferation or decrease in cell death) (Cohen and Ellwein, 1990, 1991; Cohen et al., 1991; Bellamy et al., 1995). Thus, measurements of these events contribute to the weight of the evidence for cancer hazard prediction and to mode of action understanding. In studies of proliferative effects distinctions should be made between mitogenesis and regenerative proliferation (Cohen and Ellwein, 1990, 1991; Cohen et al., 1991).

In applying information from studies on cell proliferation and apoptosis to risk assessment, it is important to identify the tissues and target cells involved, to measure effects in both normal and neoplastic tissue, to distinguish between apoptosis and necrosis, and to determine the dose that affects these processes. Gap-junctional intercellular communication is believed to play a role in tissue and organ development and in the maintenance of a normal cellular phenotype within tissues. A growing body of evidence suggests that chemical interference with gap-junctional intercellular communication is a contributing factor in tumor development (Swierenga and Yamasaki, 1992; Yamasaki, 1995).

2.3.5.3. *Experimental Considerations in Evaluating Data on Precursor Events*

Most testing schemes for mutagenicity and other short-term assays were designed for hazard identification purposes; thus, these assays are generally conducted using acute exposures. For data on "precursor steps" to be useful in informing the dose-response curve for tumor

induction below the level of observation, it is important that data come from in vivo studies and from studies where exposure is repeated or given over an extended period of time. Although consistency of results across different assays and animal models provides a stronger basis for drawing conclusions, it is desirable to have data on the precursor event in the same target organ, sex, animal strain, and species as the tumor data. In evaluating an agent's mode of action, it is usually not sufficient to determine that some event commences upon dosing. It is important to understand whether it is a causal event that plays a key role in the process that leads to tumor development versus an effect of the cancer process itself or simply an associated event.

2.3.5.4. *Judging Data*

Criteria that are generally applicable for judging the adequacy of mechanistically based data include

- mechanistic relevance of the data to carcinogenicity,
- number of studies of each endpoint,
- consistency of results in different test systems and different species,
- similar dose-response relationships for tumor and mode of action-related effects,
- conduct of the tests in accordance with generally accepted protocols, and
- degree of consensus and general acceptance among scientists regarding interpretation of the significance and specificity of the tests.

Although important information can be gained from in vitro test systems, a higher level of confidence is generally given to data that are derived from in vivo systems, particularly those results that show a site concordance with the tumor data.

It is important to remember that when judging and considering the use of any data, the basic standard of quality, as defined by the EPA Information Quality Guidelines, should be satisfied.

2.4. BIOMARKER INFORMATION

Various endpoints can serve as biological markers of events in biological systems or samples. In some cases, these molecular or cellular effects (e.g., DNA or protein adducts, mutation, chromosomal aberrations, levels of thyroid-stimulating hormone) can be measured in blood, body fluids, cells, and tissues to serve as biomarkers of exposure in both animals and humans (Callemen et al., 1978; Birner et al., 1990). As such, they can

- act as an internal surrogate measure of chemical dose, representing, as appropriate, either recent exposure (e.g., serum concentration) or accumulated (e.g., hemoglobin adducts) exposure;

- help identify doses at which elements of the carcinogenic process are operating;

- aid in interspecies extrapolations when data are available from both experimental animal and human cells; and,

- under certain circumstances, provide insights into the possible shape of the dose-response curve below levels where tumor incidences are observed (e.g., Choy, 1993).

Genetic and other findings (such as changes in proto-oncogenes and tumor suppressor genes in preneoplastic and neoplastic tissue or, possibly, measures of endocrine disruption) can indicate the potential for disease and, as such, serve as biomarkers of effect. They, too, can be used in different ways:

- The spectrum of genetic changes in proliferative lesions and tumors following chemical administration to experimental animals can be determined and compared with that in spontaneous tumors in control animals, in animals exposed to other agents of varying structural and functional activities, and in persons exposed to the agent under study.

- They may provide a linkage to tumor response.

- They may help to identify subpopulations of individuals who may be at an elevated risk for cancer, for example, cytochrome P450 2D6/debrisoquine sensitivity for lung cancer (Caporaso et al., 1989) or inherited colon cancer syndromes (Kinzler et al., 1991; Peltomäki et al., 1993).

- As with biomarkers of exposure, it may be justified in some cases to use these endpoints for dose-response assessment or to provide insight into the potential

shape of the dose-response curve at doses below those at which tumors are induced experimentally.

In applying biomarker data to cancer assessment (particularly assessments based on epidemiologic data), an assessment should consider

- routes of exposure,
- exposure to mixtures,
- time after exposure,
- sensitivity and specificity of biomarkers, and
- dose-response relationships.

2.5. MODE OF ACTION—GENERAL CONSIDERATIONS AND FRAMEWORK FOR ANALYSIS

2.5.1. General Considerations

The interaction between the biology of the organism and the chemical properties of the agent determine whether there is an adverse effect. Thus, mode of action analysis is based on physical, chemical, and biological information that helps to explain key events in an agent's influence on development of tumors. The entire range of information developed in the assessment is reviewed to arrive at a reasoned judgment. An agent may work by more than one mode of action, both at different sites and at the same tumor site. At least some information bearing on mode of action (e.g., SAR, screening tests for mutagenicity) is present for most agents undergoing assessment of carcinogenicity, even though certainty about exact molecular mechanisms may be rare.

Inputs to mode of action analysis generally include tumor data in humans and animals and among structural analogues as well as the other key data. The more complete the data package and the generic knowledge about a given mode of action, the more confidence one has and the more one can replace or refine default positions with relevant information. Reasoned judgments are generally based on a data-rich source of chemical, chemical class, and tumor type-specific information. Many times there will be conflicting data and gaps in the information base; it is important to carefully evaluate these uncertainties before reaching any conclusion.

In making decisions about potential modes of action and the relevance of animal tumor findings to humans (Ashby et al., 1990), very often the results of chronic animal studies may give important clues. Some of the important factors to review include

- tumor types, for example, those responsive to endocrine influence or those produced by reactive carcinogens (Ashby and Tennant, 1991);

- number of tumor sites, sexes, studies, and species affected or unaffected (Tennant, 1993);

- influence of route of exposure, spectrum of tumors, and local or systemic sites;

- target organ or system toxicity, for example, urinary chemical changes associated with stone formation, effects on immune surveillance;

- presence of proliferative lesions, for example, hepatic foci, hyperplasias;

- progression of lesions from preneoplastic to benign to malignant with dose and time;

- ratio of malignant to benign tumors as a function of dose and time;

- time of appearance of tumors after commencing exposure;

- tumors invading locally, metastasizing, producing death;

- tumors at sites in laboratory animals with high or low spontaneous historical incidence;

- biomarkers in tumor cells, both induced and spontaneous, for example, DNA or protein adducts, mutation spectra, chromosome changes, oncogene activation; and

- shape of the dose-response curve in the range of tumor observation, for example, linear versus profound change in slope.

Some of the myriad ways in which information from chronic animal studies influences mode of action judgments include the following. Multisite and multispecies tumor effects are often associated with mutagenic agents. Tumors restricted to one sex or species may suggest an

influence restricted to gender, strain, or species. Late onset of tumors that are primarily benign or are at sites with a high historical background incidence or that show reversal of lesions on cessation of exposure may point to a growth-promoting mode of action. It is important to consider the possibility that an agent may act differently in different tissues or have more than one mode of action in a single tissue.

Simple knowledge of sites of tumor increase in rodent studies can give preliminary clues as to mode of action. Experience at the National Toxicology Program (NTP) indicates that substances that are DNA reactive and that produce gene mutations may be unique in producing tumors in certain anatomical sites, whereas tumors at other sites may arise from both mutagenic or nonmutagenic influences (Ashby and Tennant, 1991; Huff et al., 1991).

2.5.2. Evaluating a Hypothesized Mode of Action

2.5.2.1. *Peer Review*

This section contains a framework for evaluating a hypothesized mode of action. In reaching conclusions, the question of "general acceptance" of a mode of action should be tested as part of the independent peer review that EPA obtains for its assessment and conclusions. In some cases the mode of action may already have been established by development of a large body of research information and characterization of the phenomenon over time. In some cases there will have been development of an Agency policy (e.g., mode of action involving alpha-2u-globulin in the male rat, U.S. EPA, 1991b) or a series of previous assessments in which both the mode of action and its applicability to particular cases has been explored. If so, the assessment and its peer review can focus on the evidence that a particular agent acts in this mode. When necessary, the peer review should also evaluate the strengths and weaknesses of competing modes of action.

In other cases, the mode of action may not have previously been the subject of an Agency document. If so, the data to support both the mode of action and the associated activity of the agent should undergo EPA assessment and subsequent peer review.

2.5.2.2. *Use of the Framework*

The framework supports a full analysis of mode of action information, but it can also be used as a screen to decide whether sufficient information is available to evaluate or whether the data gaps are too substantial to justify further analysis. Mode of action conclusions are used to address the question of human relevance of animal tumor responses, to address differences in anticipated response among humans, such as between children and adults or men and women;

and as the basis of decisions about the anticipated shape of the dose-response relationship. Guidance on the latter appears in Section 3.

2.5.3. Framework for Evaluating Each Hypothesized Carcinogenic Mode of Action

This framework is intended to be an analytic tool for judging whether available data support a mode of carcinogenic action hypothesized for an agent. This mode of action framework was initiated by the International Programme for Chemical Safety (WHO, 1999). It is based upon considerations for causality in epidemiologic investigations originally articulated by Hill (1965) but later modified by others and extended to experimental studies. The original Hill criteria were applied to epidemiologic data, whereas this framework is applied to a much wider assortment of experimental data, so it retains the basic principles of Hill but is much modified in content.

The modified Hill criteria can be useful for organizing thinking about aspects of causation, and they are consistent with the scientific method of developing hypotheses and testing those hypotheses experimentally. During analysis by EPA, and as guidance for peer review, a key question is whether the data to support a mode of action meet the standards generally applied in experimental biology regarding inference of causation.

All pertinent studies are reviewed in analyzing a mode of action, and an overall weighing of evidence is performed, laying out the strengths, weaknesses, and uncertainties of the case as well as potential alternative positions and rationales. Identifying data gaps and research needs is also part of the assessment.

To show that a hypothesized mode of action is operative, it is generally important to outline the sequence of events leading to cancer, to identify key events that can be measured, and to weigh information to determine whether there is a causal relationship between events and cancer formation. It is not generally expected that the complete sequence will be known at the molecular level. Instead, empirical observations made at different levels of biological organization—biochemical, cellular, physiological, tissue, organ, and system—are analyzed.

Several important points should be considered when working with the framework:

- The topics listed for analysis should *not* be regarded as a checklist of necessary "proofs." The judgment of whether a hypothesized mode of action is supported by available data takes account of the analysis as a whole.

- The framework provides a structure for organizing the facts upon which conclusions as to mode of action rest. The purpose of using the framework is to make analysis transparent and to allow the reader to understand the facts and reasoning behind a conclusion.

- The framework does not dictate an answer. The weight of evidence that is sufficient to support a decision about a mode of action may be less or more, depending on the purpose of the analysis, for example, screening, research needs identification, or full risk assessment. To make the reasoning transparent, the purpose of the analysis should be made apparent to the reader.

- Toxicokinetic studies may contribute to mode of action analysis by identifying the active form of an agent that is central to the mode of action. Apart from contributing in this way, toxicokinetics studies may reveal effects of saturation of metabolic processes. These are not considered key events in a mode of action, but they are given separate consideration in assessing dose metrics and potential nonlinearity of the dose-response relationship.

- Generally, "sufficient" support is a matter of scientific judgment in the context of the requirements of the decision maker or in the context of science policy guidance regarding a certain mode of action.

- Even when a hypothesized mode of action is supported for a described response in a specific tissue, it may not explain other tumor responses observed, which need separate consideration in hazard and dose-response assessment.

In a risk assessment document, the analysis of a hypothesized mode of action should be presented before or with the characterization of an agent's potential hazard to humans.

For each tumor site, the mode of action analysis should begin with a description of the hypothesized mode of action and its sequence of key events (see Section 2.5.3.1). This should be followed by a discussion of various aspects of the experimental support for the hypothesized mode of action (see Section 2.5.3.2). The possibility of other modes of action also should be considered (see Section 2.5.3.3); if there is evidence for more than one mode of action, each should receive a separate analysis. Conclusions about the hypothesized mode of action should

address whether the mode of action is supported in animals and is relevant to humans and which populations or lifestages can be particularly susceptible (see Section 2.5.3.4).

2.5.3.1. *Description of the Hypothesized Mode of Action*

Summary description of the hypothesized mode of action. For each tumor site, the mode of action analysis should begin with a description of the hypothesized mode of action and its sequence of key events. If there is evidence for more than one mode of action, each receives a separate analysis.

Identification of key events. This is a consideration devised for this framework. In order to judge how well data support involvement of a key event in carcinogenic processes, the experimental definition of the event or events should be clear and repeatable. To support an association, experiments should define and measure an event consistently.

- Can a list of events be identified that are key to the carcinogenic process?
- Are the events well defined?

Pertinent observations may include, but are not limited to, receptor-ligand changes, cytotoxicity, cell cycle effects, increased cell growth, organ weight differences, histological changes, hormone or other protein perturbations, DNA and chromosome effects.

2.5.3.2. *Discussion of the Experimental Support for the Hypothesized Mode of Action*

The experimental support for the hypothesized mode of action should be discussed from several viewpoints patterned after the Hill criteria (see Section 2.2.1.3). For illustration, the explanation of each topic includes typical questions to be addressed to the available empirical data and experimental observations anticipated to be pertinent. The latter will vary from case to case. For a particular mode of action, certain observations may be established as essential in practice or policy, for example, measures of thyroid hormone levels in supporting thyroid hormone elevation as a key event in carcinogenesis.

Strength, consistency, specificity of association. A statistically significant association between events and a tumor response observed in well-conducted studies is generally supportive of causation. Consistent observations in a number of such studies with differing experimental designs increase that support, because different designs may reduce unknown biases. Studies showing "recovery," that is, absence or reduction of carcinogenicity when the event is blocked or diminished, are particularly important tests of the association. Specificity of the association,

without evidence of other modes of action, strengthens a causal conclusion. A lack of strength, consistency, and specificity of association weakens the causal conclusions for a particular mode of action.

- What is the level of statistical and biological significance for each event and for cancer?

- Do independent studies and different experimental hypothesis-testing approaches produce the same associations?

- Does the agent produce effects other than those hypothesized?

- Is the key event associated with precursor lesions?

Pertinent observations include tumor response associated with events (site of action logically relates to event[s]), precursor lesions associated with events, initiation-promotion studies, and stop/recovery studies.

Dose-response concordance. If a key event and tumor endpoints increase with dose such that the key events forecast the appearance of tumors at a later time or higher dose, a causal association can be strengthened. Dose-response associations of the key event with other precursor events can add further strength. Difficulty arises when an event is not causal but accompanies the process generally. For example, if tumors and the hypothesized precursor both increase with dose, the two responses will be correlated regardless of whether a causal relationship exists. This is similar to the issue of confounding in epidemiologic studies. Dose-response studies coupled with mechanistic studies can assist in clarifying these relationships.

- What are the correlations among doses producing events and cancer?

Pertinent observations include 2-year bioassay observation of lesions correlated with observations of hormone changes and the same lesions in shorter term studies or in interim sacrifice.

Temporal relationship. If an event is a cause of tumorigenesis, it must precede tumor appearance. An event may also be observed contemporaneously or after tumor appearance; these observations may add to the strength of association but not to the temporal association.

- What is the ordering of events that underlie the carcinogenic process?
- Is this ordering consistent among independent studies?

Pertinent observations include studies of varying duration observing the temporal sequence of events and tumorigenicity.

Biological plausibility and coherence. It is important that the hypothesized mode of action and the events that are part of it be based on current understanding of the biology of cancer to be accepted. If the body of information under scrutiny is consistent with other examples (including structurally related agents) for which the hypothesized mode of action is accepted, the case is strengthened. Because some modes of action can be anticipated to evoke effects other than cancer, the available toxicity database on noncancer effects, for example, reproductive effects of certain hormonal disturbances, can contribute to this evaluation.

- Is the mode of action consistent with what is known about carcinogenesis in general and for the case specifically?

- Are carcinogenic effects and events consistent across structural analogues?

- Is the database on the agent internally consistent in supporting the purported mode of action, including relevant noncancer toxicities?

Pertinent observations include the scientific basis for considering a hypothesized mode of action generally, given current state of knowledge of carcinogenic processes; previous examples of data sets showing the mode of action; data sets on analogues; and coherence of data in this case from cancer and noncancer toxicity studies.

2.5.3.3. *Consideration of the Possibility of Other Modes of Action*

The possible involvement of more than one mode of action at the tumor site should be considered. Pertinent observations that are not consistent with the hypothesized mode of action can suggest the possibility of other modes of action. Some pertinent observations can be consistent with more than one mode of action. Futhermore, different modes of action can operate in different dose ranges; for example, an agent can act predominantly through cytotoxicity at high doses and through mutagenicity at lower doses where cytotoxicity does not occur.

If there is evidence for more than one mode of action, each should receive a separate analysis. There may be an uneven level of experimental support for the different modes of action. Sometimes this can reflect disproportionate resources spent on investigating one particular mode of action and not the validity or relative importance of the other possible modes of action.

2.5.3.4. ***Conclusions About the Hypothesized Mode of Action***

Conclusions about the hypothesized mode of action should address the following questions:

(a) Is the hypothesized mode of action sufficiently supported in the test animals? Associations observed between key events and tumors may or may not support an inference of causation. The conclusion that the agent causes a sequence of key events that results in tumors is strengthened as more aspects of causation are satisfied and weakened as fewer are satisfied. Consistent results in different experiments that test the hypothesized mode of action build support for that mode of action. Replicating results in a similar experiment does not generally meaningfully strengthen the original evidence, and discordant results generally weaken that support. Experimental challenge to the hypothesized mode of action, where interrupting the sequence of key events suppresses the tumor response or enhancement of key events increases the tumor response, creates very strong support for the mode of action.

(b) Is the hypothesized mode of action relevant to humans? If a hypothesized mode of action is sufficiently supported in the test animals, the sequence of key precursor events should be reviewed to identify critical similarities and differences between the test animals and humans. The question of concordance can be complicated by cross-species differences in toxicokinetics or toxicodynamics. For example, the active agent can be formed through different metabolic pathways in animals and humans. Any information suggesting quantitative differences between animals and humans is flagged for consideration in the dose-response assessment. This includes the potential for different internal doses of the active agent or for differential occurrence of a key precursor event.

"Relevance" of a potential mode of action is considered in the context of characterization of hazard, not level of risk. Anticipated levels of human exposure are not used to determine whether the hypothesized mode of action is relevant to humans. Exposure information is integrated into the overall risk characterization.

The question of relevance considers all populations and lifestages. It is possible that the conditions under which a mode of action operates exist primarily in a particular population or

lifestage, for example, in those with a pre-existing hormonal imbalance. Other populations or lifestages may not be analogous to the test animals, in which case the question of relevance would be decided by inference.

