EPA/630/R-03/001
August 7, 2002

Summary Report of the Technical Peer Review Workshop on the Draft Document Entitled *A Review of the Reference Dose and Reference Concentration Processes*

U.S. Environmental Protection Agency
Washington, DC
June 19, 2002

Risk Assessment Forum
U.S. Environmental Protection Agency
Washington, DC 20460

NOTICE

This document has been reviewed in accordance with U.S. Environmental Protection Agency (EPA) policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

This report was prepared by Versar, Inc., an EPA contractor (Contract No. 68-C-99-238, Task Order No. 71), as a summary of the discussion held at the Technical Peer Review Workshop On the Draft Document Entitled *A Review of the Reference Dose and Reference Concentration Processes* (June 19, 2002). This report captures the main points and highlights of the meeting. It is not a complete record of all details discussed, nor does it embellish, interpret, or enlarge upon matters that were incomplete or unclear. Statements represent the individual views of each workshop participant.

FOREWORD

EPA's Technical Panel should be congratulated for *A Review of the Reference Dose and Reference Concentration Processes*, which provides a major first step forward and recommendations for improvements in the processes used to derive reference values. The peer reviewers commend EPA for incorporating new studies and approaches into this text, which will enhance the scientific basis for this important EPA program. The criticisms and comments provided by the thirteen peer reviewers are intended to assist EPA with additional improvements to the document. EPA is encouraged to consider the reviewers' recommendations, re-assess those recommendations with which Panel members dissented, as well as those that require additional clarification and evaluation, and to pursue future directions to further the reference value processes.

EPA's commitment to improve its risk assessment methods using the best available science, which is widely heralded as being a major goal of the Agency, is both articulated and exemplified in the draft document. This commitment is important because the risk assessment methods adopted by the Agency have a profound influence on the approaches used elsewhere in this country and abroad. As a result, consideration should be given to the wider impact of the changes contemplated in this report. As EPA moves further into improving the reference value process, review and involvement by additional outside groups would be useful. For example, the reviewers endorse a recommendation for involvement of the NAS/NRC in developing some of the ideas, particularly to evaluate toxicity testing approaches and strategies, based on this and other reports. In addition, once EPA has refined and clarified its recommendations for changes in the processes, it might be useful to have those reviewed by NAS/NRC. This type of consultation enhances the scientific rigor and credibility of the Agency's risk assessment processes. Moreover, dialogue with groups such as IPCS, OECD, IARC, ATSDR, NIOSH, as well as other stakeholders, should be encouraged in order to enhance the scientific credibility of these processes and obtain input from groups that are likely to be impacted by these changes.

The enthusiasm with which the peer reviewers offer constructive comments reflects our interest in improving risk assessment science, which is the foundation for EPA's programs. Considerable feedback, both positive and negative, was provided by the peer reviewers during the workshop. This report summarizes the reviewers' criticisms, suggestions, and recommendations, which are intended to help EPA with improving the overall reference value process. Further development of these new approaches will help EPA to achieve the goal of protecting public health and the environment.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The Technical Peer Review Workshop On the Draft Document Entitled *A Review of the Reference Dose and Reference Concentration Processes* was held on June 19, 2002, in Arlington, VA. Versar, Inc., a contractor to EPA's Risk Assessment Forum, organized the workshop and assembled 13 peer reviewers to provide comments on the document. The reviewers provided suggestions and recommendations in response to a series of eight charge questions prepared by EPA in advance of the meeting that addressed issues involved with deriving reference doses (RfDs) and reference concentrations (RfCs).

The meeting was opened by Versar staff, after which the Chair briefly discussed the process for the meeting. After an opening presentation by EPA staff providing background information on the draft document, the reviewers engaged in discussions on the eight charge questions. Recommendations and comments were provided by the reviewers on the document. This session continued through the majority of the day, with two time periods set aside for observer comment. The meeting concluded with a brief session where reviewers raised issues on other topics of interest related to the document.

