Summary Response to Public Comments on EPA's Draft Cancer Risk Assessment Guidelines of 1996, 1999, and 2003

This summary response to public comments covers comments received on three distinct drafts of proposed revisions to EPA's cancer guidelines since the last time the guidelines were finalized in 1986. In this time period, these guidelines have undergone major changes in content, in part due to the comments made by the public and due in part to the advances have occurred in toxicology, carcinogenesis, and risk assessment. Taken together, these changes have overtaken some of the comments provided on the earlier revisions. Thus, this summary response to comments focuses on the major themes that were raised over the years and the responses to them that often resulted in significant and substantial changes in the guidelines.

Several commentors requested specific examples of how various aspects of the guidelines would be implemented. These requests included: how the selection of descriptors and their associated narratives are influenced by different data sets, application of the MOA framework for data rich and data poor situations, determining points of departure (PODs) with tumor data and precursor data for both linear and nonlinear extrapolations, and the use of precursor or toxicokinetic information in dose-response assessments. Earlier drafts of the guidelines contained a limited number of examples, commentors generally liked the idea of examples but expressed frustration that the examples were not sufficiently detailed or did not sufficiently captured the nuances of the example. Another problem with providing examples within the guidelines is that the examples have a tendency to become dated and not reflect current Agency practice, especially with regard to advances in the field. Therefore, EPA is not including examples within the guidelines, rather it is EPA's intent to develop a website that contains examples from actual Agency-approved analyses of chemicals, rather than hypothetical examples. This web site (http://www.epa.gov/cancerguidelines) can provide a more robust discussion of these examples and can be more easily updated to reflect current practice.

Default Assumptions

Prior to the 2003 draft guidelines, EPA's position was often to use default assumptions and models unless there was sufficient reason to move from the default position. The 2003 draft and the current guidelines specify that a critical analysis of available information is the starting point for evaluation and that defaults should only be invoked when necessary. This change in position obviates a direct response to inquiries regarding the development of specific guidance and criteria for moving from default assumptions.

Commentors were divided between thinking that defaults should continue to be used because they constitute health-protective assumptions and methodologies versus thinking that default options should be avoided to the greatest extent possible. EPA believes the current approach optimizes the use of relevant data while using health-protective defaults when data are absent. A number of commentors supported the revised approach, but noted that the text had not

been changed consistently throughout the document and that the approach to the use of default options was inconsistently presented in various sections. These parts of the text have been revised to eliminate the inconsistencies.

A continuing concern was the degree of conservatism attained by aggregate use of defaults, as well as the need for an estimate of the magnitude of the impact of use of specific default assumptions on the final risks estimates. This issue is one that commonly arises with regard to many of EPA's and other regulatory agencies' risk assessment procedures; it was discussed in EPA's recent review of its risk assessment practices (available at http://www.epa.gov/osa/ratf.htm). The Agency's risk assessors and risk managers are continuing to re-examine methods for appropriately addressing uncertainty and variability when data are not available, but will continue to use risk-protective, i.e., conservative, methods while attempting to prevent overly conservative risk estimates when possible.

The remaining comments appeared to be directed toward default assumptions associated with protection of susceptible subpopulations, most notably the fetus, infant, and child. These comments will be discussed in a separate section.

Human Studies

Issue: Preference of human data over animal data.

Commentors for all three drafts expressed concern about whether human data would really be given preference over animal data and requested a more detailed discussion, further guidance, and case examples. Commentors agreed that preference should be given to human data over animal or other data. Commentors were concerned several statements that infer that animal data could be the major determinant of the WOE, even though the guidelines note the preference of human data over animal data.

The guidelines text on epidemiology has been extensively rewritten and reorganized to address such issues. The preference for good and useful data from human exposure is more explicit, and use of supplementary data to complement animal data are also emphasized.

Issue: Adequacy and quality of epidemiology studies.

For all drafts, commentors requested additional or explicit guidelines, including criteria for study adequacy, choosing exposure metrics and sources of bias and confounding. A better explanation of the evaluation of the lack of agreement in endpoints in epidemiology studies was requested.

This section of the guidelines is not intended to be a primer on epidemiology; many excellent examples exist and are larger than this document. Rather, the functions of this section are to emphasize the importance of human data when available; highlight some of the critical

considerations in evaluating epidemiological data for the purposes of risk assessment, e.g., those that might limit the utility of such data; and provide a basis for the joint evaluation of data from human and laboratory animal exposure(s). Nevertheless, this section of the guidelines has been expanded, and concepts from the Surgeon General's report that was issued subsequent to the 2003 draft have been included and the report cited in these guidelines.

Issue: Epidemiological data that are not positive.

Some comments stated that both positive and not positive epidemiological data of acceptable quality should be considered as equally valuable for assessment of potential human cancer hazard. There was concern that epidemiologic studies that were not positive would be discounted and greater weight given to positive studies. Other comments requested the guidelines be modified to state that conclusions from null results be restricted, if made at all, to the dose level and populations studied.

The Agency intends to use all relevant, available data in its analysis. Furthermore, as exemplified by the MOE framework and the definitions that distinguish between conflicting and differing data, risk assessors are encouraged to weigh the data within the appropriate legal or regulatory context. The Surgeon General's report, mentioned in the previous response, discusses the value and utility of epidemiological data that are not positive. It also discusses how to determine the weight of evidence for human data. A discussion of these issues has been added to the text of the guidelines, and the reference has been provided so that the complete analysis can be examined by a risk assessor.