Special attention should be paid to whether tumors can arise from childhood exposure, considering various aspects of development during these lifestages. Because the studies that support a mode of action are typically conducted in mature animals, conclusions about relevance during childhood generally rely on inference. There is currently no suggested default to settle the question of whether tumors arising through the hypothesized mode of action are relevant during childhood; understanding the mode of action implies that there are sufficient data (on either the specific agent or the general mode of action) to form a confident conclusion about relevance during childhood.

(c) Which populations or lifestages can be particularly susceptible to the hypothesized mode of action? If a hypothesized mode of action is judged relevant to humans, information about the key precursor events is reviewed to identify populations or lifestages that can be particularly susceptible to their occurrence. Although agent-specific data would provide the strongest indication of susceptibility, this review may rely on general knowledge about the precursor events and characteristics of individuals susceptible to these events. Any information suggesting quantitative differences between populations or lifestages should be flagged for consideration in the dose-response assessment (see Section 3.5). This includes the potential for a higher internal dose of the active agent or for an increased occurrence of a key precursor event. Quantitative differences may result in separate risk estimates for susceptible populations or lifestages.

The possibility that childhood is a susceptible period for exposure should be explicitly addressed. Generic understanding of the mode of action can be used to gauge childhood susceptibility, and this determination can be refined through analysis of agent-specific data.

2.6. WEIGHT OF EVIDENCE NARRATIVE

The *weight of evidence narrative* is a short summary (one to two pages) that explains an agent's human carcinogenic potential and the conditions that characterize its expression. It can stand alone and it can be useful to risk managers and nonexpert readers.

Weight of evidence should be presented as a narrative laying out the complexity of information that is essential to understanding the hazard and its dependence on the circumstances of exposure or the traits of an exposed population. For example, route of exposure can be used to qualify a hazard if an agent is not absorbed by some routes. Other examples are when an

agent's mode of action occurs only on reaching a minimum dose or a minimum duration. Another example is when a hazard is expressed disproportionately in individuals possessing a specific gene; such characterizations may follow from a better understanding of the human genome. Similarly, a hazard can be attributable to exposures during a susceptible lifestage on the basis of our understanding of human development.

To capture this complexity, a weight of evidence narrative generally includes

- conclusions about human carcinogenic potential and the conditions that characterize its expression (route, magnitude, and duration of exposure; susceptible populations or lifestages),

- a summary of the key evidence supporting these conclusions,

- a summary of the key default options invoked when the available information is inconclusive, and

- a summary of potential modes of action and how they reinforce the conclusions.

To provide some measure of clarity and consistency in an otherwise free-form narrative, weight of evidence descriptors should be used in the first sentence of the narrative. Applying a descriptor is a matter of judgment and cannot be reduced to a formula. Each descriptor may be applicable to a wide variety of potential data sets and weights of evidence. Descriptors represent points along a continuum of evidence; consequently, there are gradations and borderline cases that are clarified by the full narrative. Using descriptors within a narrative preserves and presents the complexity that is an essential part of the hazard characterization. **Risk managers should consider the entire range of information included in the narrative rather than focusing simply on the descriptor.** These narratives are intended to permit sufficient flexibility to accommodate new scientific understanding and new testing methods as they are developed and accepted by the scientific community and the public.

In borderline cases, the narrative explains the case for choosing one descriptor and discusses the arguments for considering but not choosing another. For example, between "suggestive" and "likely" or between "suggestive" and "inadequate," the explanation clearly communicates the information needed to appropriately consider the agent's carcinogenic potential in subsequent decisions.

Multiple descriptors can be used for a single agent when carcinogenesis is dose- or route-dependent. For example, if an agent causes point-of-contact tumors by one exposure route but adequate testing is negative by another route, then the agent could be described as likely to be carcinogenic by the first route but not likely to be carcinogenic by the second. Another example is when the mode of action is sufficiently understood to conclude that a key event in tumor development would not occur below a certain dose range. In this case, the agent could be described as likely to be carcinogenic above a certain dose range but not likely to be carcinogenic below that range.

Descriptors can be used when an agent has not been tested in a cancer bioassay but toxicokinetic and mode of action information are available to make a strong logical case through scientific inference. For example, if an agent is one of a well-defined class of agents that are understood to operate through a common mode of action, then in the narrative the untested agent would have the same descriptor as the class. Another example is when an untested agent's effects are understood to be caused by a human metabolite, in which case in the narrative the untested agent would have the same descriptor as the metabolite. As new testing methods are developed and used, assessments may increasingly be based on inferences from toxicokinetic and mode of action information in the absence of tumor studies in animals or humans.

When tumors occur at a site other than the point of initial contact, the descriptor generally applies to all exposure routes that have not been adequately tested at sufficient doses. An exception occurs when there is convincing toxicokinetic information that absorption does not occur by another route.

When a well-studied agent produces tumors only at a point of initial contact, the descriptor generally applies only to the exposure route producing tumors unless the mode of action is relevant to other routes. The rationale for this conclusion would be explained in the narrative.

Dose can represent a qualitative limitation on hazard. In some cases reaching a certain dose range can be a precondition for effects to occur, as when cancer is secondary to another toxic effect that appears only above a certain dose. In other cases exposure duration can be a precondition for hazard if effects occur only after exposure is sustained for a certain duration. These qualitative considerations differ from the quantitative issues of relative absorption or potency at different dose levels, which are addressed in the dose-response assessment.

When multiple bioassays have led to a borderline case, mode of action studies are likely to hold the key to resolution of the more appropriate descriptor. When bioassays are few, further

bioassays to replicate a study's results or to investigate the potential for effects in another sex, strain, or species may be useful.

When there are few pertinent data, the descriptor makes a statement about the database, for example, "inadequate information to assess carcinogenic potential" or a database that provides "suggestive evidence of carcinogenic potential." With more information, the descriptor expresses a conclusion about the agent's carcinogenic potential to humans. If the conclusion is positive, the agent could be described as "likely to be carcinogenic to humans" or (with strong evidence) "carcinogenic to humans." If the conclusion is negative, the agent could be described as "not likely to be carcinogenic to humans."

The following descriptors may be of value as part of the weight-of-evidence narrative:

"Carcinogenic to Humans"

This descriptor indicates strong evidence of human carcinogenicity. It covers different combinations of evidence:

- This descriptor is appropriate when there is convincing epidemiologic evidence of a causal association between human exposure and cancer.

- Exceptionally, this descriptor is equally appropriate with a lesser weight of epidemiologic evidence that is strengthened by other lines of evidence. It can be used when all of the following conditions are met: (a) there is strong evidence of an association between human exposure and either cancer or the key precursor events of the agent's mode of action but not enough for a causal association, and (b) there is extensive evidence of carcinogenicity in animals, and (c) the mode(s) of carcinogenic action and associated key precursor events have been identified in animals, and (d) the key precursor events that precede the cancer response in animals are anticipated to occur in humans and progress to tumors, based on available biological information.

"Likely to Be Carcinogenic to Humans"

This descriptor is appropriate when the weight of the evidence is adequate to demonstrate carcinogenic potential to humans but does not reach the weight of evidence for the descriptor "carcinogenic to humans." Adequate evidence consistent with this descriptor covers a broad

spectrum. Some examples to illustrate the broad range of data combinations that are covered by this descriptor include

- an agent with some evidence of an association between human exposure and cancer, with or without evidence of carcinogenicity in animals.

- an agent that has tested positive in more than one species, sex, strain, site, or exposure route, with or without evidence of carcinogenicity in humans;

- a positive study that indicates a highly significant result, for example, an uncommon tumor, a high degree of malignancy, or an early age at onset;

- a positive study that is strengthened by other lines of evidence, for example, some evidence of an association between human exposure and cancer (but not enough to infer a causal association), or evidence that the agent or an important metabolite causes events generally known to be associated with tumor formation (such as DNA reactivity or effects on cell growth control) likely to be related to the tumor response in this case; or

- a robust animal tumor response in a single experiment that is assumed to be relevant to humans.

Although the term "likely" can have a probabilistic connotation in other contexts, its use as a weight of evidence descriptor does not correspond to a quantifiable probability. This is because the data that support cancer assessments generally are not suitable for numerical calculations of the probability that an agent is a carcinogen. The weight of evidence descriptor "likely to be carcinogenic to humans" may be taken loosely to imply that an agent is more likely than not—but is not certain—to cause cancer in humans. Other health agencies have expressed a comparable weight of evidence using terms such as "reasonably anticipated to be a human carcinogen" (NTP) or "probably carcinogenic to humans" and "possibly carcinogenic to humans" (International Agency for Research on Cancer).

"Suggestive Evidence of Carcinogenic Potential"

This descriptor of the database is appropriate when the weight of evidence is suggestive of carcinogenicity; a concern for potential carcinogenic effects in humans is raised, but the data are judged not sufficient for a stronger conclusion. This descriptor covers a spectrum of evidence associated with varying levels of concern for carcinogenicity, ranging from a positive result in the only study on an agent to a single positive result in an extensive database that includes negative studies in other species. Depending on the extent of the database, additional studies may or may not provide further insights. Some examples include

- a marginal increase in tumors observed only in a single animal or human study;
- a slight increase in a tumor with a high background rate in that sex and strain;
- a statistically significant increase at one dose only but no significant response at the other doses or trend overall; or
- evidence of a response in a study whose power, design, or conduct limits the ability to draw a confident conclusion.

"Inadequate Information to Assess Carcinogenic Potential"

This descriptor of the database is appropriate when available data are judged inadequate for applying one of the other descriptors. Additional studies generally would be expected to provide further insights. Some examples include

- little or no pertinent information.
- conflicting evidence, that is, some studies provide evidence of carcinogenicity but other studies of equal quality in the same sex and strain are negative. (*Differing results*, that is, positive results in some studies and negative results in one or more different experimental systems, do not constitute *conflicting evidence,* as the term is used here. Depending on the overall weight of evidence, differing results can be considered either suggestive evidence or likely evidence.)

- negative results that are not sufficiently robust for the descriptor, "not likely to be carcinogenic to humans."

"Not Likely to Be Carcinogenic to Humans"

This descriptor is appropriate when the available data are considered robust for deciding that there is no basis for human hazard concern. In some instances, there can be positive results in experimental animals when there is strong, consistent evidence that each mode of action in experimental animals does not operate in humans. The judgment may be based on

- animal evidence that demonstrates lack of carcinogenic effect in well-designed and well-conducted studies in at least two appropriate animal species (in the absence of other animal or human data suggesting a potential for cancer effects),

- extensive experimental evidence showing that the only carcinogenic effects observed in animals are not relevant to humans,

- convincing evidence that carcinogenic effects are not likely by a particular exposure route (see Section 2.3.3), or

- convincing evidence that carcinogenic effects are not likely below a defined dose range.

A descriptor of "not likely" applies only to the circumstances supported by the data. For example, an agent may be "not likely to be carcinogenic" by one route but not necessarily by another.

Multiple Descriptors

As discussed previously, more than one descriptor can be used when an agent's effects differ by dose or exposure route. For example, an agent may be "carcinogenic" by one exposure route but "not likely to be carcinogenic" by a route by which it is not absorbed. Another example is when an agent is "likely to be carcinogenic" above a specified dose but "not likely to be carcinogenic" below that dose because a key event in tumor formation does not occur below that dose.

Some examples show how a single positive study can lead to a wide range of descriptors, depending on the information provided by other studies: If there is only a single study and it is positive, a descriptor of "likely to be carcinogenic to humans" would generally be appropriate. If there are also some negative studies in another species ("differing results"), a descriptor of "suggestive evidence of carcinogenic potential" might be appropriate. If the negative studies are in the same test system as the positive study and are of equal quality ("conflicting evidence"), the descriptor "inadequate information to assess carcinogenic potential" could be used. If there are adequate negative studies in two species and if there is sufficient evidence to determine that the tumors in the positive study are caused by a mode of action that is not relevant to humans, the descriptor "not likely to be carcinogenic to humans" can be considered. In each case, the descriptor is determined by the overall weight of evidence, taking into account all the studies.

2.7. HAZARD CHARACTERIZATION

The *hazard characterization* contains the hazard information needed for a full risk characterization (U.S. EPA, 2000b). It presents the results of the hazard assessment and explains how the weight of evidence conclusion was reached. The hazard characterization summarizes, in plain language, conclusions about the agent's potential effects, whether they can be expected to depend qualitatively on the circumstances of exposure, and who can be expected to be especially susceptible. It discusses the extent to which these conclusions are supported by data or are the result of default options invoked because the data are inconclusive. It explains how complex cases with differing results in different studies were resolved. The hazard characterization highlights the major issues addressed in the hazard assessment and discusses alternative interpretations of the data and the degree to which they are supportable scientifically and are consistent with EPA guidelines.

When the conclusion is supported by mode of action information, the hazard characterization also provides a clear summary of the mode of action conclusions (see Section 2.5.3.4), including the completeness of the data, the strengths and limitations of the inferences made, the potential for other modes of action, and the implications of the mode of action for selecting viable approaches to the dose-response assessment. The hazard characterization also discusses the extent to which mode of action information is available to address the potential for disproportionate risks in specific populations or lifestages or the potential for enhanced risks on the basis of interactions with other agents or stressors.

Topics that should be addressed in a hazard characterization include

- summary of the results of the hazard assessment;

- identification of susceptible populations and lifestages, especially attending to children, infants, and fetuses;

- conclusions about the agent's mode of action and implications for selecting approaches to the dose-response assessment;

- identification of the available lines of evidence (animal bioassays, epidemiologic studies, toxicokinetic information, mode of action studies, and information about structural analogues or metabolites), highlighting data quality and coherence of results from different lines of evidence; and
- strengths and limitations of the hazard assessment, highlighting significant issues in interpreting the data, alternative interpretations that are considered equally plausible, critical data gaps, and default options invoked when the available information is inconclusive.

3. DOSE-RESPONSE ASSESSMENT

Dose-response assessment estimates potential risks to humans at exposure levels of interest. Dose-response assessments are useful in many applications: estimating risk at different exposure levels, estimating the risk reduction for different decision options, estimating the risk remaining after an action is taken, providing the risk information needed for benefit/cost analyses of different decision options, comparing risks across different agents or health effects, and setting research priorities. The purpose of the assessment should consider the quality of the data available, which will vary from case to case.

An analysis is developed from each study that reports quantitative data on dose and response. Alternative measures of dose are available for analyzing human and animal studies (see Section 3.1). A two-step approach distinguishes analysis of the dose-response data from inferences made about lower doses. The first step is an analysis of dose and response in the range of observation of the experimental and epidemiologic studies (see Section 3.2). Modeling is encouraged to incorporate a wide range of experimental data into the dose-response assessment (see Sections 3.1.2, 3.2.1, 3.2.2, 3.3.3). The modeling yields a POD that marks the boundary between the range of observation and the range of extrapolation to lower doses (see Sections 3.2.4, 3.2.5). The second step is extrapolation to lower doses (see Section 3.3). The extrapolation approach considers what is known about the agent's mode of action (see Section 3.3.1). Both linear and nonlinear approaches are available (see Sections 3.3.4, 3.3.3). When multiple estimates can be developed, they are combined in a way that best represents human cancer risk (see Section 3.3.5). Special consideration is given to describing dose-response differences attributable to different human exposure scenarios (see Section 3.4) and to susceptible populations and lifestages (see Section 3.5). It is important to discuss significant uncertainties encountered in the analysis (see Section 3.6) and to characterize other important aspects of the dose-response assessment (see Section 3.7).

The scope, depth, and use of a dose-response assessment vary in different circumstances. Although the quality of dose-response data is not necessarily related to the weight of evidence descriptor, dose-response assessments are generally completed for human carcinogens and likely human carcinogens. When there is suggestive evidence, the Agency generally would not attempt a dose-response assessment, as the nature of the data generally would not support one; however, when the evidence includes a well-conducted study, quantitative analyses may be useful for some purposes, for example, providing a sense of the magnitude and uncertainty of potential risks, ranking potential hazards, or setting research priorities. In each case, the rationale for the

quantitative analysis is explained, considering the uncertainty in the data and the suggestive nature of the weight of evidence. These analyses generally would not be considered Agency consensus estimates. Dose-response assessments are generally not done when there is inadequate evidence, although calculating a bounding estimate from a nonpositive epidemiologic study can indicate the study's level of sensitivity and capacity to detect risk levels of concern.

Cancer is a collection of several diseases that develop through cell and tissue changes over time. Dose-response assessment procedures based on tumor incidence have seldom taken into account the effects of key precursor events within the whole biological process due to lack of empirical data and understanding about these events. In this discussion, response data include measures of key precursor events considered integral to the carcinogenic process in addition to tumor incidence. These responses may include changes in DNA, chromosomes, or other key macromolecules; effects on growth signal transduction, including induction of hormonal changes; or physiological or toxic effects that include proliferative events diagnosed as precancerous but not pathology that is judged to be cancer. Analysis of such responses may be done along with that of tumor incidence to enhance the tumor dose-response analysis. If dose-response analysis of nontumor key events is more informative about the carcinogenic process for an agent, it is used in lieu of, or in conjunction with, tumor incidence analysis for the overall dose-response assessment.

As understanding of mode of action improves and new types of data become available, dose-response assessment will continue to evolve. These guidelines encourage the development and application of new methods that improve dose-response assessment by reflecting new scientific understanding and new sources of information.

3.1. ANALYSIS OF DOSE

For each effect observed, dose-response assessment should begin by determining an appropriate *dose metric*. The objective is to use the available data to estimate, as closely as possible, the delivered dose of the active agent at the target organ or cell of the species studied. When the delivered dose cannot be determined with confidence, dose-response assessment proceeds with another dose metric, for example, the average daily dose of the administered agent.

Selection of an appropriate dose metric considers what data are available and what is known about the agent's mode of action at the target site. The dose metric specifies

- the active agent (administered agent or a metabolite),

- proximity to the target site (exposure concentration, potential dose, internal dose, or delivered dose,[5] reflecting increasing proximity), and

- the time component of the effective dose (cumulative dose, average dose, peak dose, or body burden).

Analyses can be based on estimates of animal dose metrics or human dose metrics. The assessment should describe the approach used to select a dose metric and the reasons for this approach. The final analysis, however, should determine an equivalent human dose. This facilitates comparing results from different datasets and effects by using equivalent human dose as a common metric. When appropriate, it may be necessary to convert doses across exposure routes.

Timing of exposure can also be important. When there is a susceptible lifestage, doses during the susceptible period are not equivalent to doses at other times, and they would be analyzed separately.

3.1.1. Standardizing Different Experimental Dosing Regimens

Complex dosing regimens are often present in experimental and epidemiologic studies. The resulting internal dose depends on many variables, including concentration, duration, frequency of administration, and duration of recovery periods between administrations. Internal dose also depends on variables that are intrinsic to the exposed individual, such as lifestage and rates of metabolism and clearance. To facilitate comparing results from different study designs and to make inferences about human exposures, a summary estimate of dose may be derived for a complex dosing regimen.