Most, but not all, reviewers submitted premeeting written reviews in which they commented on the overall quality of the document as well as responded to the eight charge questions prepared by EPA prior to the meeting. The premeeting comments were distributed among the reviewers in advance of the meeting to facilitate discussion at the workshop. Those reviewers who did not prepare premeeting comments were asked to submit written comments following the meeting, so their views would be represented in the appendix to this report. Also, reviewers who did prepare premeeting comments were given the opportunity to amend their written comments following the meeting, if they so desired. This workshop report focuses on the comments, suggestions, and recommendations made by the reviewers during the workshop.

Chair's Summary

EPA's text was well thought-out and presented. EPA's arguments and recommendations were logical and for the most part convincing. The reviewers were grateful that EPA had incorporated new studies and theory into this text. The work of many experts within EPA was evident in its deliberative and insightful text.

Nonetheless, reviewers had a number of suggestions for improvement, many of which were similar to, or gained support from, other reviewers. Reviewers disagreed with EPA and with each other in a few areas. Both the suggested improvements and areas of disagreement are shown below in summary, with a more detailed accounting to follow in the body of the workshop summary report, particularly in Section 3.

The Chair thanks Dr. Carole Kimmel and her EPA colleagues and Mr. David Bottimore and his Versar staff for the opportunity to review such a fine piece of work with such eminent reviewers.

Please comment on the recommendation to derive less-than-lifetime reference values in addition to the chronic RfD and RfC? Is the need for these values adequately justified in the report? Most reviewers thought that EPA's derivation of less-than-lifetime reference values was valuable, but that more history, guidance and justification were needed. In addition, some reviewers expressed concern over implementation of these values, pointing out that since the process would be resource intensive, it would likely increase the time to complete IRIS documents. Other reviewers disagreed with this latter concern and stated that if EPA considered this a priority, it would devote the necessary resources to IRIS. Several reviewers also expressed concern over whether enough data existed to derive less-than-lifetime values, resulting in the generation of values for many chemicals with less supporting data. Reviewers also discussed the utility of extrapolating from one route of exposure with robust data to another route, and encouraged EPA to continue to use caution with such extrapolations (e.g., EPA's RfC methods have a good discussion of such caution) in order to avoid results that may not make biological sense. Developmental toxicity was recognized as an effect that should be considered with each less-than-lifetime reference value. Moreover, latent effects from low doses at critical developmental time periods are also of concern, so that the estimation of less-than-lifetime values needs to carefully consider other than high doses. Reviewers recognized that one purpose of less-than-lifetime values was practical, (e.g., clean up at waste sites may be less costly because higher concentrations are often acceptable in less-than-lifetime exposures). In other cases, a less-than-lifetime value might be more appropriate than a chronic value, used inappropriately, for example when infrequent exposure is given a daily average and compared to a chronic reference value.

Please comment on the revised definitions for reference values. In general, reviewers agreed that the definitions for reference values need further work, particularly with respect to terminology, notation, and presentation of qualitative and quantitative uncertainty information. For example, many reviewers expressed concern with the phrase "…intended to provide an estimate centered within an order of magnitude…" and favored a more qualitative statement. One reviewer suggested that EPA propose a quantitative definition for RfD. Some reviewers thought that the word "acute" could be confusing, and suggested that labels such as 15-minute or 1-hour exposures be used instead. Several reviewers thought that the subscripts were too confusing and that word modifiers, such as "short term" RfD be used instead. One reviewer noted that the development of RfVs for four durations is just as arbitrary as the three commonly used, and EPA should be concerned over the potential added burden on testing. Several reviewers supported a suggestion to have one definition and qualify the development of any RfV with a particular time duration. Reviewers in general liked the harmonization of cancer and non-cancer risk assessment methods, but pointed out that EPA needs to clarify its goal for harmonization. For example, if cancer is an endpoint of concern, then EPA's recommended definitions may not be adequate. Moreover, if a latent effect from a less-than-lifetime exposure is expected, then the definitions

may again fall short. Reviewers agreed with the use of the term "adverse" instead of "deleterious" in each of the definitions.