Issue: Meta-analysis.

The same discussion of meta-analysis was presented in all three drafts. Additional guidance was requested for application of meta-analysis. Some recommended that this approach be used with caution and emphasized the difficult of conducting valid meta-analyses.

The Agency views meta-analysis as one of the useful tools for analyzing and combining multiple data sets. Like all tools, it must be used carefully by a knowledgeable practitioner. As mentioned in the received comments, not all data are amenable to meta-analysis, and various methods for weighting the data sets may add, rather than reduce, uncertainty. The use of meta-analytic techniques will be considered and justified on a case-by-case basis.

Issue: Causality.

The discussion of causality is the same in the 1996 and 1999 drafts. In the 2003 discussion, additional criteria of experimental evidence (from human populations) and analogy, e.g., structure-activity relationships (SAR) were added. The addition of explicit language in the 2005 guidelines addresses a comment on the 1996 draft that the Bradford-Hill criteria are but a starting point to draw conclusions of causality. The current guidelines state that, although these

considerations are usually called “criteria,” they do not all need to be fulfilled to establish causality. A discussion of this issue, as presented in the Surgeon General’s report is emphasized in the guidelines. The guidelines acknowledge that biological significance can be present without statistical significance and vice versa. Both positive and not positive results will be evaluated in the WOE process.

Animal Studies

Issue: Maximum Tolerated Dose (MTD).

Commentors requested expanded discussions regarding high doses (e.g., the MTD) in animal carcinogenicity. Several commentors thought that many cancers in animal bioassays results from high-dose effects, and thus, artifacts of use of the MTD and not applicable to low-dose environmental exposures in humans.

EPA acknowledges the potential limitations of using the MTD. The guidelines highlight some issues to consider if the dose is excessively high. The guidelines also recognize that toxicological mechanisms, mode(s) of action, and effects may differ with dose. By expanding the types of biological and toxicological information that can be included in the WOE analysis and allowing different MOAs for different levels of exposure, the Agency expects that some of these concerns will be resolved. The data available that are available generally result from standard protocols that usually use doses used are fractions of the MTD. Thus, the need to estimate risks a ambient levels of exposure may require extrapolation from data that are fractions of the MTD.

Issue: Site concordance.

Questions remain on the relationship between site concordance and MOA. Commentors disagreed whether target organ concordance should be required to assume relevance to humans. The issues of MOA and site concordance are clearly interrelated. When no MOA has been determined, the Agency will assume that the MOA may not be site dependent and will evaluate tumors from all sites. To the extent that the MOA is linked to one or more specific sites, site concordance would support the determination. To the extent that MOA may not be site dependant, e.g., mutagenesis, site concordance would be less of a concern.

Issue: Combining tumors.

Commentors question when it is appropriate to combine benign and malignant tumors and when to combine across tumor sites. Commentors disagreed as to whether the discussion of how to assess benign tumor relevance was good or confusing.

In analyzing animal bioassay data on the occurrence of multiple tumor types, these guidelines outline a number of biological and other factors to consider. The objective is to use

these factors to select response data (including nontumor data as appropriate) that best represent the biology observed. Appropriate options include use of a single data set, combining data from different experiments, showing a range of results from more than one data set, showing results from analysis of more than one tumor response based on differing modes of action, representing total response in a single experiment by combining animals with tumors, or a combination of these options. The approach judged to best represent the data is presented with the rationale for the judgment, including the biological and statistical considerations involved. EPA has considered the approach of summing tumor incidences and decided not to adopt it. While multiple tumors may be independent, in the sense of not arising from metastases of a single malignancy, it is not clear that they can be assumed to represent different effects of the agent on cancer processes. In this connection, it is not clear that summing incidences provides a better representation of the underlying mode(s) of action of the agent than combining animals with tumors or using another of the several options noted above.

Issue: Statistical tests.

Questions remain regarding biological significance of some statistically significant results, use and interpretation of additional statistical tests, and appropriate level(s) of statistical significance for tumor analysis. The guidelines seek to address some of the general considerations in the statistical analysis of the data, as well as to highlight some of the important issues. As data sets will differ, the types of statistical analyses that are appropriate, as well as their interpretation, may be specific to that situation and will be explained in each risk assessment.

Issue: Historical controls.

The lack of recognition of the importance of historical control data was a concern, as was whether historical control data would actually be used in the assessment. Concern was raised by commentors that the guidelines undervalue the usefulness of historical control tumor data and are too restrictive, making it more difficult to use historical control data in cases where concurrent control data may be unusually low. There was also concern that historical control data would be used in a manner inconsistent with a WOE approach and might be construed to yield a positive result. Commentors suggested that historical control data be reviewed on a case-by-case basis so that data from other laboratories or longer periods of time could be considered, particularly in cases where a laboratory might have a small historical control database.

The Agency agrees with the last comment.

Mode of Action (MOA)

Many of the major issues that have been raised regarding regulatory cancer risk assessments are redirected by EPA's use of mode of action (MOA) for carcinogenesis as a method for addressing those issues. For example, MOA is used to determine whether linear or nonlinear methods will be use for extrapolation to lower levels of responses, as well as to consider changes in the MOA with dose.