Toxicokinetic modeling is the preferred approach for estimating dose. Toxicokinetic models generally consider a dose profile over time. More complex models can reflect sources of intrinsic variation, such as polymorphisms in metabolism and clearance rates. When a robust

[5] *Exposure* is 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 an 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 (e.g., 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. (U.S. EPA, 1992a)

model is not available, or when the purpose of the assessment does not warrant developing a model, simpler approaches may be used.

For chronic-dosing studies, the cumulative dose received should be expressed as an average over the duration of the study. This implies that a high dose received over a short duration at any period in life is equivalent to a commensurately lower dose spread over a lifetime. Uncertainty increases as the duration becomes shorter or the intermittent doses become more intense. Moreover, doses during a susceptible period are not equivalent to doses at other times. For these reasons, cumulative dose may be replaced by a more appropriate dose metric when indicated by the data.

For mode of action studies, the daily dose should be calculated over a duration that reflects the time to occurrence of the key precursor effects. Mode of action studies are often of limited duration, as the precursors can be observed after less-than-chronic dosing. When the experimental dosing regimen is specified on a weekly basis (for example, 4 hours a day, 5 days a week), the daily dose may be averaged over the week.

Doses in studies at the cellular or molecular level can be difficult to relate to organ- or organism-level dose metrics. Toxicokinetic modeling can sometimes be used to relate doses at the cellular or molecular level to doses at higher levels of organization.

3.1.2. Toxicokinetic Modeling

Physiologically based toxicokinetic modeling is potentially the most comprehensive way to account for biological processes that determine internal dose. Models are based on blood flow between physiological compartments and simulate the relationship between applied dose and internal dose. Toxicokinetic models generally need data on absorption, distribution, metabolism, and elimination of the administered agent and its metabolites.

Additionally, in the case of inhaled dose, models can explicitly characterize the geometry of the respiratory tract and the airflow through it, as well as the interaction of this airflow with the entrained particles or fibers and gases (Kimbell et al., 2001; Subramaniam et al., 2003). Because of large interspecies differences in airway morphometry, such models are particularly useful in interspecies extrapolations. Nonetheless, because of large interindividual differences in airway morphometry, particularly in humans, such models may not be representative of human populations.

Toxicokinetic models can improve dose-response assessment by revealing and describing nonlinear relationships between applied and internal dose. Nonlinearity observed in a dose-response curve often can be attributed to toxicokinetics (Hoel et al., 1983; Gaylor et al., 1994),

involving, for example, saturation or induction of enzymatic processes at high doses. Toxicokinetic processes tend to become linear at low doses (Hattis, 1990).

A discussion of confidence should accompany the presentation of model results and 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?). When a delivered dose measure is used in animal-to-human extrapolation, the assessment discusses the confidence of the target tissue and its toxicodynamics being the same in both species (see Section 3.6). Toxicokinetic modeling results may be presented as the preferred method of estimating equivalent human doses or in parallel with default procedures (see Section 3.1.3), depending on the confidence in the modeling.

3.1.3. Cross-species Scaling Procedures

Standard cross-species scaling procedures are available when the data are not sufficient to support a toxicokinetic model or when the purpose of the assessment does not warrant developing one. The aim is to define dose levels for humans and animals that are expected to produce the same degree of effect (U.S. EPA, 1992b), taking into account differences in scale between test animals and humans in size and in lifespan.

For oral exposures, doses should be scaled from animals to humans on the basis of equivalence of $mg/kg^{3/4}$-d (milligrams of the agent normalized by the 3/4 power of body weight per day) (U.S. EPA, 1992b). The 3/4 power is consistent with current science, including empirical data that allow comparison of potencies in humans and animals, and it is also supported by analysis of the allometric variation of key physiological parameters across mammalian species. It is generally more appropriate at low doses, where sources of nonlinearity such as saturation of enzyme activity are less likely to occur. This scaling is intended as an unbiased estimate rather than a conservative one. Equating exposure concentrations in food or water is an alternative version of the same approach, because daily intakes of food or water are approximately proportional to the 3/4 power of body weight.

The aim of these cross-species scaling procedures is to estimate administered doses in animals and humans that result in equal lifetime risks. It is useful to recognize two components of this equivalence: *toxicokinetic equivalence*, which determines administered doses in animals and humans that yield equal tissue doses, and *toxicodynamic equivalence*, which determines

tissue doses in animals and humans that yield equal lifetime risks (U.S. EPA, 1992b). Toxicokinetic modeling (see Section 3.1.2) addresses factors associated with toxicokinetic equivalence, and toxicodynamic modeling (see Section 3.2.2) addresses factors associated with toxicodynamic equivalence. When toxicokinetic modeling is used without toxicodynamic modeling, the dose-response assessment develops and supports an approach for addressing toxicodynamic equivalence, perhaps by retaining some of the cross-species scaling factor (e.g., using the square root of the cross-speies scaling factor or using a factor of 3 to cover toxicodynamic differences between animals and humans, as is done in deriving inhalation reference concentrations (EPA 1994)).

When assessing risks from childhood exposure, the $mg/kg^{3/4}$-d scaling factor does not use the child's body weight (U.S. EPA, 1992b). This reflects several uncertainties in extrapolating risks to children:

- The data supporting the $mg/kg^{3/4}$-d scaling factor were derived for differences across species and do not apply as well to differently sized individuals of the same species or to different lifestages.

- Using the child's body weight in the $mg/kg^{3/4}$-d scaling factor would erroneously imply that the child's intake, metabolism, and clearance are well described by such scaling.

- In addition to metabolic differences, there are also important toxicodynamic differences; for example, children have faster rates of cell division than do adults.

For inhalation exposures, an equivalent human concentration should be calculated using EPA's methods for deriving inhalation reference concentrations (U.S. EPA, 1994), which give preference to the use of toxicokinetic modeling. Otherwise, mathematical dosimetry models of particle or gas deposition in animals and humans are applied to yield a human-equivalent concentration. As with oral exposures, when toxicokinetic modeling or dosimetry modeling is used without toxicodynamic modeling, the dose-response assessment develops and supports an approach for addressing toxicodynamic equivalence.

The dosimetry models typically use a default breathing rate and respiratory tract dimensions for an adult (U.S. EPA, 1994). Children and adults breathing the same concentration of an agent (such as a reactive gas) may receive different doses to the body or lungs (U.S.

EPA, 2002b). A generalized approach to assessing such differences between children and adults is a comparison of breathing rates relative to size of the body or lungs (U.S. EPA, 2002b). The human respiratory system passes through several distinct stages of maturation and growth during the first several years of life and into adolescence (Pinkerton and Joad, 2000) during which the ratio of breathing rate to lung surface area may be markedly different (U.S. EPA, 2002b). With certain assumptions, the models can be adapted by scaling breathing rates and respiratory tract dimensions for a child's size (U.S. EPA, 1994). This scaling should be undertaken with caution because of the correlations between breathing rate, respiratory tract dimensions, and body weight. Properly done, the comparison of human-equivalent concentrations for an adult and child can indicate whether it is important to carry both concentrations forward in the dose-response assessment or whether a verbal characterization of the difference between the two will suffice.

3.1.4. Route Extrapolation

Often an assessment based on studies of one exposure route is applied to another exposure route. Route-to-route extrapolation has both qualitative and quantitative aspects. For the qualitative aspect, the assessor should weigh the degree to which positive results by one exposure route support a judgment that similar results would be expected by another route. In general, confidence in making such a judgment is strengthened when tumors are observed at a site distant from the portal of entry and when absorption is similar through both routes. In the absence of contrary data, a qualitative default option can be used that if the agent is absorbed through an exposure route to give an internal dose, it may be carcinogenic by that route.

When a qualitative extrapolation can be supported, quantitative extrapolation may still be problematic in the absence of adequate data. The differences in biological processes among routes of exposure (oral, inhalation, dermal) can be great because of, for example, first-pass effects and different results from different exposure patterns. There is no generally applicable method for accounting for these differences in uptake processes in a quantitative route-to-route extrapolation of dose-response data in the absence of good data on the agent of interest. Therefore, route-to-route extrapolation of dose data relies on a case-by-case analysis of available data. When good data on the agent itself are limited, an extrapolation analysis can be based on expectations from physical and chemical properties of the agent, properties and route-specific data on structurally analogous compounds, or in vitro or in vivo uptake data on the agent.

Route-to-route uptake models may be applied if model parameters are suitable for the compound of interest. Such models are currently considered interim methods; further model

development and validation is awaiting the development of more extensive data. For screening or hazard ranking, route-to-route extrapolation may be based on assumed quantitative comparability as a default, as long as it is reasonable to assume absorption by compared routes. When route-to-route extrapolation is used, the assessor's degree of confidence in both the qualitative and quantitative extrapolation is discussed in the assessment and highlighted in the dose-response characterization.

Toxicokinetic modeling can be used to compare results of studies by different exposure routes. Results can also be compared on the basis of internal dose for effects distant from the point of contact.

Route extrapolation can be used to understand how internal dose and subsequent effects depend on exposure route. Route extrapolation can also determine whether testing by different exposure routes has achieved similar internal doses, which can be important in determining whether testing is adequate to conclude that an agent causes effects by one route but not by another.

3.2. ANALYSIS IN THE RANGE OF OBSERVATION

The principle underlying these guidelines is to use approaches that include as much information as possible. Quantitative information about key precursor events can be used to develop a toxicodynamic model. Alternatively, such information can be fitted by empirical models to extend the dose-response analysis of tumor incidence to lower doses and response levels. The analysis in the range of observation is used to establish a POD that marks the boundary between the range of observation and the range of extrapolation to lower doses (see Section 3.3).

3.2.1. Analysis of Epidemiologic Studies

Analysis of epidemiologic studies depends on the type of study and quality of data, particularly the availability of quantitative measures of exposure. The objective is a dose-response curve that estimates the incidence of cancer attributable to exposure to the agent. In some cases, estimation of the number of cancer cases expected in a population (sometimes called "population risk") may be appropriate. Also in some cases, the agent can have discernible interactive effects with another agent, making it possible to estimate the contribution of each agent as a risk factor for the effects of the other. The analysis is tailored to the nature of each study, with due consideration of the consequences of study design. For example:

- Many studies collect information from death certificates, which leads to estimates of mortality rather than incidence. Because survival rates vary for different cancers, the analysis can be improved by adjusting mortality figures to reflect the relationship between incidence and mortality.

- Competing risks in a study population can limit the observed occurrence of cancer. The analysis can be improved by correcting for competing risks that are not similar in exposed and comparison groups.

- Comparison groups that are not free from exposure to the agent can bias the risk estimates toward zero. The analysis can be improved by considering background exposures in the exposed and comparison groups.

Some study designs can yield only a partial characterization of the overall risk, as, for example, in studies that

- investigate only one effect (typical of many case-control studies),
- include only one population segment (e.g., male workers), or
- include only one lifestage (e.g., childhood leukemia following maternal exposure to contaminated drinking water).

To obtain a fuller characterization that includes risks of other cancers, estimates from these studies can be supplemented with estimates from other studies that investigated other cancers, population segments, or lifestages (see Section 3.3.5).

The latent period for cancer implies that exposures immediately preceding the detection of a tumor would be less likely to have contributed to its development and, therefore, may count less in the analysis. Study subjects who were first exposed near the end of the study would not have had adequate time since exposure for cancer to develop, therefore, analysis of their data may be similar to analysis of data for those who were not exposed.

For epidemiologic studies, analysis by linear models in the range of observation should be used unless the fit is poor. This is justified by the relatively small dose range observed in many epidemiologic studies, which makes it difficult to discern the shape of the dose-response curve. Exposure misclassification and errors in exposure estimation also obscure the shape of the dose-response curve. When these errors are unsystematic, or random, the result is to bias the

risk estimates toward zero. When a linear model fits poorly, more flexible models that allow for low-dose linearity, for example, a linear-quadratic model or a Hill model (Murrell et al 1998), are considered next.

When several studies are available for dose-response analysis, *meta-analysis* can provide a systematic approach to weighing positive and nonpositive studies and calculating an overall risk estimate with greater precision. Issues considered include the comparability of studies, heterogeneity across studies, and the potential for a single large study to dominate the analysis. Confidence in a meta-analysis is increased when it considers study quality, including definition of the study population and comparison group, measurement of exposure, potential for exposure misclassification, adequacy of follow-up period, and analysis of confounders (see Section 2.2.1.2).

3.2.2. Toxicodynamic ("Biologically Based") Modeling

Toxicodynamic modeling can be used when there are sufficient data to ascertain the mode of action (see Section 2.5) and quantitatively support model parameters that represent rates and other quantities associated with the key precursor events of the mode of action. Toxicodynamic modeling is potentially the most comprehensive way to account for the biological processes involved in a response. Models reflect the sequence of key precursor events that lead to cancer. Toxicodynamic models can improve dose-response assessment by revealing and describing nonlinear relationships between internal dose and cancer response. Such models are generally the better approach for analysis in the range of observation, provided the purpose of the assessment justifies the effort involved.

If a new model is developed for a specific agent, extensive data on the agent are important for identifying the form of the model, estimating its parameters, and building confidence in its results. Conformance to the observed tumor incidence data does not establish a model's predictive validity, as a model can be overparameterized to fit a given dataset. Peer review, including both an examination of the scientific basis supporting the model and an independent evaluation of the model's performance, is an essential part of evaluating the new model.

If a standard model already exists for the agent's mode of action, the model can be adapted for the agent by using agent-specific data to estimate the model's parameters. An example is the two-stage clonal expansion model developed by Moolgavkar and Knudson (1981) and Chen and Farland (1991). These models continue to be improved upon.

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 tumor incidence data, it is often the case that different combinations of parameter estimates can yield similar results in the observed range. For this reason, critical parameters (e.g., mutation rates and cell birth and death rates) are estimated from laboratory studies and not by curve-fitting to tumor incidence data (Portier, 1987). This approach reduces model uncertainty (see Section 3.6) 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.

Toxicodynamic modeling can provide insight into the relationship between tumors and key precursor events. For example, a model that includes cell proliferation can be used to explore the extent to which small increases in the cell proliferation rate can lead to large lifetime tumor incidences (Gaylor and Zheng, 1996). In this way, toxicodynamic modeling can be used to select and characterize an appropriate precursor response level (see Section 3.2.4, 3.2.5).

3.2.3. Empirical Modeling ("Curve Fitting")

When a toxicodynamic model is not available or when the purpose of the assessment does not warrant developing such a model, empirical modeling (sometimes called "curve fitting") should be used in the range of observation. A model can be fitted to data on either tumor incidence or a key precursor event. Goodness-of-fit to the experimental observations is not by itself an effective means of discriminating among models that adequately fit the data (OSTP, 1985). Quantitative data on precursors can be used in conjunction with, or in lieu of, data on tumor incidence to extend the dose-response curve to lower doses. Caution is used with rates of molecular events such as mutation or cell proliferation or signal transduction. Such rates can be difficult to relate to cell or tissue changes overall. The timing of observations of these phenomena, as well as the cell type involved, need to be linked to other precursor events to ensure that the measurement is truly a key event (see Section 2.5).

For incidence data on either tumors or a precursor, an established empirical procedure should be used to provide objectivity and consistency among assessments. The procedure models incidence, corrected for background, as an increasing function of dose. The model is sufficiently flexible in the observed range to fit linear and nonlinear datasets. Additional judgment and perhaps an alternative analysis are used when the procedure fails to yield reliable results. For example, when the model fit is poor, the highest dose is often omitted in cases where it is judged that the highest dose reflects competing toxicity that is more relevant at high doses

than at lower doses. Another example is when there are large differences in survival across dose groups; here, a more detailed model that includes time-to-tumor or time-to-event information may be useful.

For continuous data on key precursor effects, an empirical model can be chosen on the basis of the structure of the data. The rationale for the choice of model, the alternatives considered and rejected, and a discussion of model uncertainty are included in the dose-response characterization.

3.2.4. Point of Departure

For each tumor response, a POD from the observed data should be estimated to mark the beginning of extrapolation to lower doses. The POD is an estimated dose (expressed in human-equivalent terms) near the lower end of the observed range without significant extrapolation to lower doses.

The POD is used as the starting point for subsequent extrapolations and analyses. For linear extrapolation, the POD is used to calculate a *slope factor* (see Section 3.3.4), and for nonlinear extrapolation the POD is used to calculate a *reference dose* or *reference concentration* (see Section 3.3.3). In a risk characterization, the POD is part of the determination of a *margin of exposure* (see Section 5.4). With appropriate adjustments, it can also be used as the basis for *hazard rankings* that compare different agents or health effects.

The goal is to use the lowest POD that is adequately supported by the data. If the POD is above some data points, it can fail to reflect the shape of the dose-response curve at the lowest doses and can introduce bias into subsequent extrapolations (see Figure 3-1). On the other hand, if the POD is far below all observed data points, it can introduce model uncertainty and parameter uncertainty (see Section 3.6) that increase with the distance between the data and the POD. Use of a POD at the lowest level supported by the data seeks to balance these considerations. It uses information from the model a small distance below the observed range rather than discarding this information and invoking default extrapolation procedures in a range where the model can provide some useful information. Statistical tests involving the ratio of the central estimate and its lower bound (i.e., ED_{xx}/LED_{xx}) can be useful for evaluating how well the data support model estimates at a particular response level. (Note that the ability to model at a particular response level is not the same as the study's ability to identify an increase at that response level as statistically significant.)

For applications that involve extrapolation, Agency practice has been to use a lower bound as the POD. This reflects the Agency's appraisal of the relative consequences of

overestimating or underestimating the POD. It also ensures that the POD considers the variance of the estimated dose, which can depend on a study's design, sample size, and quality. *For applications that do not involve extrapolation* (e.g., hazard rankings), Agency practice has been to use a central estimate as the POD. In either case, both the central estimate and its lower bound are presented to convey a sense of the uncertainty in the POD.

When tumor data are used, a POD is obtained from the modeled tumor incidences. Conventional cancer bioassays, with approximately 50 animals per group, generally can support modeling down to an increased incidence of 1–10%; epidemiologic studies, with larger sample sizes, below 1%. Various models commonly used for carcinogens yield similar estimates of the POD at response levels as low as 1% (Krewski and Van Ryzin, 1981; Gaylor et al., 1994). Consequently, response levels below 10% can often be used as the POD. As a modeling convention, the lower bound on the doses associated with standard response levels of 1, 5, and 10% can be analyzed, presented, and considered. For making comparisons at doses within the observed range, the ED_{10} and LED_{10} are also reported as a common POD that can be used, with appropriate adjustments, in hazard rankings that compare different agents or health effects (U.S. EPA, 2002c).

When precursor data are available and of good quality, models that include both tumors and their precursors are generally most useful for deriving a POD. Such models can provide insight into quantitative relationships between tumors and precursors (see Section 3.2.2), possibly suggesting the precursor response level that is associated with a particular tumor response level. The goal is to use precursor data to extend the observed range below what can be observed in tumor studies. If the precursor data are drawn from small samples or if the quantitative relationship between tumors and precursors is not well defined, then the tumor data may provide a more reliable POD. Precursor effects may or may not be biologically adverse in themselves; the intent is to consider not only tumors but also damage that can lead to subsequent tumor development by this or another agent. Special attention is needed when analyzing continuous precursor data; Murrell et al. (1998) discuss alternative approaches to deriving a POD from continuous data. A *no-observed-adverse-effect level* generally is not used for assessing the potential for carcinogenic response when a model can be fitted to the data.