Please comment on the recommendation that endpoint-specific reference values should not be derived. Reviewers agreed that endpoint-specific reference values should not be developed, and encouraged EPA to continue, and to enhance, the development of narrative descriptions that explains the endpoint on which the RfV is based, other endpoints that were considered, areas that were covered by testing, and any remaining deficiencies found on EPA's IRIS summaries and in its supporting documents. Reviewers acknowledged that such text is also planned for the confidence statements associated with the various RfVs. Reviewers also agreed that the intent of the RfV is to protect the most sensitive or critical, endpoints, and these values will cover anything less sensitive (including developmental toxicity). EPA was asked to resolve potential differences with its chemical mixtures RAF technical panel on this issue.

Please comment on the life-stage approach taken in this review. Overall, reviewers felt that EPA did a good job summarizing and discussing the different testing methods and the gaps that exist. However, reviewers suggested that the overall testing topic gain the full attention of another panel. For example, reviewers agreed that a request, which is broad in scope and time frame, be made to the NAS or another technical panel, to address testing issues such as latency, life stage, pharmacokinetics, and mode of action. Other reviewers suggested that the data of well-studied chemicals be evaluated to see how extra data affected the RfV wrought from fewer data and larger uncertainty factors. Reviewers also emphasized the importance of mode of action research and the anticipated advances in these techniques, and agreed with EPA that mode of action should be discussed as part of the harmonization of the risk assessment process between cancer and noncancer endpoints.

Please comment on the recommendation to include duration-adjustment for inhalation developmental toxicity as for other health endpoints. Reviewers agreed with EPA that duration adjustment is important and should be attempted. They also felt that the document should stress that, in addition to the impacts to the developing fetus and children, developmental effects may be latent and not obvious from most testing protocols. Another comment from several reviewers emphasized the importance of matching the exposure duration with the data used to derive the RfV, particularly for intermittent exposure scenarios (e.g., fish consumption, hot spot/peak releases).

Please comment on the recommendations in the report for using a weight-of-evidence approach for hazard characterization. Reviewers were generally supportive of using a weight-of-evidence approach for hazard characterization, because of the need to educate risk managers and to provide a transparent explanation of the quality of the data. Reviewers also acknowledged that EPA is already doing this, as evident in the more recent IRIS toxicological review documents, but also to a more limited extent in prior IRIS assessments. However, reviewers suggested that EPA's guidance emphasize a weight-of-evidence approach and leave room for scientific judgment.

Please comment on the recommendations for use of an exposure-response array and derivation of sample reference values to select the point of departure for the final reference value. In general, most of the reviewers liked a visual display of the data, because it gives a sense of the size of the database and of dose-response relationships. However, several concerns were expressed by reviewers including: the equal treatment of effects, the lack of explicit dose response curves, and a realization that the reader must have a minimal knowledge of toxicology to understand the graphics. Reviewers disagree over the usefulness of the suggested format of toxicity tables that showed the development of the RfV for each endpoint and the usefulness of the concept of critical effect. Other reviewers suggested the use of meta-analysis of all relevant toxicity studies.

EPA recommended limiting the total UF applied to a chronic reference value for any particular chemical to 3,000. Many reviewers expressed concern over this recommendation, and asked for clarification. Some reviewers thought it would be more meaningful to state that such situations trigger concern and caution on the part of the assessor as to whether or not the database is sufficient to derive a value. Several questions were posed as to why uncertainty factors were multiplied and not added, or some other variant such as multiplying by the square root. Reviewers suggested that EPA take a fresh look as to how uncertainties are combined since a better way might exist (e.g., perhaps using distributions). Also, EPA should continue its emphasis of providing, and perhaps enhancing, its narrative on choices of uncertainty factors as summarized on IRIS and found in IRIS background documentation.

EPA stated that a reduction of the intraspecies UF should be considered only if data are sufficient to support the conclusion that the data set is representative of the susceptible subpopulation(s). Several reviewers suggested that the recommendation clarify that it refers to the usual default uncertainty factor that accounts for interhuman variability. Some reviewers stated that the recommendation was too proscriptive and didn't allow for use of scientific judgment to modify the usual default factor of 10. One reviewer wondered how far EPA needed to go to protect individuals within sensitive sub-populations, suggesting that independent guidance be developed, or that EPA make a probabilistic statement with each RfV. The reviewers agreed with EPA that not much information on population variability exists and additional information is needed on variability in children as well as in adults and in the aging.