General considerations

The use of MOA data was overwhelmingly favored by most commentors. Over the years, comments on this issue included: placing emphasis on the MOA, i.e., the more general concept of the critical step(s), rather than the mechanism of action, i.e., a complete knowledge of the process, so that mode of action can be determined even if the precise mechanism of action is not fully understood, and presenting a more balanced consideration of epidemiologic and toxicologic evidence for and against specific MOA. Several commentors voiced concerns that the 2003 draft was not as comprehensive, in terms of clarity, balance, completeness, and precision, as the 1999 draft, and that little consideration was given to incorporating advances in cancer biology into MOA determination. Some suggested that IPCS draft conceptual framework for evaluating MOA be incorporated into the guidelines.

EPA's framework for determining MOA is not exclusive. The concept of mode of action as an organizing framework for consideration of hazard characterization and dose-response has been discussed and adopted by other toxicology and risk assessment groups. References to these other activities are provided. EPA's framework for determining the MOA is consistent with, although somewhat different than, some of the other published frameworks, but all have certain common themes, such as the need to demonstrate the MOA in the test species and then to further justify its selection as the MOA in humans. The EPA framework is more flexible than some because it allows for different MOAs for different routes of exposure or different patterns, including intensity, of exposures. It also examines potentially susceptible lifestages and populations.

Issue: Evaluating an Hypothesized Mode of Action

Commentors supported independent peer reviews of chemical-specific MOAs. Comments were generally similar, except that some maintained that panelists include representatives of various stakeholders (potentially including those with financial interests at stake) and others held that panelists should be independent and clearly lacking any conflict of interest. EPA recognizes the importance of independent peer review, especially for new or controversial issues. EPA intends to follow contemporaneous Agency policies concerning peer review and selection of independent peer reviewers.

Effects on DNA

Initially, as with the 1986 guidelines, various versions of the revised guidelines focused on DNA effects and the results of mutagenicity and genotoxicity testing. Even though these remain an important consideration in carcinogenesis, the current guidelines explicitly recognize a greater range of MOAs, some of which may not have effects on DNA as the dominant consideration for risk assessment. These guidelines have also been designed to accommodate future advances in toxicology and gathering of biological information, e.g., genomics. The MOA approach allows differential consideration of promoting or initiating effects, indirect carcinogenic effects arising from effects on immune function or nutrition, and saturation kinetics and other dose dependencies that had been raised as concerns in previous drafts.

Dose Extrapolation

This section was expanded in the 2003 draft guidelines to include a discussion of linearity and nonlinearity for the dose-response function as determined by the MOA. Concerns included the basic premise that carcinogens that act by a mutagenic mode of action are appropriately modeled using a linear.

The Agency, along with other Federal regulatory agencies, continues to support the use of a linear extrapolation as a risk-adverse default for carcinogens when there are insufficient data to determine the MOA(s). The current guidelines allow for a non-linear extrapolation, when the data justify such a procedure.

Precursor Events

Several commentors agreed with the definition of precursor events and the use of these events in evaluating MOA. Concerns were raised about assessing subtle pre-neoplastic precursor events in the same way as overt tumor responses. Other commentors expressed concerns about selection of appropriate precursors for MOA. Several commentors also thought that precursor events that are not clearly adverse should not be considered key events in hazard and dose-response analysis.

The guidelines are clear about the relationship between a precursor and the associated tumor (emphasis added): "For data on 'precursor steps' to be useful in informing the dose-response curve for tumor induction below the level of observation, it is often useful for data to 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 necessary 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." The use of precursor events in cancer risk assessment is an evolving process. Precursors were used both to demonstrate the MOA and to determine the regulatory toxicity

value in the IRIS assessments of chloroform that can be found through the IRIS website at http://www.epa.gov/iris/index.html.

Judging Data

The importance of determining the human relevance of animal data was reiterated in several sections throughout the comments. Rather than having a checklist for judging data, commentors suggested that a weight-of-evidence approach be used to arrive at a "reasoned judgment," using all data of acceptable quality and including both human and animal data, with human relevance of animal data being assessed. Commentors expressed concerns that these guidelines for judging data have raised the "bar" or level of evidence for deviation from the linear default low-dose assumption acceptance to a level that will be impossible or very difficult to achieve. Commentors also requested that EPA present case studies demonstrating the application of the criteria presented for judging data.

Given the growing complexity and amount of data that may be available for evaluating the MOA(s), it is not reasonable to attempt to speculate on all possible data sets, the potential quality of each of the possible types of data, and the possible mutually supportive or conflicting effects of the data base when judged as a whole.

Weight of Evidence (WOE) Descriptors

Issue: WOE Narrative Approach.

The proposed approach that uses both a descriptor and narrative to characterize carcinogenicity by route was generally considered to be an improvement over the previous alpha-numeric system. Commentors thought the descriptors provided a better understanding of the placement of an agent, making it clear that the placement is based on weight of evidence (WOE) for carcinogenicity, and allowing a simpler explanation for risk communication. However, some commentors felt that descriptors perpetuate a black-and-white distinction between carcinogens and noncarcinogens. As with the section on MOA, a number of the comments requested examples, definitions of qualifying adjectives (e.g., "conclusive," "convincing," "compelling," "extensive," "marginal," or "suggestive") or definitive criteria for the various descriptors.

EPA's guidelines utilize five 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." 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."