3.2.5. Characterizing the POD: the POD Narrative

As a single-point summary of a single dose-response curve, the POD alone does not convey all the critical information present in the data from which it is derived. To convey a measure of uncertainty, the POD should be presented as a central estimate with upper and lower

confidence bounds. A POD narrative summarizes other important features of the database and the POD that are important to account for in low-dose extrapolations or other analyses.

(a) Nature of the response. Is the POD based on tumors or a precursor? If on tumors, does the POD measure incidence or mortality? Is it a lifetime measure or was the study terminated early? The relationships between precursors and tumors, incidence and mortality, and lifetime and early-termination results vary from case to case. Modeling can provide quantitative insight into these relationships, for example, linking a change in a precursor response to a tumor incidence (see Section 3.2.2). This can aid in evaluating the significance of the response at the POD and adjusting different PODs to make them comparable.

(b) Level of the response. What level of response is associated with the POD, for example, 1% cancer risk, 10% cancer risk, or 10% change in a precursor measure?

(c) Nature of the study population. Is the POD 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 lifestage?

(d) Slope of the dose-response curve at the POD. How does response change as dose is reduced below the POD? A steep slope indicates that risk decreases rapidly as dose decreases. On the other hand, a steep slope also indicates that errors in an exposure assessment can lead to large errors in estimating risk. Both aspects of the slope are important. The slope also indicates whether dose-response curves for different effects are likely to cross below the POD. For example, in the ED_{01} study where 2-acetylaminofluorene caused bladder carcinomas and liver carcinomas in mice (Littlefield et al., 1980), the dose-response curves for these tumors cross between 10% and 1% response (see Figure 3-2). This crossing, which can be inferred from the slopes of the curves at a 10% response, shows how considering the slope can lead to better inferences about the predominant effects expected at lower doses. Mode of action data can also be useful; quantitative information about key precursor events can be used to describe how risk decreases as dose decreases below the POD.

(e) Relationship of the POD with other cancers. How does the POD for this cancer relate to PODs for other cancers observed in the database? For example, a POD based on male workers would not reflect the implications of mammary tumors in female rats or mice.

(f) Extent of the overall cancer database. Have potential cancer responses been adequately studied (e.g., were all tissues examined), or is the database limited to particular effects, population segments, or lifestages? Do the mode of action data suggest a potential for

cancers not observed in the database (e.g., disruption of particular endocrine pathways leading to related cancers)?

3.2.6. Relative Potency Factors

Relative potency factors can be used for a well-defined class of agents that operate through a common mode of action. A complete dose-response assessment is conducted for one well-studied member of the class that serves as the *index chemical* for the class. The other members of the class are tied to the index chemical by relative potency factors that are based on characteristics such as relative toxicological outcomes, relative metabolic rates, relative absorption rates, quantitative SARs, or receptor binding characteristics (U.S. EPA, 2000c). Examples of this approach are the *toxicity equivalence factors* for dioxin-like compounds and the relative potency factors for some carcinogenic polycyclic aromatic hydrocarbons.

3.3. EXTRAPOLATION TO LOWER DOSES

The purpose of low-dose extrapolation is to provide as much information as possible about risk in the range of doses below the observed data. The most versatile form of low-dose extrapolation is a dose-response model that characterizes risk as a probability over a range of environmental exposure levels. These risk probabilities allow estimates of the risk reduction under different decision options and estimates of the risk remaining after an action is taken and provide the risk information needed for benefit/cost analyses of different decision options.

When a dose-response model is not developed for lower doses, another form of low-dose extrapolation is a safety assessment that characterizes the safety of one lower dose, with no explicit characterization of risks above or below that dose. Although this type of extrapolation may be adequate for evaluating different decision options, it may not be adequate for other purposes (e.g., benefit/cost analyses) that require a quantitative characterization of risks across a range of doses. At this time, safety assessment is the default approach for tumors that arise through a nonlinear mode of action; however, EPA continues to explore methods for quantifying dose-response relationships over a range of environmental exposure levels for tumors that arise through a nonlinear mode of action (U.S. EPA, 2002c). EPA program offices that need this more explicit dose-response information may develop and apply methods that are informed by the methods described in these guidelines.

3.3.1. Choosing an Extrapolation Approach

The approach for extrapolation below the observed data considers the understanding of the agent's mode of action at each site (see Section 2.5). Mode of action data can suggest the likely shape of the dose-response curve at lower doses. The extent of inter-individual variation is also considered, with greater variation spreading the response over a wider range of doses.

Linear extrapolation should be used when there are data to indicate that the dose-response curve has a linear component below the POD, as when

- the agent is DNA-reactive and has direct mutagenic activity or the agent operates through another mode of action that is expected to be linear at low doses, or

- human exposure or body burden is high and near doses associated with key precursor events in the carcinogenic process, so that background exposures to this and other agents operating through a common mode of action are in the increasing, approximately linear, portion of the dose-response curve.

Linear extrapolation can also be used as a default approach when the available data fall short of establishing the mode of action at a tumor site, because linear extrapolation generally is considered to be a health-protective approach for addressing uncertainty about the mode of action.

A *nonlinear approach* should be selected when there are sufficient data to ascertain the mode of action and conclude that it is not linear at low doses and the agent does not demonstrate mutagenic or other activity consistent with linearity at low doses. Special attention is needed when the data support a nonlinear mode of action but there is also a suggestion of mutagenicity (either the evidence of mutagenicity is weak, or the mutagenic effect is weak, or mutagenicity is expected only at high doses). Depending on the strength of the suggestion of mutagenicity, the assessment may justify a conclusion that mutagenicity is not operative at low doses and focus on a nonlinear approach, or alternatively, the assessment may use both linear and nonlinear approaches.

Both linear and nonlinear approaches may be used when there are multiple modes of action:

- If there are multiple tumor sites, one with a linear and another with a nonlinear mode of action, then the corresponding approach is used at each site.

- If there are multiple modes of action at a single tumor site, one linear and another nonlinear, then both approaches are used to decouple and consider the respective contributions of each mode of action in different dose ranges. For example, an agent can act predominantly through cytotoxicity at high doses and through mutagenicity at lower doses where cytotoxicity does not occur. Modeling to a low response level can be useful for estimating the response at doses where the high-dose mode of action would be less important.

Nonlinear approaches generally should not be used in cases where the mode of action has not been ascertained.

3.3.2. Extrapolation Using a Toxicodynamic Model

The better approach is to develop a toxicodynamic model of the agent's mode of action and use that model for extrapolation to lower doses (see Section 3.2.2). The extent of extrapolation is governed by an analysis of *model uncertainty*, where alternative models that fit similarly in the observed range can diverge below that range (see Section 3.6). Substantial divergence is likely when model parameters are estimated from tumor incidence data, so that different combinations of parameter estimates yield similar fits in the observed range but have different implications at lower doses. An analysis of model uncertainty can be used to determine the range where extrapolation using the toxicodynamic model is supported and where further extrapolation would be based on either a linear or a nonlinear default, as appropriate (see Sections 3.3.4, 3.3.3).

3.3.3. Nonlinear Extrapolation to Lower Doses

A nonlinear default can be used for cases with sufficient data to ascertain the mode of action and conclude that it is not linear at low doses but not enough data to support a toxicodynamic model at low doses. Currently, nonlinear default approaches do not estimate risk probabilities or provide a dose-response curve at low doses, because there is considerable model uncertainty (see Section 3.6) inherent in the extrapolation of nonlinear models: different nonlinear models that fit the observed data can lead to a wide range of results at lower doses, with no basis to choose among them. EPA is continuing to explore methods for quantifying dose-response relationships over a range of environmental exposure levels for tumors that arise through a nonlinear mode of action.

For cases where the tumors arise through a nonlinear mode of action, an oral *reference dose* or an inhalation *reference concentration*, or both, should be developed in accordance with EPA's established practice for developing such values, taking into consideration the factors summarized in the characterization of the POD (see Section 3.2.5). This approach expands the past focus of such reference values (previously reserved for effects other than cancer) to include carcinogenic effects determined to have a nonlinear mode of action. As with other health effects of concern, it is important to put cancer in perspective with the overall health impact of an exposure by comparing reference value calculations for cancer with those for other health effects.

For effects other than cancer, reference values have been described as being based on the assumption of biological thresholds. The Agency's more current guidelines for these effects, however, do not use this assumption, citing the difficulty of empirically distinguishing a true threshold from a dose-response curve that is nonlinear at low doses (U.S. EPA, 1998b,1996a).

Economic and policy analysts need to know how the probability of cancer varies at exposures above the reference dose and whether, and to what extent, there are health benefits from reducing exposures below the reference dose. The risk assessment community is working to develop better methods to provide more useful information to economic and policy analysts.

3.3.4. Extrapolation Using a Low-dose Linear Model

Linear extrapolation should be used in two distinct circumstances: (1) when there are data to indicate that the dose-response curve has a linear component below the POD, or (2) as a default for a tumor site where the mode of action is not established (see Section 3.3.1). For linear extrapolation, a line should be drawn from the POD to the origin, corrected for background. This implies a proportional (linear) relationship between risk and dose at low doses. (Note that the dose-response curve generally is not linear at higher doses.)

The slope of this line, known as the *slope factor*, is an upper-bound estimate of risk per increment of dose that can be used to estimate risk probabilities for different exposure levels. The slope factor is equal to $0.01/LED_{01}$ if the LED_{01} is used as the POD.

Unit risk estimates express the slope in terms of μg/L drinking water or $\mu g/m^3$ air. In general, the drinking water unit risk is derived by converting a slope factor from units of mg/kg-d to units of μg/L, whereas an inhalation unit risk is developed directly from a dose-response analysis using equivalent human concentrations already expressed in units of $\mu g/m^3$. Unit risk estimates often assume a standard intake rate (L/day drinking water or m^3/day air) and body weight (kg), which may need to be reconciled with the exposure factors for the population

of interest in an exposure assessment (see Section 4.4). Although unit risks have not been calculated in the past for dermal exposures, both exposures that are absorbed into the systemic circulation and those that remain in contact with the skin are also important.

Risk-specific doses are derived from the slope factor or unit risk to estimate the dose associated with a specific risk level, for example, a one-in-a-million increased lifetime risk.

3.3.5. Comparing and Combining Multiple Extrapolations

When multiple estimates can be developed, all datasets should be considered and a judgment made about how best to represent the human cancer risk. Some options for presenting results include

- adding risk estimates derived from different tumor sites (NRC, 1994),
- combining data from different datasets in a joint analysis (Stiteler et al., 1993; Vater et al., 1993),
- combining responses that operate through a common mode of action,
- representing the overall response in each experiment by counting animals with any tumor showing a statistically significant increase,
- presenting a range of results from multiple datasets (in this case, the dose-response assessment includes guidance on how to choose an appropriate value from the range),
- choosing a single dataset if it can be justified as most representative of the overall response in humans, or
- a combination of these options.

Cross-comparison of estimates from human and animal studies can provide a valuable risk perspective:

- Calculating an animal-derived slope factor and using it to estimate the risk expected in a human study can provide information about the human study design, for example, adequacy of exposure level and sample size.

- Calculating an upper-bound slope factor from a nonpositive human study with good exposure information and comparing it to an animal-derived slope factor can indicate whether the positive animal and nonpositive humans studies are consistent.

3.4. EXTRAPOLATION TO DIFFERENT HUMAN EXPOSURE SCENARIOS

Often, an assessment based on human or animal studies of long-term, constant exposure is applied to different human exposure scenarios, for example, less-than-lifetime durations or intermittent patterns of exposure. The dose-response assessment provides recommendations to exposure assessors who will evaluate such scenarios. In developing these recommendations, tumor studies involving less-than-lifetime dosing or follow-up are often not informative, as these studies can be limited by inadequate power or insufficient allowance for latency.

For lifetime human exposure scenarios that involve intermittent or varying levels of exposure, the prevailing practice has been to assess exposure by calculating a *lifetime average daily dose* (U.S. EPA, 1992a). This approach assumes that an intermittent exposure scenario is equivalent to constant lifetime exposure at the average level, which matches the dosing regimen used in conventional cancer bioassays.

For less-than-lifetime human exposure scenarios, too, the lifetime average daily dose has been used. This implies that less-than-lifetime exposure is associated with a proportional reduction of the lifetime risk, regardless of when exposures occur. The appeal of this default lies in its simplicity and its appearance of being risk neutral with regard to timing of exposure. It is not, however, compatible with current dose-response models of carcinogenesis: both the multistage model and the two-stage clonal expansion model predict that short-duration risks are not necessarily proportional to exposure duration and can depend on the nature of the carcinogen and the timing of exposure (Goddard et al., 1995). In some circumstances, use of a lifetime average daily dose would underestimate cancer risk by two- to fivefold (Murdoch et al., 1992). Consistent with these theoretical results, an empirical comparison of results from parallel chronic and stop-exposure studies conducted by NTP suggests that use of the lifetime average daily dose is more likely to lead to an underestimate than an overestimate of risk (Halmes et al., 2000). As methodological research focuses on new approaches for estimating risks from less-than-lifetime exposures, methods and defaults can be expected to change.

This highlights the importance for each dose-response assessment to critically evaluate all information pertaining to less-than-lifetime exposure. For example, detailed stop-exposure studies can provide information about the relationship between exposure duration, precursor effects, potential for reversibility, and tumor development. Toxicokinetic modeling can investigate differences in internal dose between short-term and long-term exposure or between intermittent and constant exposure. Persistence in the body can be useful in explaining long-term effects resulting from shorter-term exposures.

The use of lifetime average daily dose described above was adopted with low-dose linear cancer assessments in mind. *For nonlinear cancer analyses*, it is appropriate to assess exposure by calculating a *daily dose* that is not averaged over a lifetime (see Section 3.1.1). This reflects an expectation that the precursor effects on which the analysis is based can result from less-than-lifetime exposure, bringing consistency to the methods used for dose-response assessment and exposure assessment in such cases. The dose-response assessment provides a recommendation to exposure assessors about the averaging time that is appropriate to the mode of action.

3.5. EXTRAPOLATION TO SUSCEPTIBLE POPULATIONS AND LIFESTAGES

The dose-response assessment strives to derive separate estimates for susceptible populations and lifestages so that these risks can be explicitly characterized. For a susceptible population, higher risks can be expected from exposures anytime during life, but this applies to only a portion of the general population (e.g., those bearing a particular genetic susceptibility). In contrast, for a susceptible lifestage, higher risks can be expected from exposures during only a portion of a lifetime, but this applies to the entire population. Exposures during a susceptible period are not equivalent to exposures at other times; consequently, it is useful to estimate the risk attributable to exposures during each period.

Depending on the data available, a tiered approach should be used to address susceptible populations and lifestages.

1. When there is an epidemiologic study or an animal bioassay that reports quantitative results for susceptible individuals, the data should be analyzed to provide a separate risk estimate for those who are susceptible. If susceptibility pertains to a lifestage, it is useful to characterize the portion of the lifetime risk that can be attributed to the susceptible lifestage.

2. When there are data on some risk-related parameters that allow comparison of the general population and susceptible individuals, the data should be analyzed with an eye toward adjusting the general population estimate for susceptible individuals. This analysis can range from toxicokinetic modeling that uses parameter values representative of susceptible individuals to more simply adjusting a general population estimate to reflect differences in important rate-governing parameters. Care is taken to not make parameter adjustments in isolation, as the appropriate adjustment can depend on the interactions of several parameters; for example, the ratio of metabolic activation and clearance rates can be more appropriate than the activation rate alone (U.S. EPA, 1992b).

3. In the absence of such agent-specific data, there is some general information to indicate that childhood can be a susceptible lifestage for exposure to some carcinogens (EPA 2003); this warrants explicit consideration in each assessment. The potential for susceptibility from early-life exposure is expected to vary among specific agents and chemical classes. In addition, the concern that dose-averaging generally used for assessing less-than-lifetime exposure is more likely to understate than overstate risk (Halmes et al., 2000, see also Section 3.4) contributes to the suggestion that alternative approaches be considered for assessing risks from less-than-lifetime exposure that occurs during childhood. Accompanying these guidelines is supplemental guidance that the Agency will use to assess risks from early-life exposure to potential carcinogens (U.S. EPA, 2003). **[This draft Supplemental Guidance presents, at this time, a possible approach for assessing cancer susceptibility from early-life exposure to carcinogens. The final guidance will reflect public comment and recommendations from the Science Advisory Board's review of the Supplemental Guidance.]** The supplemental guidance may be updated to reflect new data and new understanding that may become available in the future.

3.6. UNCERTAINTY

The NRC (1983, 1994, 1996, 2002) has repeatedly advised that proper characterization of uncertainty is essential in risk assessment. An assessment that omits or underestimates uncertainty can leave decision-makers with a false sense of confidence in estimates of risk. On the other hand, a high level of uncertainty does not imply that a risk assessment or a risk

management action should be delayed (NRC, 2002). Uncertainty in dose-response assessment can be classified as either *model uncertainty* or *parameter uncertainty*. A related concept is *human variation*, discussed below. Assessments should discuss the significant uncertainties encountered in the analysis, distinguishing, if possible, between model uncertainty, parameter uncertainty, and human variation.

Model uncertainty refers to a lack of knowledge needed to determine which scientific theory a model is based upon is correct. 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 characterization can be obtained by comparing the results of alternative models. Model uncertainty is expressed through comparison of separate analyses from each model, coupled with a subjective probability statement, where feasible and appropriate, of the likelihood that each model might be correct (NRC, 1994).

Some aspects of model uncertainty that should be addressed in an assessment include the use of animal models as a surrogate for humans, the influence of cross-species differences in metabolism and physiology, the use of effects observed at high doses as an indicator of the potential for effects at lower doses, the effect of using linear or nonlinear extrapolation to estimate risks, the use of using small samples and subgroups to make inferences about entire human populations or subpopulations with differential susceptibilities, and the use of experimental exposure regimens to make inferences about different human exposure scenarios (NRC, 2002).

Toxicokinetic and toxicodynamic models are generally premised on *site concordance* across species, modeling, for example, the relationship between administered dose and liver tissue concentrations to predict increased incidences of liver cancer. This relationship, which can be observed in animals, is typically only inferred for humans. There are, however, numerous examples of an agent causing different cancers in different species. The assessment should discuss the relevant data that bear on this form of model uncertainty.

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.

Human variation refers to person-to-person differences in biological susceptibility or in exposure. Although both human variation and uncertainty can be characterized 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 characterized, 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.