Sound scientific judgment should be used in the application of UFs to derive reference values which are applied to the value chosen for the POD (NOAEL, LOAEL, or BMDL) derived from the available database. All of the reviewers agreed with this recommendation. They also agreed that the collection of additional data would enhance our understanding of the various uncertainty covered by EPA's five UFs.

EPA considers that its current five uncertainty factors will be adequate in most cases to cover concerns and uncertainties about children's health risks. Reviewers were generally pleased that EPA had taken a systematic look to see how the FQPA factor relates to the existing uncertainty factors. Most reviewers agreed with EPA that dividing the POD by another 10x

factor was not a good idea, because the standard factors appear to cover the underlying scientific issues and provide an adequate margin of safety. Moreover, there was general agreement among the reviewers that the Agency should have the ability to judge this on a case-by-case basis, and the addition of another factor would make such judgment more difficult. Similar to what EPA states, the reviewers felt that developmental toxicity is a unique database deficiency at each RfV, and that generally more relevant data collection is needed. If developmental toxicity data are missing and considered relevant, then that gap needs to be filled, and EPA needs to address this data gap, as it already does (and is doing better), with its database uncertainty factor.

EPA recommends the discontinuance in use of the MF. Most reviewers agreed with EPA's recommendation to discontinue use of the modifying factor because of its infrequent use and inclusion in the other uncertainty factors.

EPA will develop its own guidance on Compound Specific Adjustment Factors (CSAF) based on some of the available methods (e.g., IPCS). Reviewers generally agreed with this recommendation and also agreed that criteria were needed on how much data are enough and what kind of data are needed to justify a CSAF. Several reviewers added that the IPCS report on methods for CSAF already contains criteria that EPA might review for its purposes.

Reviewers also raised other issues that they felt would improve the document, but because of time limitations, reviewers agreed to make these comments part of their post meeting submission (either as an addendum to their premeeting comments or as revised written comments). However, one issue, that of the use of an uncertainty factor for severity of effect, was discussed in part. One reviewer felt that such an uncertainty factor was needed because not all critical effects for different chemicals are of similar severity. An added benefit of using this approach is that it also promotes harmonization with the risk assessment of cancer endpoints, if such endpoints are considered as frankly toxic, or their precursors are considered as minimally toxic. Not all reviewers agreed with the suggestion to use an uncertainty factor for severity, as reference values are usually based on LOAELs or NOELs or precursor effects. The reviewers did not have sufficient time to more fully explore this suggestion, but individual reviewers were encouraged to express their thoughts in their individual written comments.

1.0 INTRODUCTION

1.1 Workshop Purpose

The Technical Peer Review Workshop on the Draft Document Entitled *A Review of the Reference Dose and Reference Concentration Processes* was held on June 19, 2002, in Arlington, VA. The workshop was sponsored by the U.S. Environmental Protection Agency (EPA) Risk Assessment Forum and was organized by Versar, Inc. The purpose of the meeting was to provide a scientific peer review of the draft document.

1.2 Workshop Participants

A group of 13 experts, from different disciplines and types of organizations, was assembled by Versar to peer review EPA's draft document entitled *A Review of the Reference Dose and Reference Concentration Processes.* Versar selected experts with experience in development and use of RfDs and RfCs from a variety of perspectives: academia, consulting, environmental groups, industry, and state government. In addition, the experts were selected such that the following topic areas would be covered: toxicology (including immunotoxicology and reproductive and developmental toxicology), pharmacokinetic modeling, dose-response procedures, dose-adjustment procedures, uncertainty factor issues, approaches to testing for human health assessment, setting less-than-lifetime reference values, susceptible populations, and related areas of importance to human health risk assessment. Versar identified more than 75 candidate reviewers, from which final peer reviewers were selected. The list of 13 peer reviewers is presented in Appendix A. In addition to these reviewers, about 45 observers attended the workshop. The list of observers is presented in Appendix B.