In 1996 when EPA first proposed revisions to the cancer guidelines, three descriptors were proposed: "known/likely," "cannot be determined," and "not likely." Many commentors felt that these descriptors were too broad to be meaningful. Among the suggested changes was to split "known" from "likely" and to create another descriptor "suggestive" to separate it from "cannot be determined." Every set of descriptors that has been proposed over the years, has had its supporters and detractors. Some prefer more descriptors in an attempt to capture within the descriptor the underlying basis for the choice of the descriptor, while others prefer fewer. The Agency has been applying these descriptors in its assessments during the interim period while the guidelines were being finalized.

Having examined the various comments on these assessments and other referenced systems, the Agency has determined that the current descriptors, with all of the attendant qualifications and requirement for an accompanying narrative, are appropriately useful. Moreover, the narrative that, according to the guidelines, should not be separated from the descriptor, will contain the types of information that were suggested by some for inclusion in the descriptors themselves. The guidelines text has been revised to emphasize these points: "Choosing 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. These descriptors and 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. Descriptors represent points along a continuum of evidence; consequently, there are gradations and borderline cases that are clarified by the full narrative. Descriptors represent points along a continuum of evidence; consequently, there are gradations and borderline cases that are clarified by the full narrative. Descriptors, as well as an introductory paragraph, are a short summary of the complete narrative that preserves the complexity that is an essential part of the hazard characterization."

Issue: Policy implications and risk communication of WOE.

Commentors requested that EPA recognize that the descriptors influence public behavior and decisions about use of a chemical. Some of the descriptors could pose legal and communication issues. EPA acknowledges that the descriptors may be misused in many ways, especially without their complementary narrative. If used correctly, however, the WOE procedure and the resulting narrative and descriptor, afford an appropriate mechanism for providing risk managers and stakeholders with a summary of the data, results, and analyses that form the basis for EPA's conclusions. The guidelines emphasize the importance of not separating the descriptor from the narrative (emphasis original): "**Users of these guidelines and of the risk assessments that result from the use of these guidelines should consider the entire range of information included in the narrative rather than focusing simply on the descriptor.**" Likewise, the guidelines state: "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 of whether the chemical is carcinogenic. 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."

Issue: Route Specificity.

There was general support for the route-specific approach to characterizing carcinogenic hazard, although some comments indicated that qualifiers should be added to the descriptors so that cancer by another route is not qualitatively ruled out. On the other hand, use of multiple descriptors were thought to be confusing to the public when used for other reasons, such dose-dependent effects. EPA believes that the current WOE procedure is sufficiently flexible that these issues can be resolved on a case-by-case basis.

Issue: Comparison of EPA's Descriptors with IARC's Categories.

Additional information was requested for inclusion concerning the relationship between EPA's weight of evidence (WOE) descriptors and IARC's categories for evaluating carcinogens. Two of the frequently asked questions are: (1) how does the IARC classification scheme compare to the EPA cancer guidelines descriptors and (2) why are there differences? Although both of these questions seem like they should have a simple answer, they do not.

Through the mid-1980's, the majority of carcinogens were evaluated based on studies of tumor incidence in humans or laboratory animals with supplementary information from *in vivo* or *in vitro* mutagenicity studies. Within this framework, it was relatively simple to devise a basic checklist of possible results, e.g., the table in EPA's 1986 cancer guidelines, although even those guidelines allowed for adjustments within this table (either up or down) based on other data.

Since then, more types of data pertaining to an understanding of the processes of carcinogenesis have become available including (1) toxicokinetics to predict the target-organ dose of the proximate carcinogen and (2) non-mutagenic modes of carcinogenesis. Prospects for additional, useful information are likely, e.g., from genomics or proteomics. Thus, the weight-of-evidence data reviewed have changed from a relatively consistent data set with a limited number of results to a highly variable data base with a large number of results that can either form a consistent story or may present conflicting or incomplete information. To structure such data sets, while allowing for improvements in understanding of carcinogenesis in the future, the 2005 guidelines evaluate the total weight of the evidence that can be informed by evaluation through a mode of action (MOA) framework. The MOA/WOE approach allows consideration of all available, relevant, quality data. It also allows each result to build upon the previous information in a manner to highlight consistencies or contradictions within the biological hypothesis. This approach, however, also restricts the ability to determine the data that are necessary to use a weight-of-evidence descriptor, as the matrix of potential results is too large to consider all possible combinations *a priori*. The complexity of the decision is one of the reasons that the

guidelines emphasize that the narrative that is a summary paragraph, not just a descriptor, is necessary to characterize fully the weight of the evidence decision. The use of a different framework for decisions as well as inclusion of additional types of data make a one-to-one correspondence with IARC's process problematic, especially given the differences noted below.

The two processes retain many similarities. For example, the documentation at the beginning of IARC's CD of their evaluations clearly states that the weight-of-evidence descriptors are neither probabilities nor should they be applied outside the context and purpose for which they were developed. Similarly, EPA's guidelines state, "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 of whether the chemical is carcinogenic. 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." EPA's guidelines also specifically discuss the use of descriptors within the context of broader narrative and caution users of the guidelines to consider the entire range of information included in the narrative. The overall similarities can also be observed in this table of EPA's descriptors and IARC's categories.