Probabilistic risk assessment, informed by expert judgment, 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 application 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 (see U.S. 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. With appropriate data and expert judgment, formal approaches to probabilistic risk assessment can be applied to provide insight into the overall extent and dominant sources of human variation and uncertainty. In doing this, 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. Specification of joint probability distributions is appropriate when variables are not independent of each other. In each case, the assessment should carefully consider the questions of uncertainty and human variation and discuss the extent to which there are data to address them.

Probabilistic risk assessment has been used in dose-response assessment to determine and distinguish the degree of uncertainty and variability in toxicokinetic and toxicodynamic modeling. Although this field is less advanced that probabilistic exposure assessment, progress is being made and these guidelines are flexible enough to accommodate continuing advances in these approaches.

3.7. DOSE-RESPONSE CHARACTERIZATION

A *dose-response characterization* extracts the dose-response information needed in a full risk characterization (U.S. EPA, 2000b), including

- presentation of the recommended estimates (slope factors, reference doses, reference concentrations),
- a summary of the data supporting these estimates,
- a summary of the modeling approaches used,
- the POD narrative (see Section 3.2.5),
- a summary of the key defaults invoked,
- identification of susceptible populations or lifestages and quantification of their differential susceptibility, and
- a discussion of the strengths and limitations of the dose-response assessment, highlighting significant issues in developing risk estimates, alternative approaches considered equally plausible, and how these issues were resolved.

All estimates should be accompanied by the weight of evidence descriptor (see Section 2.6) to convey a sense of the qualitative uncertainty about whether the agent may or may not be carcinogenic.

Slope factors generally represent an upper bound on the average risk in a population or the risk for a randomly selected individual but not the risk for a highly susceptible individual or group. Some individuals face a higher risk and some face a lower risk. The use of upper bounds generally is considered to be a health-protective approach for covering the risk to susceptible individuals, although the calculation of upper bounds is not based on susceptibility data. Similarly, exposure during some lifestages can contribute more or less to the total lifetime risk than do similar exposures at other times. The dose-response assessment characterizes, to the extent possible, the extent of these variations.

Depending on the supporting data and modeling approach, a slope factor can have a mix of traits that tend to either estimate, overestimate, or underestimate risk.

Some examples of traits that tend to overestimate risk include:

- The slope factor is derived from data on a highly susceptible animal strain.
- Linear extrapolation is used as a default and extends over several orders of magnitude.
- The largest of several slope factors is chosen.

Some examples of traits that tend to underestimate risk include:

- Several tumor types were observed, but the slope factor is based on a subset of them.
- The study design does not include exposure during a susceptible lifestage, for example, perinatal exposure.
- The study population is of less-than-average susceptibility, for example, healthy adult workers.
- There is random exposure misclassification or random exposure measurement error in the study from which the slope factor is derived.

Some examples of traits that neither overestimate nor underestimate risk include:

- The slope factor is derived from data in humans or in an animal strain that responds like humans.
- Linear extrapolation is appropriate for the agent's mode of action.
- Environmental exposures are close to the observed data.

- Several slope factors for the same tumor are averaged or a slope factor is derived from pooled data from several studies.

- The slope factor is derived from the only suitable study.

The dose-response characterization weighs these traits and discusses the degree to which the slope factor, on balance, would tend to yield an overestimate, underestimate, or central estimate of risk. If the overall tendency appears to underestimate risk, then the assessment may adjust the slope factor so that risk is not likely to be underestimated.

Figure 3-1. Compatability of alternative points of departure with observed and modeled tumor incidences

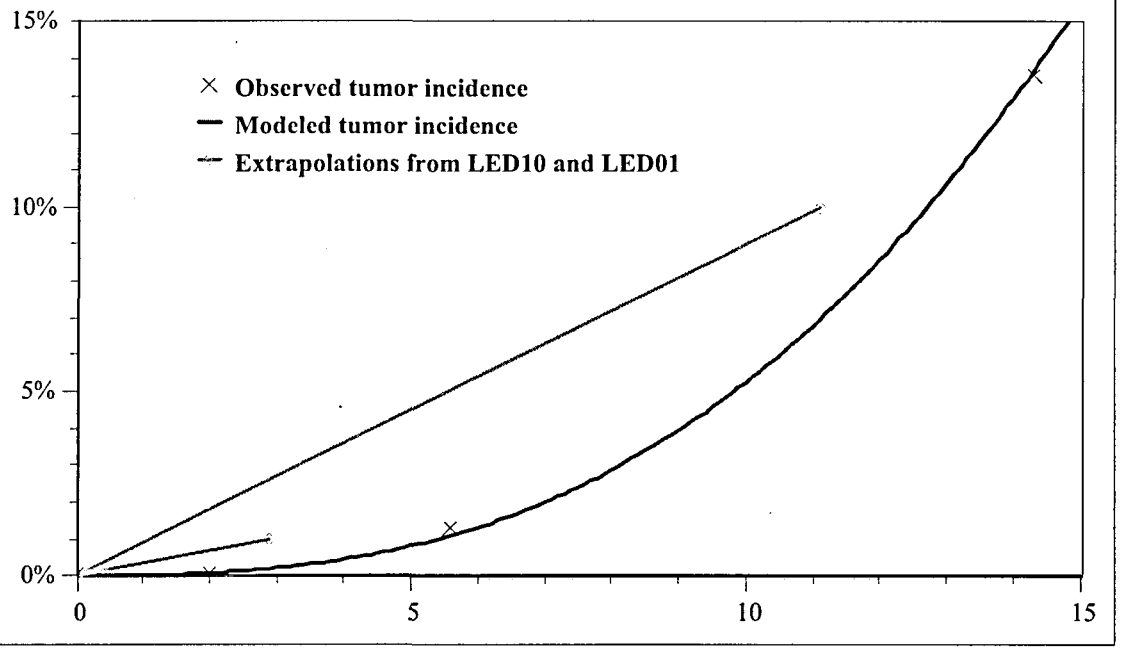
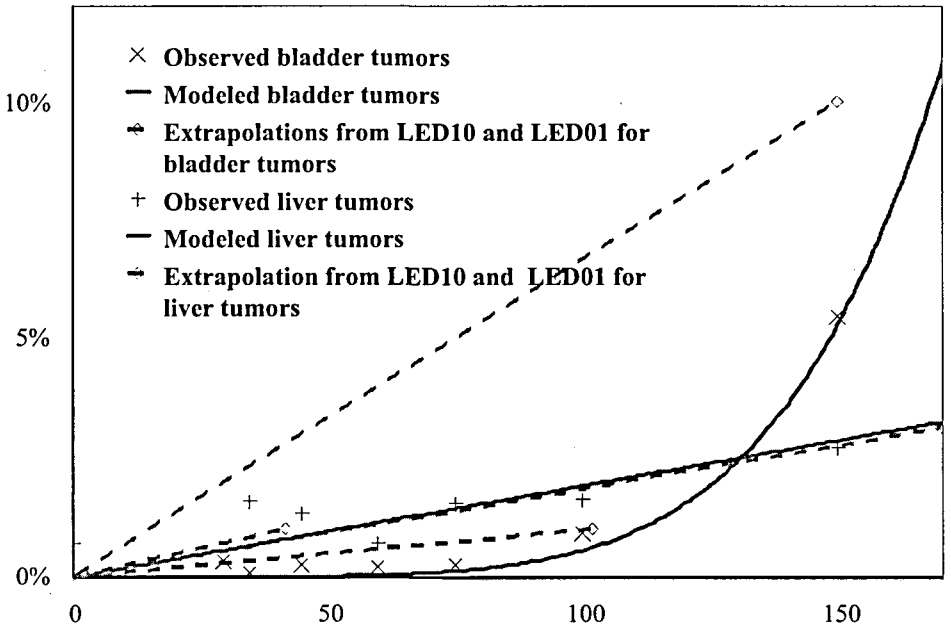


Figure 3-2. Crossing--between 10% and 1%--of dose-response curves for bladder carcinomas and liver carcinomas induced by 2-AAF



4. EXPOSURE ASSESSMENT

Exposure assessment is the determination (qualitative and quantitative) of the magnitude, frequency, and duration of exposure (U.S. EPA, 1992a). This section provides a brief overview of exposure assessment principles, with an emphasis on issues related to carcinogenic risk assessment. The information presented here should be used in conjunction with other guidance documents, including *Guidelines for Exposure Assessment* (U.S. EPA, 1992a), *Policy and Guidance for Risk Characterization* (U.S. EPA, 2000b), *Exposure Factors Handbook* (U.S. EPA, 1997c), the 1997 *Policy for Use of Probabilistic Analysis in Risk Assessments* (U.S. EPA, 1997d), and the 1997 *Guiding Principles for Monte Carlo Analysis* (U.S. EPA, 1997e). In addition, program-specific guidelines for exposure assessment should be consulted.

Exposure assessment generally consists of four major steps: defining the assessment questions, selecting or developing the conceptual and mathematical models, collecting data or selecting and evaluating available data, and exposure characterization. Each of these steps is briefly described below.

4.1. DEFINING THE ASSESSMENT QUESTIONS

In providing a clear and unambiguous statement of the purpose and scope of the exposure assessment (U.S. EPA, 1997e), consider the following:

- The management objectives of the assessment will determine whether deterministic screening level analyses are adequate or whether full probabilistic exposure characterization is needed.

- Identify and include all important sources (e.g., pesticide applications), pathways (e.g., food or water), and routes (e.g., ingestion, inhalation, and dermal) of exposure in the assessment. If a particular source, pathway, or route is omitted, a clear and transparent explanation should be provided.

- Separate analyses should be conducted for each definable subgroup within the population of interest. In particular, subgroups that are believed to be highly exposed or susceptible to a particular health effect should be studied. These include people with certain diseases or genetic susceptibilities and others whose behavior or

physiology may lead to higher exposure or susceptibility. Consider the following examples:

— Physiological differences between men and women (e.g., body weight and inhalation rate) may lead to important differences in exposures. See, for example, the discussion in *Exposure Factors Handbook*, Appendix 1A (U.S. EPA, 1997c).

— Pregnant and lactating women may have exposures that differ from the general population (e.g., slightly higher water consumption) (U.S. EPA, 1997c). Further, exposure to pregnant women may result in exposure to the developing fetus (NRC, 1993b).

— Children consume more food per body weight than do adults while consuming fewer types of foods (ILSI, 1992; NRC, 1993b; U.S. EPA, 1997c). In addition, children engage in crawling and mouthing (i.e., putting hands and objects in the mouth) behaviors, which can increase their exposures.

— The elderly and disabled may have important differences in their exposures due to a more sedentary lifestyle (U.S. EPA, 1997c). In addition, the health status of this group may affect their susceptibility to the detrimental effects of exposure.

For further guidance, see *Guidelines for Exposure Assessment*, § 3 (U.S. EPA, 1992a).

4.2. SELECTING OR DEVELOPING THE CONCEPTUAL AND MATHEMATICAL MODELS

Carcinogen risk assessment models are generally based on the premise that risk is proportional to total lifetime dose. For lifetime human exposure scenarios, therefore, the exposure metric used for carcinogenic risk assessment is the lifetime average daily dose (LADD). The LADD is typically used in conjunction with the slope factor to calculate individual excess cancer risk. It is an estimate of the daily intake of a carcinogenic agent throughout the entire life of an individual. Depending on the objectives of the assessment, the LADD may be calculated deterministically (using point estimates for each factor to derive a

point estimate of the exposure) or stochastically (using probability distributions to represent each factor and such techniques as Monte Carlo analysis to derive a distribution of the LADD) (U.S. EPA, 1997e). Stochastic analyses may help to identify certain population segments that are highly exposed and may need to be assessed as a special subgroup. For further guidance, see *Guidelines for Exposure Assessment*, § 5.3.5.2 (U.S. EPA, 1992a).

When the route of exposure is inhalation or dermal contact, derivation of the LADD often needs an approach to "route-to-route extrapolation." The slope factor and other measures of toxicity are typically derived from oral administered doses in animal studies. Therefore, for ingestion exposures in a human population it is not usually necessary to make adjustments to account for route-specific differences in absorption and uptake. However, for inhalation and dermal exposures, such adjustments may be necessary. For further guidance, see *Guidelines for Exposure Assessment*, § 2.1.4 (U.S. EPA, 1992a).

For less-than-lifetime human exposure scenarios, use of an LADD is more likely to lead to an underestimate than an overestimate of risk (see Section 3.4). As methodological research focuses on new approaches for estimating risks from less-than-lifetime exposures, methods and defaults can be expected to change.

There may be cases where the mode of action indicates that dose rates are important in the carcinogenic process. In these cases, short-term, less-than-lifetime exposure estimates may be more appropriate than the LADD for risk assessment. This is typically the case when a nonlinear dose-response approach is used (see Section 3.4).

4.3. COLLECTING DATA OR SELECTING AND EVALUATING AVAILABLE DATA

After the assessment questions have been defined and the conceptual and mathematical models have been developed, it is important to compile and evaluate existing data or, if necessary, to collect new data. Depending on the exposure scenario under consideration, data on a wide variety of exposure factors may be needed. EPA's *Exposure Factors Handbook* (U.S. EPA, 1997c) contains a large compilation of exposure data, with some analysis and recommendations. Some of these data are organized by age groups to assist with assessing such subgroups as children. See, for example, *Exposure Factors Handbook*, Volume 1, Chapter 3 (U.S. EPA, 1997c). When using these existing data, it is important to evaluate the quality of the data and the extent to which the data are representative of the population under consideration. EPA's *Guidance for Data Quality Assessment* (U.S. EPA, 2000d) and program-specific guidances can provide further assistance for evaluating existing data.

When existing data fail to provide an adequate surrogate for the needs of a particular assessment, it is important to collect new data. Such data collection efforts should be guided by the references listed above (e.g., *Guidance for Data Quality Assessment* and program-specific guidance). Once again, subgroups of concern are an important consideration in any data collection effort.

4.3.1. Adjusting Unit Risks for Highly Exposed Populations and Lifestages

Unit risk estimates developed in the dose-response assessment often assume standard adult intake rates. When an exposure assessment focuses on a population with higher exposure, good exposure assessment practice would replace the standard intake rates with values representative of the exposed population.

> *For example*, to adjust the drinking water unit risk for an active population that drinks 4 L/day (instead of 2 L/day), multiply the unit risk by 2.

Because children eat more food, drink more water, and breathe more air relative to body weight than do adults (U.S. EPA, 2002d), adjustments to unit risk estimates are warranted whenever they are applied in an assessment of childhood exposure.

> *For example*, to adjust the drinking water unit risk for a 9-kg infant who drinks 1 L/day (instead of a 70-kg adult who drinks 2 L/day), multiply the unit risk by [(1 L/day) / (9 kg)] / [(2 L/day) / (70 kg)] = 3.9.

Air unit risks are typically expressed in terms of air concentrations rather than daily intakes (U.S. EPA, 1994). Children and adults breathing the same concentration of an agent (such as a reactive gas) may receive different doses to the body or lungs (U.S. EPA, 2002b). For effects other than cancer, reference concentrations derived from a human-equivalent concentration are made to cover the general population by applying uncertainty factors for human variation or an incomplete database (U.S. EPA, 1994). For cancer unit risks, the human-equivalent concentration has been used without applying uncertainty factors, necessitating another approach to ensure that other groups are represented. One such approach would be to

calculate separate human-equivalent concentrations for children and adults in the dose-response assessment (see Section 3.1.3), leading to separate unit risks for children and adults.

However, if only adult-based values were presented in the dose-response assessment, a comparison of breathing rates relative to respiratory tract dimensions in children and adults can be undertaken (U.S. EPA, 2002b, 1994) in the exposure assessment to decide whether risk values for children can be improved by (1) substituting child-specific parameter values in the dosimetry model, (2) applying an adjustment or uncertainty factor, or (3) determining whether a verbal characterization of exposure differences between children and adults will suffice.

Any adjustments are made consistent with the dosimetry model that was used. In these models, dose is generally proportional to air intake rate adjusted for surface area (for respiratory tract effects) or body weight (for effects elsewhere in the body). The dosimetry model that was used can be adapted for healthy children by substituting the child's air intake rate, surface area, and body weight in place of the adult default (U.S. EPA, 1994).

An exception occurs for gases that are not reactive and not water soluble. In this case, dose is proportional to the blood:gas partition coefficient, independent of intake rate and surface area, assuming equilibrium between ambient air, blood, and body compartments (U.S. EPA, 1994). Consequently, it is important to determine whether children reach equilibrium as quickly as do adults. Under nonequilibrium conditions, dose can depend on intake rate and body size.

The dose-response assessment discusses the key sources of uncertainty in estimating children's doses, including use of dosimetry models that are based on the dimensions of the adult respiratory tract. This is particularly crucial for particle dosimetry, because a different distribution of particle sizes would be expected in a child's smaller respiratory tract. Children's dose is also affected by other anatomical and metabolic differences between children and adults.

4.4. EXPOSURE CHARACTERIZATION

The exposure characterization is a technical characterization that presents the assessment results and supports the risk characterization. It provides a statement of the purpose, scope, and approach used in the assessment, identifying the exposure scenarios and population subgroups covered. It provides estimates of the magnitude, frequency, duration, and distribution of exposures among members of the exposed population as the data permit. It identifies and compares the contribution of different sources, pathways, and routes of exposure. In particular, a qualitative discussion of the strengths and limitations (uncertainties) of the data and models are presented.

The discussion of uncertainties is a critical component of the exposure characterization. Uncertainties can arise out of problems with the conceptual and mathematical models. Uncertainties can also arise from poor data quality and data that are not quite representative of the population or scenario of interest. Consider the following examples of uncertainties.

- National data (i.e., data collected to represent the entire U.S. population) may not be representative of exposures occurring within a regional or local population.

- Use of short-term data to infer chronic, lifetime exposures should be done with caution. Use of short-term data to estimate long-term exposures has the tendency to underestimate the number of people exposed while overestimating the exposure levels experienced by those in the upper end (i.e., above the 90^{th} percentile) of the exposure distribution. For further guidance, refer to *Guidelines for Exposure Assessment*, § 5.3.1 (U.S. EPA, 1992a).

- Children's behavior may lead to relatively high but intermittent exposures. This pattern of exposure, "one that gradually declines over the developmental period and which remains relatively constant thereafter" is not accounted for in the LADD model (ILSI, 1992). Further, the physiological characteristics of children may lead to important differences in exposure. Some of these differences can be accounted for in the LADD model. For further guidance, see *Guidelines for Exposure Assessment*, § 5.3.5.2 (U.S. EPA, 1992a).

Overall, the exposure characterization should provide a full description of the sources, pathways, and routes of exposure. The characterization also should include a full description of the populations assessed. In particular, highly exposed or susceptible subgroups should be discussed. For further guidance on the exposure characterization, consult *Guidelines for Exposure Assessment* (U.S. EPA, 1992a), the *Policy and Guidance for Risk Characterization* (U.S. EPA, 2000b,1995) and EPA's *Rule Writer's Guide to Executive Order 13045* (especially Attachment C: Technical Support for Risk Assessors—Suggestions for Characterizing Risks to Children) (U.S. EPA, 1998d).