1.3 Charge Questions

A list of charge questions, presented in Appendix C, was prepared by EPA and distributed to the peer reviewers prior to the meeting to stimulate feedback on technical issues related to the draft document. These charge questions addressed eight topic areas of interest in EPA's document that include a series of recommendations to improve the RfD and RfC processes. The peer reviewers were asked to review the recommendations of EPA's Technical Panel and to provide comments regarding the scientific rationale for the recommendations. Each of the charge questions included the question, an expanded description of the issue, and EPA's recommendation. At the peer review workshop, the eight charge questions served as a guide for dialogue. Prior to the meeting, reviewers were asked to prepare premeeting written comments in response to the charge questions. Most, but not all reviewers submitted premeeting written reviews in which they commented on the overall quality of the document as well as responded to the eight charge questions. The premeeting comments were distributed among the reviewers in advance of the

meeting to facilitate discussion at the workshop. Those reviewers who did not prepare premeeting comments were asked to submit written comments following the meeting, so their views would be represented in the appendix to this report. Also, reviewers who did prepare premeeting comments were given the opportunity to amend their written comments following the meeting, if they so desired.

1.4 Agenda

The workshop agenda is presented in Appendix D. The meeting began with opening remarks from Versar, Inc., including an overview of the agenda for the two-day meeting. The Chair provided a review of the objectives and introduced the process and ground rules for the meeting. This was followed by a presentation from the Chair of the EPA Risk Assessment Forum's Reference Dose/Reference Concentration Technical Panel, who provided background of the draft document. The agenda was organized around the eight charge questions. After the presentations, most of the day was devoted to the Chair leading the reviewers through discussion in response to the eight charge questions. The end of the day included a discussion among reviewers of other issues of concern related to deriving reference values. Two time periods were set aside for observer comment.

1.5 Workshop Summary Report

This report summarizes the presentations and discussion from the peer review workshop. It should be noted that this peer review (and the report) does not represent consensus positions; the comments, criticisms, and suggestions are those of the individual reviewers. As a result, the summaries presented in the body of the report reflect the statements of individual reviewers (i.e., the text notes when several reviewers agreed on issues or when individuals had dissenting opinions). Following the main body of the report are appendices that provide handouts, materials used in presentations, and written comments from the peer reviewers. The remainder of the report is organized as follows:

- Section 2 of this report summarizes the opening presentations. Overheads used by the presenters are provided in Appendix E.
- Section 3 provides summaries of the comments and suggestions of the reviewers in response to the charge questions.
- Section 4 summarizes the observer comments and subsequent discussion.
- The appendices to this report present the handouts from the meeting (e.g., lists of peer reviewers and observers, agenda, charge questions, and presentation materials/slides) as well as written comments from the peer reviewers.

2.0 SUMMARY OF OPENING REMARKS

2.1 Welcome

David Bottimore of Versar, Inc., opened the meeting by welcoming participants and observers. He presented an overview of the agenda, introduced the participants, and described the goals and intended outcome of the workshop. During his opening remarks, he emphasized that the meeting objective was to promote dialogue among the reviewers and provide comments on technical issues associated with the draft document. Mr. Bottimore talked about the effort to assemble a group of peer reviewers with diverse backgrounds and expertise related to RfDs and RfCs. He also noted that in addition to the main discussion sessions, time would be set aside for observer comments. His opening remarks were concluded by introducing the peer reviewers, including the Chair, Michael Dourson.

2.2 Chair's Introduction

Michael Dourson, from Toxicology Excellence for Risk Assessment, was the Chair for the peer review workshop and served as facilitator. He started by describing the peer review process and setting the ground rules for the workshop. The goal of the meeting was restated to emphasize that technical input was sought from each participant, noting that there would be no attempt to achieve consensus through this meeting. Rather, the discussion should bring out the diverse perspectives of individual experts in the group. He informed the reviewers that he would keep the discussion focused and on schedule, because of the volume of material to cover during the one-day workshop. During his introduction and overview of the peer review process, he distinguished the roles of the peer reviewers from the observers (including EPA staff) and encouraged reviewers to focus their thoughts and comments on the charge questions, the background text, and suggestions from other peer reviewers. Finally, he mentioned post meeting activities to prepare a workshop report that summarizes the discussion and made the request that each participant's individual comments be represented by either their pre- or post-meeting comments, which will be appended to the workshop report.