EPA Descriptors	**IARC Categories**
Carcinogenic to humans	Carcinogenic to humans (Group 1)
Likely to be carcinogenic to humans	Probably carcinogenic to humans (Group 2A)
Suggestive evidence of carcinogenic potential	Possibly carcinogenic to humans (Group 2B)
Inadequate information to assess carcinogenic potential	Not classifiable as to its carcinogenicity to humans (Group 3)
Not likely to be carcinogenic to humans	Probably not carcinogenic to humans (Group 4)

Both systems have five delineations. EPA's descriptors and IARC's categories may have roughly the same meaning to a layman. The slight differences in language may alert the risk assessor and risk manager to the differences that exist in the two methods (described below). The top descriptor/category of both is the same, and the greatest similarity for evaluation is in this evaluation. Both organizations use mode of action (MOA) information in their final evaluation of the relevance of animal data to human carcinogenicity (see discussion below).

Differences exist in the evaluation processes. IARC use a two-step process to determine the WOE for carcinogenesis (see Cogliano et al., 2004[1]). It first evaluates separately the WOE

[1] Cogliano, VJ; Baan, RA;Straif, K; Grosse, Y; Secretan, MB; El Ghissassi, F; Kleihues, P. (2004). The Science and Practice of Carcinogen Identification and Evaluation. Environ Health Perspect 12:1269-1274.

for human carcinogenicity and for animal carcinogenicity as sufficient evidence, limited evidence, inadequate evidence, or evidence suggesting lack of carcinogenicity. These decisions can be made by separate expert panels. In the second step, these partial evaluations are combined in a rather formulaic procedure to derive a preliminary default evaluation using one of the descriptors in the table above. Only after this preliminary evaluation is determined are the "other data relevant to an evaluation of carcinogenicity and its mechanisms" considered. These additional data may modify the default evaluation up or down. In contrast, EPA performs one integrative analysis to examine what the biological data reveal as a whole about carcinogenic effects and mode of action of the agent, and their implications for human hazard and dose-response evaluation.

In addition to differences in how some data sets might be described by EPA and IARC, two fundamental differences exist in the overall use of the descriptors/categories. First, EPA allows different descriptors for different doses, routes of exposure, tumor sites, etc., e.g., hexavalent chromium by inhalation versus ingestion. IARC only uses one category per chemical regardless of such factors. Thus, while EPA can determine that chloroform is only carcinogenic at "high" doses that are toxic to cells, IARC must use one category for all levels of exposure. Second, the IARC Preamble does not indicate how negative studies will be used in the evaluation of chemicals. EPA's guidelines distinguish between conflicting data (positive and negative results in the same test system) and differing data (positive and negative results in different test systems) and give some indication of how they might be used in the WOE evaluation.

These differences in procedures may have consequences for the choice of descriptor or category.

- For IARC, any positive study, e.g., one positive animal study, regardless of mode of action may result in a classification of at least "not classifiable." In contrast, if the mode of action is not relevant to human carcinogenesis, e.g., alpha-2-mu-globulin in aged rat kidneys, EPA would describe it as "not likely."

- IARC generally requires negative evidence in both human and animal studies to determine that a chemical is "not likely" to be carcinogenic. If negative evidence is only available from animal studies, the chemical would most likely be classified as "not classifiable." In contrast, EPA might describe a chemical as "not likely" if there were two or more negative, high quality studies in two species of animals.

Dose-Response Assessment

In the 2003 draft and the current guidelines, MOA information is used to determine whether a linear or nonlinear approach to extrapolation is implemented. As this procedure allows consideration of various nonlinear dose-response curves, comments on the previous drafts that requested such consideration are resolved.

Issue: Combining Tumors

Comments were received about combining tumors across multiple sites. The guidelines use the MOA framework to provide a biological basis for combining tumors: "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 a characterization of the 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)."

Similarly, some comments were received on combining benign and malignant tumors. EPA believes the procedures in the guidelines provide a reasonable biological and statistical method for the evaluation of multiple tumor sites, absent data that allows and justifies a chemical-specific procedure. As noted in the guidelines, "The incidence of benign and malignant lesions of the same cell type, usually within a single tissue or organ, are considered separately but may be combined *when scientifically defensible* ..." (emphasis added). Moreover, the MOA framework provides information, when data are available, upon which to base a decision of whether the tumors are likely to arise from dependent or independent processes.

Issue: Toxicokinetic Modeling/Analysis.

There was general support for EPA's encouragement for use of toxicokinetic modeling in the current guidelines. The same discussion of toxicokinetic modeling was presented in the 1996 and 1999 draft guidelines. The discussion in the current guidelines is not extensively changed, except that a requested section on inhalation models was added. A more thorough discussion of PBPK model accuracy and precision, structure error and sensitivity analyses, and methods for validation were requested. EPA believes that the guidelines are not the appropriate place for such analyses. Each model, existing or future, should undergo its own evaluation and validation.

Issue: Interspecies Scaling (Analysis of Dose)

Commentors requested more comprehensive explanations on how to apply dose analysis to subpopulations and how toxicokinetic/toxicodynamics should be taken into account when scaling across species. The empirical basis for adopting the oral scaling factor of body weight to the 3/4 power, as a default when data are lacking, is presented in U.S. EPA (1992)[2]. It was considered appropriate, as a matter of policy, for the agencies to agree on one factor. The use of a toxicokinetic model is encouraged when data become available. Toxicokinetics have been used

[2] U.S. EPA (U.S. Environmental Protection Agency). (1992) 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.

in some Agency analyses. The default for inhalation exposure is a model that is constructed to become better as more agent-specific data become available. The Agency continues to examine and evaluate alternative methods for interspecies scaling, and this may be the subject of future guidance.

Issue: Route Extrapolation.