5. RISK CHARACTERIZATION

5.1. PURPOSE

EPA has developed general guidance on risk characterization for use in its risk assessment activities. The core of EPA's risk characterization policy (U.S. EPA, 2000b, 1995) includes the following:

> Each risk assessment prepared in support of decision making at EPA should include a risk characterization that follows the principles and reflects the values outlined in this policy. A risk characterization should be prepared in a manner that is clear, transparent, reasonable, and consistent with other risk characterizations of similar scope prepared across programs in the Agency. Further, discussion of risk in all EPA reports, presentations, decision packages, and other documents should be substantively consistent with the risk characterization. The nature of the risk characterization will depend upon the information available, the regulatory application of the risk information, and the resources (including time) available. In all cases, however, the assessment should identify and discuss all the major issues associated with determining the nature and extent of the risk and provide commentary on any constraints limiting fuller exposition.

Risk characterization should be carried out in accordance with the EPA and OMB information quality guidelines. EPA's Risk Characterization Handbook (U.S. EPA, 2000b) provides detailed guidance to Agency staff. The discussion below does not attempt to duplicate this material, but it summarizes its applicability to carcinogen risk assessment.

The risk characterization process includes an integrative analysis of the major results of the risk assessment that is summarized for the risk manager in a nontechnical discussion that minimizes the use of technical terms. It is an appraisal of the science that informs the risk manager in public health decisions, as do other decision-making analyses of economic, social, or technology issues. It also serves the needs of other interested readers. The summary is an information resource for preparing risk communication information, but being somewhat technical, is not itself the usual vehicle for communication with every audience.

The integrative analysis brings together the assessments of hazard, dose response, and exposure to make risk estimates for the exposure scenarios of interest. This analysis is generally

much more extensive than the risk characterization summary. It may be peer reviewed or subject to public comment along with the summary in preparation for an Agency decision. The integrative analysis may be titled differently by different EPA programs (e.g., "Staff Paper" for criteria air pollutants), but it typically will identify exposure scenarios of interest in decision making and present risk analyses associated with them. Some of the analyses may concern scenarios in several media; others may examine, for example, only drinking water risks. The integrative analysis also may be the document that contains quantitative analyses of uncertainty.

The values supported by a risk characterization throughout the process are *transparency* in environmental decision making, *clarity* in communication, *consistency* in core assumptions and science policies from case to case, and *reasonableness*. While it is appropriate to err on the side of protection of health and the environment in the face of scientific uncertainty, common sense and reasonable application of assumptions and policies are essential to avoid unrealistic estimates of risk (U.S. EPA, 2000b, 1995). Both integrative analyses and the risk characterization summary present an integrated and balanced picture of the analysis of the hazard, dose-response, and exposure. The risk analyst should provide summaries of the evidence and results and describe the quality of available data and the degree of confidence to be placed in the risk estimates. Important features include the constraints of available data and the state of knowledge, significant scientific issues, and significant science and science policy choices that were made when alternative interpretations of data exist (U.S. EPA, 2000b, 1995). Choices made about using default options or data in the assessment are explicitly discussed in the course of analysis, and if a choice is a significant issue, it is highlighted in the summary.

5.2. APPLICATION

Risk characterization is a necessary part of generating any Agency report on risk, whether the report is preliminary—to support allocation of resources toward further study—or comprehensive—to support regulatory decisions. In the former case, the detail and sophistication of the characterization are appropriately small in scale; in the latter case, appropriately extensive. Even if a document covers only parts of a risk assessment (hazard and dose-response analyses, for instance), the results of these are characterized.

Risk assessment is an iterative process that grows in depth and scope in stages from screening for priority making to preliminary estimation to fuller examination in support of complex regulatory decision making. Default options are typically used at every stage because no database is ever complete, but they are predominant at screening stages and are used less as more data are gathered and incorporated at later stages. Various provisions in EPA-administered

statutes require decisions based on differing findings for which differing degrees of analysis are appropriate. There are close to 30 provisions within the major statutes that require decisions based on risk, hazard, or exposure assessment. For example, Agency review of pre-manufacture notices under Section 5 of the Toxic Substances Control Act relies on screening analyses, whereas requirements for industry testing under Section 4 of that Act rely on preliminary analyses of risk or simply of exposure. In comparison, air quality criteria under the Clean Air Act rest on a rich data collection and required by statute to undergo periodic reassessment. There are provisions that require ranking of hazards of numerous pollutants—which may be addressed through a screening level of analysis—and other provisions for which a full assessment of risk is more appropriate.

Given this range in the scope and depth of analyses, not all risk characterizations can or should be equal in coverage or depth. The risk assessor should carefully decide which issues in a particular assessment are important to present, choosing those that are noteworthy in their impact on results. For example, health effect assessments typically rely on animal data because human data are rarely available. The objective of characterization of the use of animal data is not to recount generic issues about interpreting and using animal data; Agency guidance documents cover these issues. Rather, the objective is to call out any significant issues that arose within the particular assessment being characterized and inform the reader about significant uncertainties that affect conclusions.

5.3. PRESENTATION OF THE RISK CHARACTERIZATION SUMMARY

The presentation is a nontechnical discussion of important conclusions, issues, and uncertainties that uses the hazard, dose response, exposure, and integrative analyses for technical support. The primary technical supports within the risk assessment are the hazard characterization, dose-response characterization, and exposure characterization described in this guideline. The risk characterization is derived from these. The presentation should fulfill the aims outlined in the purpose section above.

5.4. CONTENT OF THE RISK CHARACTERIZATION SUMMARY

Specific guidance on hazard, dose-response, and exposure characterization appears in previous sections. Overall, the risk characterization routinely includes the following, capturing the important items covered in hazard, dose response, and exposure characterization:

- Primary conclusions about hazard, dose response, and exposure, including equally plausible alternatives.

- Nature of key supporting information and analytic methods.

- Risk estimates and their attendant uncertainties, including key uses of default options when data are missing or uncertain.

  — With linear extrapolations, risk is typically approximated by multiplying the slope factor by an estimate of exposure [Risk = Slope factor x Exposure]. For exposure levels above the POD, the dose-response model is used instead of this approximation.

  — With nonlinear extrapolations, hazard can be expressed as a *hazard quotient* (HQ), defined as the ratio of an exposure estimate over the reference dose (RfD) (HQ = Exposure / RfD). From the hazard quotient, it can generally be inferred whether the nonlinear mode of action is relevant at the environmental exposure level in question.

- Statement of the extent of extrapolation of risk estimates from observed data to exposure levels of interest and its implications for certainty or uncertainty in quantifying risk. The extent of extrapolation can be expressed as a *margin of exposure* (MOE), defined as the ratio of the POD over an exposure estimate (MOE = POD / Exposure).

- Significant strengths and limitations of the data and analyses, including any major peer review issues.

- Appropriate comparison with similar EPA risk analyses or common risks with which people may be familiar.

- Comparison with assessment of the same problem by another organization.

When a cancer risk assessment is prepared in a context where the results are likely to be used by Agency economists and policy analysts, it is important that the resulting characterizations include expected estimates of risk, as stipulated in OMB and EPA guidelines for benefit-cost analysis. Statutory mandates, such as the Safe Drinking Water Act, the Food Quality Protection Act, and the Clean Air Act, call for the Agency to generate specific kinds of risk information and thus these updated cancer assessment guidelines should be read in conjunction with the Agency's statutory mandates regarding risk assessment.

APPENDIX: MAJOR DEFAULT OPTIONS

This discussion covers the major default options commonly employed in a cancer risk assessment and adopted in these guidelines. These options are predominantly inferences that are needed to use the data observed under empirical conditions in order to estimate events and outcomes under environmental conditions. Several inferential issues arise when effects seen in a subpopulation of humans or animals are used to infer potential effects in the population of environmentally exposed humans. Several more inferential issues arise in extrapolating the exposure-effect relationship observed empirically to lower-exposure environmental conditions. The following issues cover the major default areas. Typically, an issue has some sub-issues; they are introduced here but are discussed in greater detail in later sections.

- Is the presence or absence of effects observed in a human population predictive of effects in another exposed human population?

- Is the presence or absence of effects observed in an animal population predictive of effects in exposed humans?

- How do metabolic pathways relate across species and among different age groups and between sexes in humans?

- How do toxicokinetic processes relate across species and among different age groups and between sexes in humans?

- What is the correlation of the observed dose-response relationship to the relationship at lower doses?

Is the Presence or Absence of Effects Observed in a Human Population Predictive of Effects in Another Exposed Human Population?

When cancer effects in exposed humans are attributed to exposure to an exogenous agent, the default option is that the resulting data are predictive of cancer in any other exposed human population. Studies either attributing cancer effects in humans to exogenous agents or reporting no effects are often studies of occupationally exposed humans. By sex, age, and general health, workers are not representative of the general population exposed environmentally

to the same agents. In such studies there is no opportunity to observe subpopulations who are likely to be under represented, such as fetuses, infants and children, women, or people in poor health, who may respond differently from healthy workers. Therefore, it is understood that this option could still underestimate the response of certain human subpopulations. (NRC, 1993b, 1994).

There is not yet enough knowledge to form a basis for any generally applicable qualitative or quantitative inference to compensate for this knowledge gap. In these guidelines, this problem is left to analysis in individual cases, to be attended to with further general guidance as future research and information allow. When information on a susceptible subpopulation or lifestage exists, it will be used. For example, an agent such as diethylstilbestrol (DES) causes a rare form of vaginal cancer (clear-cell adenocarcinoma) (Herbst, 1971) in about 1 per 1000 of adult women whose mothers were exposed during pregnancy (Hatch et al., 1998). *When cancer effects are not found in an exposed human population, this information by itself is not generally sufficient to conclude that the agent poses no carcinogenic hazard to this or other populations of potentially exposed humans, including susceptible subpopulations or lifestages.* This is because epidemiologic studies usually have low power to detect and attribute responses and typically evaluate cancer potential in a restricted population (e.g., by age, occupation). The topic of susceptibility and variation is addressed further in the discussion below of quantitative default options about dose-response relationships.

Is the Presence or Absence of Effects Observed in an Animal Population Predictive of Effects in Exposed Humans?

The default option is that positive effects in animal cancer studies indicate that the agent under study can have carcinogenic potential in humans. Thus, if no adequate human data are present, positive effects in animal cancer studies are a basis for assessing the carcinogenic hazard to humans. This option is a public health-conservative policy, and it is both appropriate and necessary given that we do not test for carcinogenicity in humans. The option is supported by the fact that nearly all of the agents known to cause cancer in humans are carcinogenic in animals in tests that have adequate protocols (IARC, 1994; Tomatis et al., 1989; Huff, 1994). Moreover, almost one-third of human carcinogens were identified subsequent to animal testing (Huff, 1993). Further support is provided by research on the molecular biology of cancer processes, which has shown that the mechanisms of control of cell growth and differentiation are remarkably homologous among species and highly conserved in evolution. Nevertheless, the same research tools that have enabled recognition of the nature and commonality of cancer

processes at the molecular level also have the power to reveal differences and instances in which animal responses are not relevant to humans (Lijinsky, 1993; U.S. EPA, 1991b). Under these guidelines, available mode of action information is studied for its implications in both hazard and dose-response assessment and its effect on default options.

There may be instances in which the use of an animal model would identify a hazard in animals that is not truly a hazard in humans (e.g., the alpha-2u-globulin association with renal neoplasia in male rats [U.S. EPA, 1991b]). The extent to which animal studies may yield false positive indications for humans is a matter of scientific debate. To demonstrate that a response in animals is not relevant to any human situation, adequate data to assess the relevancy issue must be available.

The default option is that effects seen at the highest dose tested are appropriate for assessment, but it is necessary that the experimental conditions be scrutinized. Animal studies are conducted at high doses in order to provide statistical power, the highest dose being one that is minimally toxic (maximum tolerated dose). Consequently, the question often arises of whether a carcinogenic effect at the highest dose may be a consequence of cell killing with compensatory cell replication or of general physiological disruption rather than inherent carcinogenicity of the tested agent. There is little doubt that this may happen in some cases, but skepticism exists among some scientists that it is a pervasive problem (Ames and Gold, 1990; Melnick et al., 1993; Barrett, 1993). If adequate data demonstrate that the effects are solely the result of excessive toxicity rather than carcinogenicity of the tested agent per se, then the effects may be regarded as not appropriate to include in assessment of the potential for human carcinogenicity of the agent. This is a matter of expert judgment, with consideration given to all of the data available about the agent, including effects in other toxicity studies, structure-activity relationships, and effects on growth control and differentiation.

When cancer effects are not found in well-conducted animal cancer studies in two or more appropriate species and other information does not support the carcinogenic potential of the agent, these data provide a basis for concluding that the agent is not likely to possess human carcinogenic potential, in the absence of human data to the contrary. This default option about lack of cancer effects has limitations. It is recognized that animal studies (and epidemiologic studies as well) have very low power to detect cancer effects. Detection of a 10% tumor incidence is generally the limit of power with standard protocols for animal studies (with the exception of rare tumors that are virtually markers for a particular agent, e.g., angiosarcoma caused by vinyl chloride). In some situations, the tested animal species may not be predictive of effects in humans; for example, arsenic shows only minimal or no effect in animals, whereas it is

clearly positive in humans. Therefore, it is important to consider other information as well; absence of mutagenic activity or absence of carcinogenic activity among structural analogues can increase the confidence that negative results in animal studies indicate a lack of human hazard.

Another limitation is that standard animal study protocols are not yet available for effectively studying perinatal effects. The potential for effects on the very young generally should be considered separately. Under existing Agency policy (U.S. EPA, 1997a, b), perinatal studies accomplished by modification of existing adult bioassay protocols are required in special circumstances.

The default option is that target organ concordance is not a prerequisite for evaluating the implications of animal study results for humans. Target organs of carcinogenesis for agents that cause cancer in both animals and humans are most often concordant at one or more sites (Tomatis et al., 1989; Huff, 1994). However, concordance by site is not uniform. The mechanisms of control of cell growth and differentiation are concordant among species, but there are marked differences among species in the way control is managed in various tissues. For example, in humans, mutations of the tumor suppressor genes p53 and retinoblastoma are frequently observed genetic changes in tumors. These tumor-suppressor genes are also observed to be operating in some rodent tissues, but other growth control mechanisms predominate in other rodent tissues. Thus, an animal response may be due to changes in a control that are relevant to humans but appear in animals in a different way.

However, it is appropriate under these guidelines to consider the influences of route of exposure, metabolism, and, particularly, some modes of action that may either support or not support target organ concordance between animals and humans. When data allow, these influences are considered in deciding whether the default remains appropriate in individual instances (NRC, 1994). Another exception to the basic default of not assuming site concordance exists in the context of toxicokinetic modeling. Site concordance is inherently assumed when these models are used to estimate delivered dose in humans on the basis of animal data.

The default is to include benign tumors observed in animal studies in the assessment of animal tumor incidence if such tumors have the capacity to progress to the malignancies with which they are associated. This default is consistent with the approach of the National Toxicology Program and the International Agency for Research on Cancer and is somewhat more protective of public health than not including benign tumors in the assessment; benign and malignant tumors are treated as representative of related responses to the test agent (McConnell et al., 1986), which is scientifically appropriate. Nonetheless, in assessing findings from animal

studies, a greater proportion of malignancy is weighed more heavily than is a response with a greater proportion of benign tumors. Greater frequency of malignancy of a particular tumor type in comparison with other tumor responses observed in an animal study is also a factor to be considered in selecting the response to be used in dose-response assessment.

Benign tumors that are not observed to progress to malignancy are assessed on a case-by-case basis. There is a range of possibilities for the overall significance of benign tumors. They may deserve attention because they are serious health problems even though they are not malignant; for instance, benign tumors may be a health risk because of their effect on the function of a target tissue, such as the brain. They may be significant indicators of the need for further testing of an agent if they are observed in a short-term test protocol, or such an observation may add to the overall weight of evidence if the same agent causes malignancies in a long-term study. Knowledge of the mode of action associated with a benign tumor response may aid in the interpretation of other tumor responses associated with the same agent.

How Do Metabolic Pathways Relate Across Species and Among Different Age Groups and Between Sexes in Humans?

The default option is that there is a similarity of the basic pathways of metabolism and the occurrence of metabolites in tissues in regard to the species-to-species extrapolation of cancer hazard and risk. If comparative metabolism studies were to show no similarity between the tested species and humans and a metabolite(s) was the active form, there would be less support for an inference that the animal response(s) relates to humans. In other cases, parameters of metabolism may vary quantitatively between species; this becomes a factor in deciding on an appropriate human equivalent dose based on animal studies, optimally in the context of a toxicokinetic model. Although the basic pathways are assumed to be the same among humans, the presence of polymorphisms and the maturation of the pathways in infants need to be considered. The active form of an agent may be present to differing degrees, or it may be completely absent, which may result in greater or lesser risk for subpopulations.

How Do Toxicokinetic Processes Relate Across Species and Among Different Age Groups and Between Sexes in Humans?

A major issue is how to estimate human equivalent doses in extrapolating from animal studies. *As a default for oral exposure, a human equivalent dose for adults is estimated from data on another species by an adjustment of animal applied oral dose by a scaling factor of body weight to the 0.75 power.* This adjustment factor is used because it represents scaling of

metabolic rate across animals of different size. Because the factor adjusts for a parameter that can be improved on and brought into more sophisticated toxicokinetic modeling when such data become available, the default option of 0.75 power can be refined or replaced. *The same factor is used for children because it is slightly more protective than using children's body weight (see Section 3.1.3).*

For inhalation exposure, a human equivalent dose for adults is estimated by default methodologies that provide estimates of lung deposition and internal dose. The methodologies can be refined to more sophisticated forms with data on toxicokinetic and metabolic parameters of the specific agent. This default option, like the one for oral exposure, is selected in part because it lays a foundation for incorporating better data. *For gases and aerosols, an adjustment is made for infants and children because their breathing rate and body weight differ from those of adults (see Section 3.1.3).* For inhaled particles, the adjustment does not take into account the different size and spacing of airways of children and adults; this difference could result in children and adults retaining particles with a different size distribution and different toxicologic properties. To reduce this uncertainty, EPA is developing a default dosimetry model for children that is based on children's inhalation parameters. The use of information to improve dose estimation from applied to internal to delivered dose is encouraged, including use of toxicokinetic modeling instead of any default, where data are available.

There are important differences between infants, adults, and older adults in the processes of absorption, distribution, and elimination; for example, infants tend to absorb metals through the gut more rapidly and more efficiently than do older children or adults (Calabrese, 1986). Renal elimination is also not as efficient in infants. Although these processes reach adult competency at about the time of weaning, they may have important implications, particularly when the dose-response relationship for an agent is considered to be nonlinear and there is an exposure scenario disproportionately affecting infants, because in these cases the magnitude of dose is more pertinent than the usual approach in linear extrapolation of averaging dose across a lifetime. Efficiency of intestinal absorption in older adults tends to be generally less overall for most chemicals. Another notable difference is that, post-weaning (about 1 year), children have a higher metabolic rate than do adults (Renwick, 1998) and they may toxify or detoxify agents at a correspondingly higher rate.