2.3 Background on the Draft Document Entitled *A Review of the Reference Dose and Reference Concentration Processes*

Carole Kimmel, the Chair of the EPA Risk Assessment Forum's Reference Dose/Reference Concentration Technical Panel, provided background on the deliberations that had gone into preparing the draft document entitled *A Review of the Reference Dose and Reference Concentration Processes.* She introduced the need to revise the processes used to derive RfDs and RfCs, with particular focus on developmental toxic effects. This document is a review, rather

than guidance, and it addresses the need to incorporate new information and approaches. As an evolving process, there are new data and methods available to toxicologists that can improve the way RfDs and RfCs are derived. She stated that this review is the first step; additional work will follow such as case studies and other efforts that will lead to development of guidance to implement changes in the program. Two major themes are evident in the document and the recommendations proposed: harmonization of the cancer and noncancer health risk assessment processes and the use of this information to promote future improvements in animal testing protocols.

Dr. Kimmel's presentation included overviews of the major issues and considerations that have gone into the eight recommendations, which are the basis for the charge questions. She summarized each of the eight topic areas, providing the rationale for EPA's recommendations, and emphasizing the areas where EPA seeks the peer reviewers' input and suggestions. Her presentation set the stage for the peer reviewers to address the charge questions, one by one.

Following this presentation, questions were posed by the reviewers to help clarify EPA's intent and perspective. Many of these questions introduced topics that were discussed in more detail later in the workshop, such as: terminology/definitions, emphasizing developmental toxicity, derivation of endpoint-specific reference values, acute and short-term reference values, reversibility of effects, uncertainty factors, and the need to have more transparent presentation of the data base used to derive reference values.

3.0 SUMMARY OF COMMENTS AND RECOMMENDATIONS

3.1 Charge Question #1

Please comment on the recommendation to derive less-than-lifetime reference values in addition to the chronic RfD and RfC? Is the need for these values adequately justified in the report?

In general, most of the reviewers felt that the derivation of less-than-lifetime values was a good concept and would be a useful approach, providing many more tools to risk assessors. Most of the reviewers agreed that EPA needs to provide more guidance as to when and where less-than-lifetime reference values should be applied. The reviewers felt that justification for the use of less-than-lifetime values also needs to be made explicitly. The point also needs to be made that these values would not supplant other established values (e.g., AEGLs, AREs). One reviewer commented that some EPA offices (e.g., Superfund and Office of Water) have already developed less-than-lifetime values and this recommendation would promote consistency across programs. It was generally agreed that these values would be a worthy addition to IRIS and would make that database even more useful.

There was some concern among the reviewers, however, over implementation of the derivation of less-than-lifetime values. For example, some of the reviewers pointed out that the process would be time and resource intensive. As it stands, the amount of time to complete an IRIS document is quite lengthy (often 2 years) and adding more components to the process would only increase this length of time. Other reviewers disagreed and stated that EPA could devote the resources to IRIS to address this concern. However, there was agreement that the development of less-than-lifetime values is a step that needs to be taken, and a suggestion was made that it would be helpful to prioritize the chemicals for which values should be derived. In some cases, there may be more values needed for a certain chemical or the values may be needed sooner than for another chemical.

Several reviewers also expressed concern over whether or not there are enough data available on many chemicals to derive less-than-lifetime values. One reviewer noticed that a general assumption in the document seems to be that a sufficient amount of data are available to support the derivation of these values, which may not be the case. As a result, some reviewers were concerned that this recommendation would result in the generation of more values with less data to support them. It was noted, however, that often data are available but they are not accessible and transparent. Reviewers also discussed the utility of extrapolating across species or routes of exposure for deriving less-than-lifetime values, because a robust database might exist for one route but not for the other. Reviewers noted that caution should be used in extrapolation that produces results which may not make sense or which are not consistent.

One reviewer expressed concern over the fact that the recommendation for less-than-lifetime values did not address the possible greater uncertainty with shorter-term or acute values. This

reviewer stated that the error for exceeding the RfV becomes larger because the dose-response curve is generally steeper for acute doses. Other reviewers disagreed with this statement because the dose response curve may not actually be steeper with acute doses. Another concern expressed in deriving such values was that acute toxicity has typically been thought of in terms of high dose and short time periods. However, for developmental toxicity effects, low doses at critical time periods are also of concern, so that the derivation of short term values needs to carefully consider other than high doses.