The content of this section has been changed in the current guidelines to reflect differences in the methods for estimating RfDs and RfCs. This change addresses the major issues regarding route extrapolation (e.g., methods for quantitative extrapolation and further guidance on default assumptions) that were raised. As with interspecies extrapolation, the Agency continues to examine and evaluate alternative methods for route extrapolation, especially those that incorporate chemical-specific information.

Analysis in the Range of Observation and the point of departure (POD)

Issue: Analysis of Epidemiologic Studies

There were several comments on the epidemiologic section of the 2003 draft, including that the section discussed what the studies were but not how best to use them in risk assessment. Some commentors requested clarification of some of the text. The section on epidemiologic studies has been rearranged and some of the text rewritten to clarify the text and expand some of the topics alluded to in the draft. The guidelines also now incorporate suggestions for evaluating data from human exposures that were included in a recently released report by the Surgeon General.

Issue: Biologically-Based Models

While commentors on the drafts supported biologically-based models, they acknowledged general lack of data for development of the models. Commentors noted that as many parameters as possible should be measured directly; however, they stated that guidelines should allow the use of indirect lines of evidence pertaining to the dose-response curve or the existence of a threshold dose. Concern was expressed about the apparently stringent criteria for application of these models. Some commentors suggested the use of more generalized concepts and indirect lines of evidence.

The Agency believes that, when data and models are available, a biologically based model offers the best method for evaluating data. Such models can often highlight interspecies or dose-dependant differences that were the subject of many comments. It is expected that each model will be reviewed by an expert panel, and the issues appropriate for that model deliberated at that time. Furthermore, EPA has recently established a Council for Regulatory Environmental Modeling (CREM), and this body will develop Agency guidance for validation of models.

Issue: Empirical Modeling ("Curve Fitting")

The 2003 draft provides an expanded discussion of modeling incidence data. However, the comments for the three drafts generally raise the same issues: the choice of dose-response model; the choice of the lower bound on the dose versus the dose, i.e., LED_{10} versus ED_{10}; choice of various response levels (1%, 10%); and the method for incorporating key events with tumor incidence data.

With regard to choice of model, the guidelines state, "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[3]). 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. 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. However, in situations where there are alternative models with significant biological support, the decisionmaker can be informed by the presentation of these alternatives along with their strengths and uncertainties." This approach allows sufficient flexibility when data are available to make a biologically based choice, while providing a default method when such data are lacking.

It was noted that the shape of the curve for tumor incidence data and key precursor data may differ over the range of observations. A clear example of how to employ key events, in conjunction with tumors, in plotting the dose-response curve was requested.

EPA believes the qualitative data on precursor causality can best be established by using the MOA framework. Quantitative analysis of precursor data, as it relates to the carcinogenicity dose-response curve, will need to be examined on a case-by-case basis. When sufficient data and examples exist to consider a generalized position that would be expected to be specific to a group of chemicals, e.g., based on mechanistic or MOA data, supplemental guidance to these guidelines may be considered.

---

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

Further guidance was requested on the level of statistical rigor needed for validation of non-default mathematical models. This commentor also noted that discussion of standard models, including types of data sets they are designed to fit, and their relation to underlying modes of action, would be helpful. EPA's technical guidance documents for its benchmark dose procedures address these and other technical matters regarding dose-response curves.

Issue: Point of Departure (POD)

In the 2003 version, discussion of the POD was expanded into a separate section that provided additional information regarding POD and its narrative. In general, the comments focused on clarification of evaluation of the carcinogenic process including the following issues: determining causality and adversity, selection of the "key events," selection of the response level for the POD, and the use of precursor events to evaluate dose-response relationships at lower exposure levels. Some stated that they thought that the 2003 guidelines provided less confusing and diverse definitions of "key events" than did the 1999 guidelines. Many commentors thought that more guidance was needed for determination of a POD. Concerns were raised that the modeling might not be anchored to biology. More guidance was requested on the selection of a POD from the benchmark dose modeling including: (1) selection of a POD and evaluation of the degree of confidence in a benchmark below the ED_{10} (e.g., ratios of ED_{01}:LED_{01} and ED_{01}:lowest dose tested), (2) incorporation of a more complete spectrum of the data (e.g., Monte Carlo simulation), and (3) the use of uncertainty factors with PODs based on precursors.

This section has been rewritten to make it clear that the POD is selected based on biological, as well as statistical, considerations based on the data available. The illustrative example of an LED_{10} was based on experience with observed results from cancer bioassays due to their design. Response levels other than 10% have and can continue to be used in Agency analyses. The Agency has and will continue to estimate and present all PODs that are appropriate and useful for the analyses. A definition of a 'key event" is presented in the guidelines: "A *'key event'* is an empirically observable precursor step that is itself a necessary element of the mode of action or is a biologically based marker for such an element." Thus, the requested biological considerations are already in the definition. Current Agency practice requires justification of the uncertainty factors that are applied in any particular analysis. Use of uncertainty factors for PODs based on precursors would also be justified in the analysis.

Finally, these guidelines address uncertainty in the POD by incorporating EPA's data quality guidelines that recommend, where practicable, a central estimate as well as an upper and a lower bound on that estimate be presented. Specifically, the guidelines state, "To the extent practicable, such assessments should provide central estimates of potential risks in conjunction with lower and upper bounds (e.g., confidence limits) and a clear statement of the uncertainty associated with these estimates."