For a route-to-route exposure extrapolation, *the default option is that an agent that causes internal tumors by one route of exposure will be carcinogenic by another route if it is absorbed by the second route to give an internal dose.* This is a qualitative option and is considered to be public-health conservative. The rationale is that for internal tumors an internal

dose is significant no matter what the route of exposure. Additionally, the metabolism of the agent will be qualitatively the same for an internal dose. The issue of quantitative extrapolation of the dose-response relationship from one route to another is addressed case by case. Quantitative extrapolation is complicated by considerations such as first-pass metabolism, but it is approachable with empirical data. Adequate data are necessary to demonstrate that an agent will act differently by one route versus another route of exposure.

What Is the Correlation of the Observed Dose-Response Relationship to the Relationship at Lower Doses?

If sufficient data are available, a biologically based model for both the observed range and extrapolation below that range may be used. Although no standard biologically based models are in existence, one may be developed if extensive data exist in a particular case and the purpose of the assessment justifies the investment of the resources needed. *The default procedure for the observed range of data when a biologically based model is not used is to use a curve-fitting model for incidence data.*

In the absence of data supporting a biologically based model for extrapolation outside of the observed range, the choice of approach is based on the view of mode of action of the agent arrived at in the hazard assessment.

The basic default is to assume linearity and to use a linear default approach when the mode of action information is supportive of linearity or mode of action is not understood. The linear approach is used when a view of the mode of action indicates a linear response, for example, when a conclusion is made that an agent directly causes alterations in DNA, a kind of interaction that not only theoretically requires one reaction but also is likely to be additive to ongoing, spontaneous gene mutation. Other kinds of activity may have linear implications, for example, linear rate-limiting steps that support a linear procedure also. The linear approach is to draw a straight line between a point of departure from observed data, generally, as a default, an LED chosen to be representative of the lower end of the observed range, and the origin (zero incremental dose, zero incremental response). This approach is generally considered to be public-health protective.

The linear default is thought to generally provide an upper-bound calculation of potential risk at low doses, for example, a 1/100,000 to 1/1,000,000 risk; the straight line approach gives numerical results that are about the same as those from a linearized multistage procedure. This upper bound is thought to be public-health conservative at low doses for the range of human variation, considering the typical Agency target range for risk management of 1/1,000,000 to

1/10,000, although it may not completely be so (Bois et al., 1995) if pre-existing disease or genetic constitution place a percentage of the population at risk from any exposure above zero to xenobiotics, natural or manmade. The question of what may be the actual variation in human susceptibility is one that the NRC (1994) report discussed, as did the NRC report on pesticides in children and infants (1993b). NRC has recommended research on the question, and EPA and other agencies are conducting such research. Given the current state of knowledge, EPA will assume that the linear default procedure adequately accounts for human variation unless there is case-specific information for a given agent that indicates a particularly susceptible subpopulation or lifestage, in which case the special information will be used.

When adequate data on mode of action show that linearity is not plausible and provide sufficient evidence to support a nonlinear mode of action for the general population and any subpopulations of concern, the default changes to a different approach—a reference dose/reference concentration—that assumes that nonlinearity is more reasonable. The departure point is again generally an LED when incidence data are modeled.

A sufficient basis to support this nonlinear procedure will include data on responses that are key events integral to the carcinogenic process. This means that the point of departure mostly will be from these precursor response data, for example, hormone levels or mitogenic effects rather than tumor incidence data.

The mode of action may have specific implications to be considered for risk potential of certain exposure scenarios. For instance, stimulus of cell growth through hormonal or other signal disruption or as a result of damage from toxicity are reversible if the exposure is for a short time, because homeostasis brings a return to normal levels after cessation of exposure. Another feature of a specific exposure scenario may be the exposure of a susceptible subpopulation or lifestage. If those exposed in a particular scenario wholly or largely comprise a subpopulation or lifestage for whom evidence indicates a special susceptibility to the agent's mode of action, this needs to be considered.

When the mode of action information indicates that the dose response may be adequately described by both a linear and a nonlinear approach, then the default is to present both the linear analysis and the reference dose/reference concentration. An assessment may use both linear and nonlinear approaches if linearity is not plausible and nonlinearity has support but a mode of action is not defined or different responses are thought to result from different modes of action or a response appears to be very different at high and low doses due to influence of separate modes of action. The results may be needed for assessment of combined risk from agents that have common modes of action.

A default option is made that cumulative dose received over a lifetime, expressed as a lifetime average daily dose, is an appropriate measure of dose. This assumes that a high dose of such an agent received over a shorter period of time is equivalent to a low dose spread over a lifetime. This is thought to be a relatively public-health-protective option and has empirical support (Monro, 1992). An example of effects of short-term, high exposure that results in subsequent cancer development is treatment of cancer patients with certain chemotherapeutic agents. An example of cancer from long-term exposure to an agent of relatively low potency is smoking. When sufficient information is available to indicate that the carcinogenic mode of action supports a nonlinear dose-response approach, a different approach may be used. In these cases, short-term exposure estimates (several days to several months) may be more appropriate than the lifetime average daily dose, and both agent concentration and duration are likely to be important, because such effects are generally observed to be reversible at cessation of very short-term exposure.

REFERENCES

Allen, BC; Crump, KS; Shipp, AM. (1988) Correlation between carcinogenic potency of chemicals in animals and humans. Risk Anal 8:531–544.

Ames, BN; Gold, LS. (1990) Too many rodent carcinogens: mitogenesis increases mutagenesis. Science 249:970–971.

Ashby, J; Tennant, RW. (1991) Definitive relationships among chemical structure, carcinogenicity and mutagenicity for 301 chemicals tested by the U.S. NTP. Mutat Res 257:229–306.

Ashby, J; Tennant, RW. (1994) Prediction of rodent carcinogenicity for 44 chemicals: results. Mutagenesis 9:7–15.

Ashby, J; Doerrer, NG; Flamm, FG; et al. (1990) A scheme for classifying carcinogens. Regul Toxicol Pharmacol 12:270–295.

Ashby, J; Brady, A; Elcombe, CR; et al. (1994) Mechanistically based human hazard assessment of peroxisome proliferator-induced hepatocarcinogenesis. Hum Exper Toxicol 13:1–117.

Barrett, JC. (1992) Mechanisms of action of known human carcinogens. In: Mechanisms of carcinogenesis in risk identification. IARC Sci Pubs No. 116, 115–134. International Agency for Research on Cancer, Lyon, France.

Barrett, JC. (1993) Mechanisms of multistep carcinogenesis and carcinogen risk assessment. Environ Health Perspect 100:9-20.

Barrett, JC; Lee, TC. (1992) Mechanisms of arsenic-induced gene amplification. In: Kellems, RE, ed. Gene amplification in mammalian cells: a comprehensive guide. New York: Marcel Dekker.

Baylin, S; Bestor, TH. (2002) Altered methylation patterns in cancer cell genomes: causes or consequence? Cancer Cell 1:299–305.

Bayly, AC; Roberts, RA; Dive, C. (1994) Suppression of liver cell apoptosis in vitro by the nongenotoxic hepatocarcinogen and peroxisome proliferator nafenopin. J Cell Biol 125:197–203.

Bellamy, COC; Malcomson, RDG; Harrison, DJ; et al. (1995) Cell death in health and disease: the biology and regulation of apoptosis. Seminars in Cancer Biology, Apoptosis in Oncogenesis and Chemotherapy 6:3–16.

Biggs, PJ; Warren, W; Venitt, S; et al. (1993) Does a genotoxic carcinogen contribute to human breast cancer? The value of mutational spectra in unraveling the etiology of cancer. Mutagenesis 8:275–283.

Birnbaum, LS; Fenton, SE. (2003) Cancer and developmental exposure to endocrine disruptors. Environ Health Perspect (in press).

Birner, G; Albrecht, W; Neumann, HG. (1990) Biomonitoring of aromatic amines. III: hemoglobin binding and benzidine and some benzidine congeners. Arch Toxicol 64(2):97–102.

Blair, A; Burg, J; Foran, J; et al. (1995) Guidelines for application of meta-analysis in environmental epidemiology. Regul Toxicol Pharmacol 22:189–197.

Bois, FY; Krowech, G; Zeise, L. (1995) Modeling human interindividual variability in metabolism and risk: the example of 4-aminobiphenyl. Risk Anal 15:205–213.

Calabrese, EJ. (1986) Age and susceptibility to toxic substances. New York: Winter-Interscience Publication, John Wiley and Sons, Inc.

Callemen, CJ; Ehrenberg, L; Jansson, B; et al. (1978) Monitoring and risk assessment by means of alkyl groups in hemoglobin in persons occupationally exposed to ethylene oxide. J Environ Pathol Toxicol 2:427–442.

Caporaso, N; Hayes, RB; Dosemeci, M; et al. (1989) Lung cancer risk, occupational exposure, and the debrisoquine metabolic phenotype. Cancer Res 49:3675–3679.

Cavenee, WK; Koufos, A; Hansen, MF. (1986) Recessive mutant genes predisposing to human cancer. Mutat Res 168:3–14.

Chang, CC; Jone, C; Trosko, JE; et al. (1988) Effect of cholesterol epoxides on the inhibition of intercellular communication and on mutation induction in Chinese hamster V79 cells. Mutat Res 206:471–478.

Chen, C; Farland, W. (1991) Incorporating cell proliferation in quantitative cancer risk assessment: approaches, issues, and uncertainties. In: Butterworth, B., Slaga, T., Farland, W., et al., eds. Chemical induced cell proliferation: implications for risk assessment. New York: Wiley-Liss, pp. 481–499.

Choy, WN. (1993) A review of the dose-response induction of DNA adducts by aflatoxin B_2 and its implications to quantitative cancer-risk assessment. Mutat Res 296:181–198.

Clayson, DB; Mehta, R; Iverson, F. (1994) Oxidative DNA damage—the effects of certain genotoxic and operationally non-genotoxic carcinogens. Mutat Res 317:25–42.

Cohen, SM. (1995) Role of urinary physiology and chemistry in bladder carcinogenesis. Fd Chem Toxicol 33:715–30.

Cohen, SW; Ellwein, LB. (1990) Cell proliferation in carcinogenesis. Science 249:1007–1011.

Cohen, SM; Ellwein, LB. (1991) Genetic errors, cell proliferation and carcinogenesis. Cancer Res 51:6493–6505.

Cohen, SM; Purtilo, DT; Ellwein, LB. (1991) Pivotal role of increased cell proliferation in human carcinogenesis. Mod Pathol 4:371–375.

Connolly, RB; Andersen, ME. (1991) Biologically based pharmacodynamic models: tools for toxicological research and risk assessment. Ann Rev Pharmacol Toxicol 31:503–523.

Cresteil, T. (1998) Onset of xenobiotic metabolism in children: toxicological implications. Food Addit Contam 15, Supplement 45–51.

D'Souza, RW; Francis, WR; Bruce, RD; et al. (1987) Physiologically based pharmacokinetic model for ethylene chloride and its application in risk assessment. In: Pharmacokinetics in risk assessment: drinking water and health. Vol. 8. Washington, DC: National Academy Press.

Enterline, PE; Henderson, VL; Marsh, GM. (1987) Exposure to arsenic. Amer J Epidemiol 125:929–938.

Executive Order 13045 (1997). Protection of children from environmental health risks and safety risks, issued April 21, 1997.

Fearon, E; Vogelstein, B. (1990) A genetic model for colorectal tumorigenesis. Cell 61:959–967.

Fisher, RA. (1950) Statistical methods for research workers. Edinburgh, Scotland: Oliver and Boyd.

Flynn, GL. (1990) Physicochemical determinants of skin absorption. In: Gerrity, TR, Henry, CJ, eds. Principles of route to route extrapolation for risk assessment. New York: Elsevier Science; pp. 93–127.

Garfinkel, L; Silverberg, E. (1991) Lung cancer and smoking trends in the United States over the past 25 years. Cancer 41:137–145.

Gaylor, DW; Zheng, Q. (1996) Risk assessment of nongenotoxic carcinogens based on cell proliferation/death rates in rodents. Risk Anal 16(2):221–225.

Gaylor, DW; Kodell, RL; Chen, JJ; et al. (1994) Point estimates of cancer risk at low doses. Risk Anal 14(5):843–850.

Gibson, DP; Aardema, MJ; Kerckaert, GA; et al. (1995) Detection of aneuploidy-inducing carcinogens in the Syrian hamster embryo (SHE) cell transformation assay. Mutat.Res 343:7–24.

Ginsberg, GL. (2003) Assessing cancer risks from short-term exposures in children. Risk Anal 23(1):19-34.

Goddard, MJ; Murdoch, DJ; Krewski, D. (1995). Temporal aspects of risk characterization. Inhal Toxicol 7:1005–1018.

Goldsworthy, TL; Hanigan, MH; Pitot, HC. (1986) Models of hepatocarcinogenesis in the rat–contrasts and comparisons. CRC Crit Rev Toxicol 17:61–89.

Goodman, JI; Counts, JL. (1993) Hypomethylation of DNA: A possible nongenotoxic mechanism underlying the role of cell proliferation in carcinogenesis. Environ Health Perspect 101 Supp. 5:169–172.

Greenland, S. (1987) Quantitative methods in the review of epidemiologic literature. Epidemiol Rev 9:1–29.

Gulezian, D; Jacobson-Kram, D; McCullough, CB; et al. (2000) Use of transgenic animals for carcinogenicity testing: considerations and implications for risk assessment. Toxicol Pathol 28:482–499.

Halmes, NC; Roberts, SM; Tolson, JK; et al. (2000) Reevaluating cancer risk estimates for short-term exposure scenarios. Toxicol Sci 58:32–42.

Hammand, EC. (1966) Smoking in relation to the death rates of one million men and women. In: Haenxzel, W., ed. Epidemiological approaches to the study of cancer and other chronic diseases. National Cancer Institute Monograph No. 19. Washington, DC.

Hanahan, D; Weinberg, RA. (2000) The hallmarks of cancer. Cell 100:57–70.

Harris, CC; Hollstein, M. (1993) Clinical implications of the p53 tumor suppressor gene. N Engl. J Med 329:1318–1327.

Haseman, JK. (1983) Issues: a reexamination of false-positive rates for carcinogenesis studies. Fundam Appl Toxicol 3:334–339.

Haseman, JK. (1984) Statistical issues in the design, analysis and interpretation of animal carcinogenicity studies. Environ Health Perspect 58:385–392.

Haseman, JK. (1985) Issues in carcinogenicity testing: dose selection. Fundam Appl Toxicol 5:66–78.

Haseman, JK. (1990) Use of statistical decision rules for evaluating laboratory animal carcinogenicity studies. Fundam Appl Toxicol 14:637–648.

Haseman, JK. (1995) Data analysis: Statistical analysis and use of historical control data. Regul Toxicol Pharmacol 21:52–59.

Hatch, EE, Palmer, JR, Titus-Ernstoff, L, Noller, KL et al. (1998) Cancer risk in women exposed to diethylstilbestrol in utero. JAMA 280: 630-634.

Hattis, D. (1990) Pharmacokinetic principles for dose-rate extrapolation of carcinogenic risk from genetically active agents. Risk Anal 10:303–316.

Hayward, JJ; Shane, BS; Tindall, KR; et al. (1995) Differential in vivo mutagenicity of the carcinogen-noncarcinogen pair 2,4- and 2,6-diaminotoluene. Carcinogenesis 10:2429–2433.

Herbst, AL, Ulfelder, H, Poskanzer, DC. (1971) Adenocarcinoma of the vagina: association of maternal stilbestrol therapy with tumor appearance in young women. N Engl J Med 284:878-881.

Hill, AB. (1965) The environment and disease: association or causation? Proc R Soc Med 58:295–300.

Hoel, DG; Kaplan, NL; Anderson, MW. (1983) Implication of nonlinear kinetics on risk estimation in carcinogenesis. Science 219:1032–1037.

Holliday, R. (1987) DNA methylation and epigenetic defects in carcinogenesis. Mutat Res 181:215–217.

Huff, JE. (1993) Chemicals and cancer in humans: first evidence in experimental animals. Environ Health Perspect 100:201-210.

Huff, JE. (1994) Chemicals causally associated with cancers in humans and laboratory animals. A perfect concordance. In: Carcinogenesis. Waalkes, MP, Ward, JM, eds., New York: Raven Press; pp. 25-37.

Hulka, BS; Margolin, BH. (1992) Methodological issues in epidemiologic studies using biological markers. Am J Epidemiol 135:122–129.

IARC (International Agency for Research on Cancer). (1994) IARC monographs on the evaluation of carcinogenic risks to humans. Vol. 60, Some Industrial Chemicals. Lyon, France: IARC; pp. 13-33.

IARC. (1999) The use of short- and medium-term tests for carcinogens and data on genetic effects in carcinogenic hazard evaluation. Lyon, France.

ILSI (International Life Sciences Institute). (1992). Similarities & differences between children & adults; implications for risk assessment." Washington, DC: ILSI Press.

ILSI. (1997) Principles for the selection of doses in chronic rodent bioassays. Foran, JA, ed. Washington, DC: ILSI Press.

ILSI. (2001) Proceedings of workshop on the evaluation of alternative methods for carcinogenesis testing. Toxicol Pathol 29:1–351.

IPCS (International Programme on Chemical Safety). (1999) IPCS workshop on developing a conceptual framework for cancer risk assessment, 16-18 February 1999, Lyon, France, IPCS/99.6. IPCS, World Health Organization, Geneva.

Ito, N; Shirai, T; Hasegawa, R. (1992) Medium-term bioassays for carcinogens. In: Vainio, H, Magee, PN, McGregor, DB, et al., eds. Mechanisms of carcinogenesis in risk identifications. International Agency for Research on Cancer, Lyon, France; pp. 353–388.

Jones, PA. (1986) DNA methylation and cancer. Cancer Res 46:461–466.

Kehrer, JP. (1993) Free radicals as mediators of tissue injury and disease. Crit Rev Toxicol 23:21–48.

Kelsey, JL; Thompson, WD; Evans, AS. (1986) Methods in observational epidemiology. New York: Oxford University Press.

Kimbell, JS; Subramaniam, RP; Gross, EA; Schlosser, PM; Morgan, KT. (2001) Dosimetry modeling of inhaled formaldehyde: comparisons of local flux predictions in the rat, monkey and human nasal passages. Toxicol Sci 64(1):100-110.

Kinzler, KW; Vogelstein, B. (2002) Colorectal tumors. In: Vogelstein, B; Kinzler, KW, eds. The genetic basis of human cancer. New York: McGraw-Hill.

Kinzler, KW; Nilbert, MC; Su, L-K; et al. (1991) Identification of FAP locus genes from chromosome 5q21. Science 253:661–665.

Kraus, AL; Munro, IC; Orr, JC; et al. (1995) Benzoyl peroxide: an integrated human safety assessment for carcinogenicity. Regul Toxicol Pharmacol 21:87–107.

Krewski, D; Van Ryzin, J. (1981) Dose response models for quantal response toxicity data. In: Csorgo; Dawson; Rao; et al., eds. Statistics and related topics. Amsterdam: North-Holland, pp. 201–231.

Krewski, D; Murdoch, DJ; Withey, JR. (1987) The application of pharmacokinetic data in carcinogenic risk assessment. In: Pharmacokinetics in risk assessment: drinking water and health. Vol. 8. Washington, DC: National Academy Press; pp. 441–468.

Levine, AJ; Perry, ME; Chang, A; et al. (1994) The 1993 Walter Hubert lecture: the role of the p53 tumor-suppressor gene in tumorigenesis. Br J Cancer 69:409–416.

Lijinsky, W. (1993) Species differences in carcinogenesis. In Vivo 7:65-72.

Lilienfeld, AM; Lilienfeld, D. (1979) Foundations of epidemiology, 2nd ed. New York: Oxford University Press.

Littlefield, NA; Farmer, JH; Gaylor, DW. (1980) ED01 study. J Environ Pathol Toxicol 3:17.

Maltoni, C; Lefemine, G; Ciliberti, A; et al. (1981) Carcinogenicity bioassay of vinyl chloride monomer: a model of risk assessment on an experimental basis. Environ Health Perspect 41:3–29.

Maronpot, RR; Shimkin, MB; Witschi, HP; et al. (1986) Strain A mouse pulmonary tumor test results for chemicals previously tested in National Cancer Institute carcinogenicity test. J Natl Cancer Inst 76:1101–1112.

Marsman, DS; Popp, JA. (1994) Biological potential of basophilic hepatocellular foci and hepatic adenoma induced by the peroxisome proliferator, Wy-14,643. Carcinogenesis 15:111–117.

Mausner, JS; Kramer, S. (1985) Epidemiology, 2nd ed. Philadelphia: W.B. Saunders.

McConnell, EE. (1992) Comparative response in carcinogenesis bioassay as a function of age at first exposure. In: Guzelian, P; Henry, CJ; Olin, SS, eds. Similarities and difference between children and adults: implications for risk assessment. Washington, DC: ILSI Press; pp. 66-78.

McConnell, EE; Solleveld, HA; Swenberg, JA; et al. (1986) Guidelines for combining neoplasms for evaluation of rodent carcinogenesis studies. J Natl Cancer Inst 76:283–289.

McMichael, AJ. (1976) Standardized mortality ratios and the "healthy worker effect": scratching beneath the surface. J Occup Med 18:165–168.

Melnick, RL, Huff, JE, Barrett, JC, Maronpot, RR, Lucier, G, Portier, CJ. (1993) Cell proliferation and chemical carcinogenesis: A symposium overview. Mol Carcinog 7:135-138.

Miller, RW. (1995) Special susceptibility of the child to certain radiation-induced cancers. Environ Health Perspect 103(suppl 6):41–44.

Monro, A. (1992) What is an appropriate measure of exposure when testing drugs for carcinogenicity in rodents? Toxicol Appl Pharmacol 112:171-181.

Moolgavkar, SH; Knudson, AG. (1981) Mutation and cancer: a model for human carcinogenesis. J Natl Cancer Inst 66:1037–1052.

Morrison, V; Ashby, J. (1994) A preliminary evaluation of the performance of the muta™ mouse (lacZ) and Big Blue™ (lacI) transgenic mouse mutation assays. Mutagenesis 9:367–375.

Murdoch, DJ; Krewski, D; Wargo, J. (1992) Cancer risk assessment with intermittent exposure. Risk Anal 12(4):569–577.

Murrell, JA; Portier, CJ; Morris, RW. (1998) Characterizing dose-response I: critical assessment of the benchmark dose concept. Risk Anal 18(1):13–25.

NRC (National Research Council). (1983) Risk assessment in the federal government: managing the process. Committee on the Institutional Means for Assessment of Risks to Public Health, Commission on Life Sciences, NRC. Washington, DC: National Academy Press.

NRC. (1990) Health effects of exposure to low levels of ionizing radiation (BEIR V). Washington, DC: National Academy Press.

NRC. (1993a) Issues in risk assessment. Committee on Risk Assessment Methodology. Washington, DC: National Academy Press.

NRC. (1993b) Pesticides in the diets of infants and children. Washington, DC: National Academy Press.

NRC. (1994) Science and judgment in risk assessment. Washington, DC: National Academy Press.

NRC. (1996) Understanding risk: informing decisions in a democratic society. Washington, DC: National Academy Press.

NRC. (2002) Estimating the public health benefits of proposed air pollution regulations. Washington, DC: National Academy Press.

NTP (National Toxicology Program). (1984) Report of the ad hoc panel on chemical carcinogenesis testing and evaluation of the National Toxicology Program, Board of Scientific Counselors. Washington, DC: U.S. Government Printing Office. 1984-421-132:4726.

OECD (Organization for Economic Cooperation and Development). (1981) Guidelines for testing of chemicals. Carcinogenicity studies. No. 451. Paris, France.

OMB (Office of Management and Budget). (2002) Guidelines for ensuring and maximizing the quality, objectivity, utility, and integrity of information disseminated by federal agencies. Federal Register 67(36):8451-8460. Available from: http://www.epa.gov/oci/qualityguidelines/fr22fe02-117.htm.

OSTP (Office of Science and Technology Policy). (1985) Chemical carcinogens: review of the science and its associated principles. Federal Register 50:10372-10442.

Peltomäki, P; Aaltonen, LA; Sisonen, P; et al. (1993) Genetic mapping of a locus predisposing human colorectal cancer. Science 260:810–812.

Peto, J. (1992) Meta-analysis of epidemiological studies of carcinogenesis. In: Mechanisms of carcinogenesis in risk assessment. IARC Sci. Pubs. No. 116, Lyon, France; pp. 571–577.

Peto, J; Darby, S. (1994) Radon risk reassessed. Nature 368:97–98.

Peto, R; Gray, R; Brantom, P; et al. (1984) Nitrosamine carcinogenesis in 5120 rodents: chronic administration of sixteen different concentrations of NDEA, NDMA, NPYR and NPIP in the water of 4440 inbred rats, with parallel studies on NDEA alone of the effect of age of starting (3,6, or 20 weeks) and of species (rats, mice or hamsters). IARC Sci Publ 57:627–665.

Pinkerton, KE; Joad, J. (2000) The mammalian respiratory system and critical windows of exposure for children's health. Environ Health Perspect 108(suppl):457–462.

Portier, C. (1987) Statistical properties of a two-stage model of carcinogenesis. Environ Health Perspect 76:125–131.

Rall, DP. (1991) Carcinogens and human health: part 2. Science 251:10–11.

Regulatory Toxicology and Pharmacology. (1996) 24:126–40

Renwick, AG. (1998) Toxicokinetics in infants and children in relation to the ADI and TDI. Food Addit Contam 15, Suppl 17–35.

Rice, JM. (1979) Problems and perspective in perinatal carcinogenesis: a summary of the conference. NCI Monogr 51:271-278.

Rothman, KT. (1986) Modern Epidemiology. Boston: Little, Brown and Company.

Rouse, J; Jackson, SP. (2002) Interfaces between the detection, signaling, and repair of DNA damage. Science 297:547–551.

Shelby, MD; Zeiger, E. (1990) Activity of human carcinogens in the *Salmonella* and rodent bone-marrow cytogenetics tests. Mutat Res 234:257–261.

Sisk, SC; Pluta, LJ; Bond, JA; et al. (1994) Molecular analysis of lacI mutants from bone marrow of B6C3F1 transgenic mice following inhalation exposure to 1,3-butadiene. Carcinogenesis 15(3):471–477.

Snedecor, GW; Cochran, WG. (1967) Statistical methods, 6th ed. Ames, Iowa: Iowa State University Press.

Spalding, JW; French, JE; Stasiewicz, S; Furedi-Machacek, M; Conner, F; Tice, RR; Tennant, RW. (2000) Responses of transgenic mouse lines p53(+/-) and Tg.AC to agents tested in conventional carcinogenicity bioassays. Toxicol Sci 53(2)213-223.

Stiteler, WH; Knauf, LA; Hertzberg, RC; et al. (1993) A statistical test of compatibility of data sets to a common dose-response model. Regul Toxicol Pharmacol 18:392–402.

Subramaniam, RP; Asgharian, B; Freijer, JI; Miller, FJ; Anjilvel, S. (2003) Analysis of differences in particle deposition in the human lung. Inhalation Toxicol 15:1-21.

Swierenga, SHH; Yamasaki, H. (1992) Performance of tests for cell transformation and gap junction intercellular communication for detecting nongenotoxic carcinogenic activity. In: Mechanisms of carcinogenesis in risk identification. IARC Sci. Pubs. No. 116, Lyon, France; pp. 165–193.

Tarone, RE. (1982) The use of historical control information in testing for a trend in proportions. Biometrics 38:215–220.

Taylor, JH; Watson, MA; Devereux, TR; et al. (1994) p53 mutation hotspot in radon-associated lung cancer. Lancet 343:86–87.

Tennant, RW. (1993) Stratification of rodent carcinogenicity bioassay results to reflect relative human hazard. Mutat Res 286:111–118.

Tennant, RW; French, JE; Spalding, JW. (1995) Identifying chemical carcinogens and assessing potential risk in short-term bioassays using transgenic mouse models. Environ Health Perspect 103:942–950.

Tennant, RW; Stasiewicz, S; Mennear, J; et al. (1999) Genetically altered mouse models for identifying carcinogens. In: McGregor, DB; Rice, JM; Venitt, S, eds. The use of short- and medium-term tests for carcinogens and data on genetic effects in carcinogenic hazard evaluation. Lyon, France: International Agency for Research on Cancer.

Tinwell, H; Ashby, J. (1991) Activity of the human carcinogen MeCCNU in the mouse bone marrow mironucleus test. Environ Molec Mutagen 17:152–154.

Todd, GC. (1986) Induction of reversibility of thyroid proliferative changes in rats given an antithyroid compound. Vet Pathol 23:110–117.

Tomatis, L; Aitio, A; Wilbourn, J; et al. (1989) Human carcinogens so far identified. Jpn J Cancer Res 80:795–807.

U.S. Environmental Protection Agency (U.S. EPA). (1986a) Guidelines for carcinogen risk assessment. Federal Register 51(185):33992–34003. Available from: http://www.epa.gov/ncea/raf/.

U.S. EPA. (1986b) Guidelines for mutagenicity risk assessment. Federal Register 51(185):34006-34012. Available from: http://cfpub.epa.gov/ncea/raf/recordisplay.cfm?deid=23160.

U.S. EPA. (1989) Summary of the second workshop carcinogenesis bioassay with the dermal route. May 18-19, 1988, Research Triangle Park, NC. EPA/560/6-89/003, available from NTIS, 5284 Port Royal Road, Springfield, VA 22161 (703-487-4650).

U.S. EPA. (1991a) Guidelines for developmental toxicity risk assessment. Federal Register 56(234):63798-63826. Available from: http://cfpub.epa.gov/ncea/raf/recordisplay.cfm?deid=23162.

U.S. EPA. (1991b) Alpha-2u-globulin: association with chemically induced renal toxicity and neoplasia in the male rat. Risk Assessment Forum, Washington, DC. EPA/625/3-91/019F.

U.S. EPA. (1992a) Guidelines for exposure assessment. Federal Register 57(104):22888-22938. Available from: http://cfpub.epa.gov/ncea/raf/recordisplay.cfm?deid=15263.

U.S. EPA. (1992b) Draft report: a cross-species scaling factor for carcinogen risk assessment based on equivalence of $mg/kg^{3/4}$/day. Federal Register 57(109):24152-24173.

U.S. EPA. (1994) Methods for derivation of inhalation reference concentrations and application of inhalation dosimetry. Office of Health and Environmental Assessment, Environmental Criteria and Assessment Office, Research Triangle Park, NC. EPA/600/8-90/066F.

U.S. EPA. (1995) Policy for risk characterization. Memorandum of Carol M. Browner, Administrator, March 21, 1995, Washington, DC. Available from: http://www.epa.gov/osp/spc/2riskchr.htm.

U.S. EPA. (1996a) Guidelines for reproductive toxicity risk assessment. Federal Register 61(212):56274-56322. Available from: http://cfpub.epa.gov/ncea/raf/recordisplay.cfm?deid=2838.

U.S. EPA. (1996b) Comparison of the effects of chemicals with combined perinatal and adult exposure vs. Adult only exposure in carcinogenesis studies. Office of Pesticide Programs, October 1996.

U.S. EPA. (1997a) A proposed OPP policy on determining the need for perinatal carcinogenicity testing on a pesticide. Office of Pesticide Programs, 14 August 1997.

U.S. EPA. (1997b) A set of scientific issues being considered by the Agency in connection with the criteria for requiring in-utero cancer studies. Office of Pesticide Programs. FIFRA Scientific Advisory Panel. September 1997 meeting report. Available from: http://www.epa.gov/pesticides/SAP/archive/september/finalsep.htm.

U.S. EPA. (1997c) Exposure factors handbook. National Center for Environmental Assessment, Washington, DC. EPA/600/P-95/002F. Available from: http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=12464.

U.S. EPA. (1997d) Policy for use of probabilistic analysis in risk assessment. Memorandum of Fred Hansen, Deputy Administrator, May 15, 1997. Available from: http://www.epa.gov/osp/spc/probpol.htm.

U.S. EPA. (1997e) Guiding principles for Monte Carlo analysis. Risk Assessment Forum, Washington, DC. EPA/630/R-97/001. Available from: http://cfpub.epa.gov/ncea/raf/recordisplay.cfm?deid=29596.

U.S. EPA. (1998a) Assessment of thyroid follicular cell tumors. Risk Assessment Forum, Washington, DC. EPA/630/R-97/002. Available from: http://cfpub.epa.gov/ncea/raf/recordisplay.cfm?deid=13102.

U.S. EPA. (1998b) Guidelines for neurotoxicity risk assessment. Federal Register 63(93):26926-26954. Available from: http://cfpub.epa.gov/ncea/raf/recordisplay.cfm?deid=12479.

U.S. EPA. (1998c) Health effects test guidelines: OPPTS 870.4300 combined chronic toxicity/carcinogenicity. Office of Prevention, Pesticides and Toxic Substances, Washington, DC. EPA/712/C-98/212. Available from: http://www.epa.gov/opptsfrs/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Series/

U.S. EPA. (1998d) EPA's rule writer's guide to Executive Order 13045. Available from: http://yosemite.epa.gov/ochp/ochpweb.nsf/content/whatwe_regulate.htm

U.S. EPA. (1999a) Guidelines for carcinogen risk assessment (review draft). Risk Assessment Forum, Washington, DC. NCEA-F-0644. Available from: http://www.epa.gov/ncea/raf/cancer.htm.

U.S. EPA. (1999b) Review of revised sections of the proposed guidelines for carcinogen risk assessment. Science Advisory Board, Washington, DC. EPA/SAB/EC-99/015. Available from: http://www.epa.gov/ncea/raf/cancer.htm.

U.S. EPA. (1999c) Cancer risk coefficients for environmental exposure to radionuclides: federal guidance report no. 13. Office of Air and Radiation. EPA/402/R-99/001. Available from: http://www.epa.gov/radiation/federal.

U.S. EPA. (2000a) Science Policy Council handbook: peer review. Office of Research and Development, Office of Science Policy, Washington, DC. EPA/100/B-98/001. Available from: http://www.epa.gov/osp/spc/prhandbk.pdf.

U.S. EPA. (2000b) U.S. EPA. Science Policy Council handbook: risk characterization. EPA Science Policy Council, Washington, DC. EPA/100/B-00/002. Available from: http://www.epa.gov/osp/spc/rchandbk.pdf

U.S. EPA. (2000c) Supplementary guidance for conducting health risk assessments of chemical mixtures. Risk Assessment Forum, Washington, DC. EPA/630/R-00/002. Available from: http://cfpub.epa.gov/ncea/raf/recordisplay.cfm?deid=20533.

U.S. EPA. (2000d) Guidance for data quality assessment: practical methods for data analysis. Office of Environmental Information, Washington, DC. EPA/600/R-96/084. Available from: http://www.epa.gov/quality/qs-docs/g9-final.pdf.

U.S. EPA. (2002a) Guidelines for ensuring and maximizing the quality, objectivity, utility and integrity for information disseminated by the Environmental Protection Agency. Office of Environmental Information, Washington, DC. EPA/260/R-02/008. Available from: http://www.epa.gov/oei/qualityguidelines/index.html.

U.S. EPA. (2002b) A review of the reference dose and reference concentration process (external review draft). Risk Assessment Forum, Washington, DC. EPA/630/P-02/002A. Available from: http://cfpub.epa.gov/ncea/raf/recordisplay.cfm?deid=51717.

U.S. EPA. (2002c) Workshop on the benefits of reductions in exposure to hazardous air pollutants: developing best estimates of dose-response functions. Science Advisory Board, Washington, DC. EPA/SAB-EC/WKSHP/02/001. Available from: http://www.epa.gov/science1/fiscal02.htm.

U.S. EPA. (2002d) Child-specific exposure factors handbook (interim report). EPA/600/P-00/002B. Office of Research and Development, National Center for Environmental Assessment, Washington, DC, 448 pp. Available from: http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=55145.

U.S. EPA. (2003) Supplemental guidance for assessing cancer susceptibility from early-life exposure to carcinogens (external review draft). Risk Assessment Forum, Washington, DC. Available from: http://www.epa.gov/ncea/raf/cancer2003.htm.

Vainio, H; Magee, P; McGregor, D; et al. (1992) Mechanisms of carcinogenesis in risk identification. IARC Sci. Pubs. No. 116. Lyon, France: IARC.

Van Sittert, NJ; De Jong, G; Clare, MG; et al. (1985) Cytogenetic, immunological, and hematological effects in workers in an ethylene oxide manufacturing plant. Br J Indust Med 42:19–26.

Vater, ST; McGinnis, PM; Schoeny, RS; et al. (1993) Biological considerations for combining carcinogenicity data for quantitative risk assessment. Regul. Toxicol Pharmacol 18:403–418.

Vesselinovitch, SD; Rao, KVN; Mihailovich, N. (1979) Neoplastic response of mouse tissues during perinatal age periods and its significance in chemical carcinogenesis. NCI Monogr 51:239.

Vogelstein, B; Fearon, ER; Hamilton, SR; et al. (1988) Genetic alterations during colorectal-tumor development. N Eng J Med 319:525–532.

Whysner, J; Williams, GM. (1996) Saccharin mechanistic data and risk assessment: urine composition, enhanced cell proliferation, and tumor promotion. Pharmacol Ther 71: 225:252.

Woo, YT; Arcos, JC. (1989) Role of structure-activity relationship analysis in evaluation of pesticides for potential carcinogenicity. In: Ragsdale, NN; Menzer, RE, eds. Carcinogenicity and pesticides: principles, issues, and relationship. ACS Symposium Series No. 414. San Diego: Academic Press; pp. 175–200.

Yamasaki, H. (1995) Non-genotoxic mechanisms of carcinogenesis: Studies of cell transformation and gap junctional intercellular communication. Toxicol Lett 77:55–61.