Low-Dose Extrapolation

Issue: Linear and Nonlinear Low-Dose Extrapolation

This section was significantly revised in each of the drafts. Both the reorganization and the comments received in response to the 1996, 1999, and 2003 drafts reflect an increasing level of sophistication that is due, in part, to considerable advances in understanding of cancer biology, and recognition of the limitations in current approaches to modeling biological processes involved in carcinogenesis, as well as efforts to harmonize approaches to cancer and noncancer risk assessment. The concept of weight-of-evidence in evaluating mutagenicity information recognizes that not all mutagenic and genotoxic data are equal in evaluating potential human carcinogenic response. Assessment of human relevance is dependent on the type of assays and test systems employed to evaluate these end points as well as the dose levels used. Nevertheless, some commentors thought that low-dose linear extrapolation should always be used for such chemicals in order to be public health protective.

Most comments expressed support for the use of all available relevant scientific information for low-dose extrapolation, including biologically-based or case-specific modeling for estimating cancer risk when extrapolating below the observed range, as well as using a nonlinear approach when data are sufficient to support it.

In the 1996 draft, the margin-of-exposure (MOE) approach was briefly discussed and commentors requested that the discussion be expanded. In the 1999 draft, the MOE discussion was expanded and further clarification was requested by many commentors. In the 2003 draft, the MOE approach was mentioned but not discussed; several commentors requested that a discussion of this approach be returned to the guidelines. Additional guidance was also requested on how to determine an acceptable MOE and what types of information warrant the use of both a linear and nonlinear approach for describing dose response. Other commentors thought that the MOE methodology relaxes existing health-protective procedures and does not consider background exposure risks and cumulative exposures over time. Clarification was requested of EPA's discussion regarding development of oral reference dose or inhalation reference concentrations for tumors arising from a nonlinear MOA *"in accordance with EPA's established practice of developing such values."*

The text of the guidelines has been changed so that the nonlinear low-dose extrapolation approach for cancer effects are to be estimated in a manner similar to that used to develop reference doses or reference concentrations for noncancer effects. EPA has published guidance for estimating reference doses and concentrations (U. S. EPA, 1994[4]). An MOE approach is not precluded, but discussion of the MOE has been moved to the section on risk characterization.

---

[4] U.S. EPA (U.S. Environmental Protection Agency). (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.

This was done because most of the comments on the MOE concerned what margin was considered adequate. That issue is a risk management decision, and therefore the chemical-specific issues should be appropriately characterized to facilitate the decision-making process. The guidelines state: "With nonlinear extrapolations, the method of risk assessment depends on the procedure used. If a nonlinear dose-response function has been determined, ... 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 ... can be expressed as a *margin of exposure* (MOE), defined as the ratio of the POD over an exposure estimate (MOE = POD / Exposure)..."

Comments on nonlinear extrapolation included concerns that the amount of mechanistic information and level of evidence required for use of this approach would be too high. A request was made that the guidelines provide clear criteria for when a MOA can be considered to have a threshold. Concern was also expressed that underestimation of cancer risk may result from implementation of the low-dose nonlinear extrapolation approach.

These guidelines recognize that the nonlinear extrapolation methods have limitations. They state, "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 evaluation of some 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 cancer guidelines."

The MOA framework is the method used in these guidelines to determine if a nonlinear extrapolation procedure is appropriate. Current Agency analyses provide examples of the types of data that demonstrate that a nonlinear extrapolation should be used. An MOE approach could be used, if it is determined to be useful. As nonlinear extrapolations are a relative new procedure, the process is expected to evolve with experience.

Numerous comments were also received on "Choosing an Extrapolation Approach." Several commentors thought that using high-dose data, i.e., data in the range of experimental observation, to extrapolate to lower doses was inherently flawed because human cancer epidemiology has identified eight human cancer/risk factors, and there is no evidence environmental exposures to trace levels of chemicals contribute to human cancer rates/risks. Further, very high doses used in laboratory studies can themselves cause mutagenic/carcinogenic effects.

If data exists from exposures within the range of interest for a risk assessment, they are expected to be included in the analysis. Absent data to the contrary, however, the Agency will assume that results in laboratory animals can be used to predict potential results in humans.

More guidance was requested on how to resolve conflicting data sets on mutagenicity and suggestions were made that a weight-of-evidence approach to evaluation of mutagenicity be used, with the text expanded to discuss the complexity of these issues more comprehensively.

The guidelines use WOE for defining MOA for carcinogenicity, one of which is mutagenicity. The issue of determining a mutagenic MOA is also more fully discussed in the *Supplemental Guidance Assessing Susceptibility for Early-Life Exposure to Carcinogens*. The guidelines also define the difference between conflicting and differing data; these definitions may resolve some of the apparent difficulties of the complex data sets. The WOE narrative will justify the conclusions.

Extrapolation to Different Human Exposure Scenarios

Issue: Dose Metrics and Exposure Duration

The 2003 draft addressed some of the concerns about earlier drafts. Comments about selection of the appropriate averaging periods and for different windows of susceptibility to carcinogens, such as those occurring intermittently or during early-life stages, were noted as being important but little guidance had been given on how to quantitatively incorporate these variables into exposure assessments. Several suggested modeling approaches to quantitatively deal with less-than-lifetime exposure scenarios.

The Agency believes that the averaging period for lifestages should be assessed on a case-by-case basis. The available data may also limit the ability to select an averaging period. Clearly, information on the MOA will inform these decisions and may provide a method for a group of chemicals. Indeed, as presented for chemicals with a mutagenic MOA for carcinogenesis are the subject of the Supplemental Guidance. The Supplemental Guidance also provides examples of quantitative analyses for both lifetime and less-than-lifetime exposures to these carcinogens.

Commentors thought that risk may be overestimated when LADD was used as dose measure, because this metric assumes that a high dose of an agent is equivalent to a low dose spread over a lifetime. According to these commentors, several papers have shown that low-dose risk estimates reached by the use of LADDs may overestimate true risk by several orders of magnitude but do not underestimate true risk. In contrast, commentors provided some evidence indicating that, for nonmutagenic carcinogens with an unknown mode of action, the use of LADD would underestimate risks for children.

While the Agency understands the discussion that has occurred regarding the LADD and averaging times, limitations of the data often preclude a more sophisticated analysis. When data

are available, e.g., for early-life exposure to carcinogens with a mutagenic MOA, the LADD may be superceded, as with the Supplemental Guidance. Similarly, data necessary for time-to-tumor models are not always available. Time-to-tumor data have been used by EPA for some estimates of cancer potency. Similarly, when data demonstrate the utility and accuracy of alternative dose metrics, they have been used, e.g., body burden is used as the dose metric for PCBs.

Extrapolation to Susceptible Populations and Lifestages

Issue: Susceptible lifestages

Over the years, numerous comments were made on whether children and the elderly are more susceptible to carcinogens, and on how to adjust risk estimates to account for such potential susceptibilities. EPA has developed a separate document entitled *Supplemental Guidance Assessing Susceptibility for Early-Life Exposure to Carcinogens*, in which this susceptible lifestages is discussed and adjustment factors for potency estimates for different early lifestages are proposed for carcinogens that act through a mutagenic MOA, if potency estimates are otherwise based on adult data. Comments received in response to publication of this Supplemental Guidance document were numerous and are presented in a separate document.

Risk Characterization

Issue: Risk characterization narrative

Most comments reflect changes over the years in the development of an acceptable risk characterization summary, e.g., the requirement for reporting upper and lower bound estimates with central tendency, and reporting of risk estimates under alternative assumptions. Other comments are similar, specifically a better characterization of variability and uncertainty, consideration of the intended audience in terms of how the risk characterization summary is written (e.g., "nontechnical versus somewhat technical"), and consideration of "equally plausible" alternatives. The use of quantitative techniques such as probabilistic approaches was recommended to characterize the plausibility of different risk estimates, reflecting advances in the development and acceptability of these methods by risk assessors, risk managers, and others.

The guidelines explicitly state that both upper and lower bounds, as well as central tendency estimates, should be presented where practicable and applicable in accordance with EPA's Information Quality Guidelines. The guidelines also recommend consideration of probabilistic techniques for risk assessment and characterization, including but not limited to expert elicitation. The text of the guidelines has been revised to reflect these concerns and now reads: "*Advances in uncertainty analysis* are expected as the field develops. The cancer guidelines are intended to be flexible enough to incorporate additional approaches for characterizing uncertainty that have less commonly been used by regulatory agencies. In all scientific and engineering fields, data and research limitations often limit the application of established methods. A dearth of data is a particular problem when quantifying the *probability*

distribution of model outputs. In many of these scientific and engineering disciplines, researchers have used rigorous expert elicitation methods to overcome the lack of peer-reviewed methods and data. Although expert elicitation has not been widely used in environmental risk assessment, several studies have applied this methodology as a tool for understanding quantitative risk. For example, expert elicitation has been used in chemical risk assessment and its associated uncertainty (e.g., Richmond, 1981; Renn, 1999; Florig et al., 2001; Morgan et al., 2001; Willis et al., 2004), components of risk assessment such as hazard assessment and dose-response evaluation (e.g., Hawkins and Graham 1988; Jelovsek et al., 1990; Evans et al., 1994; IEc, 2004; U.S. EPA 2004) and exposure assessment (e.g., Whitfield and Wallsten, 1989; Hawkins and Evans, 1989; Winkler et al., 1995; Stiber et al., 1999; Walker et al., 2001, 2003; Van Der Fels-Klerx et al., 2002), and for evaluating other types of risks (e.g., North and Merkhofer, 1976; Fos and McLin, 1990). These cancer guidelines are flexible enough to accommodate the use of expert elicitation to characterize cancer risks, as a complement to the methods presented in the cancer guidelines. According to NRC (NRC, 2002), the rigorous use of expert elicitation for the analyses of risks is considered to be quality science."

The Risk Characterization section of these guidelines further state, "While it is an appropriate aim to assure protection of health and the environment in the face of scientific uncertainty, common sense, reasonable applications of assumptions and policy, and transparency are essential to avoid unrealistically high estimates. It is also important to inform risk managers of the final distribution of risk estimates (U.S. EPA, 2000b; 1995). Otherwise, risk management decisions may be made on varying levels of conservatism, leading to misplaced risk priorities and potentially higher overall risks. (Nichols and Zeckhauser,1986; Zeckhauser and Viscusi,1990).

"The risk characterization presents 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, 1995, 2000b). Choices made about using data or default options 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. In situations where there are alternative approaches for a risk assessment that have significant biological support, the decisionmaker can be informed by the presentation of these alternatives along with their strengths and uncertainties."