Most of the reviewers thought that the derivation of less-than-lifetime values will help with a movement towards harmonization of reference values for cancer and non-cancer effects. In terms of cancer versus non-cancer effects, there is a large database of single exposure data. Genotoxic and non-genotoxic effects should also be considered, which can result from short-term exposures early in life.

The discussion addressed the fact that there seem to be several different definitions of exposure duration used by various agencies in deriving these kinds of values. One reviewer suggested that the document provide more background on the history of the different values and describe why the differences in definitions came to be. An investigation might reveal why some definitions are different and why they need to be kept different. With this in mind, a single, standardized definition might not be the most efficient way to go. The reviewer, therefore, recommended that the definitions be examined but left flexible.

There was also discussion among the reviewers on the purpose of less-than-lifetime values. One reviewer suggested that the use of these values was not necessary for the protection of the public health because chronic values are protective for shorter duration exposure scenarios. Rather, these values would be used because they would make risk assessment easier and more practical. For example, a practical consideration in risk assessments for contaminated sites would be cost, where clean up may be less costly because higher concentrations might be considered acceptable in scenarios that involve shorter, less-than-lifetime exposures. Some of the other reviewers, however, disagreed with this suggestion. They pointed out that, in certain cases, chronic values are averaged out when a short term exposure is being assessed for public health protection (e.g., fish consumption advisories often dictate the infrequent consumption of fish, once per month, on the basis of an RfD; this use is not necessarily consistent with the intended use of the chronic RfD). The development of these less-than-lifetime values would make this incorrect practice unnecessary. All reviewers agreed that the “best science” should be used, but acknowledged that what constitutes best science is not always clear.

3.2 Charge Question #2

Please comment on the revised definitions for reference values.

In general, there was agreement among the reviewers that the definitions for reference values need further work, particularly with respect to terminology, notation, and presentation of qualitative and quantitative uncertainty information. There was general agreement with the use of the term adverse, instead of deleterious. However, many reviewers expressed concern with the last sentence of each definition ("The application of these factors is intended to provide an estimate centered within an order of magnitude."). Many reviewers wondered from where that part of the definition came and what the basis for it was. The Chair explained that EPA's original RfD definition used the word "perhaps" to explain the possibility that the imprecision in the reference value might span an order of magnitude, but depending on the underlying data it could also be greater than an order of magnitude, or less than an order of magnitude. The Chair went on to provide a general example of his interpretation of the revised definition, where if the reference value is "x", then the imprecision of the RfD would place its value as high as "3x" and as low as "0.3x." This imprecision exists in the RfD, primarily because the uncertainty factors are also imprecise. Several reviewers questioned whether the uncertainty would be "centered" around the value because reference doses are lower-bound estimates and most of the uncertainty would be on the upper end. The Chair acknowledged that this was another interpretation of "perhaps an order of magnitude." EPA's attempt to "center" the estimate was based on the interpretation used by most risk assessors within EPA.

One of the reviewers commented that it did not seem logical to include language on how uncertain a value is without supporting data. It is good to let people know the uncertainty of a value but this reviewer recommended that it not be included in the definition but rather, in a separate comment inserted after the definition. Many of the reviewers favored a more qualitative statement rather than having chemical-specific quantitative information. Some language was suggested and most of the reviewers supported this change ("In most instances, due to uncertainty in the underlying data and the mechanisms by which exposure causes disease, it will not be possible to say with certainty that exposures at or below the RfV are without risk, or that exposures moderately above the RfV pose any appreciable risk."). [Chair's Note to EPA: This points out the difference in using the word "uncertainty" versus "imprecision" in the definition of the RfV. The appropriate word to use here is "imprecision" because the range in the resulting RfV is not an attempt to quantify uncertainty, it simply reflects that fact that since the inputs to the RfV are imprecise (principally the uncertainty factors), so too is the estimate.]

One reviewer noted that a much larger problem needs to be addressed and that is that a quantitative definition for RfD does not exist. This reviewer felt that trying to describe the imprecision of the reference dose in qualitative terms does not make any sense if there is no quantitative definition for RfD. Several of the reviewers seemed to agree with this statement and acknowledged that in most cases, it is very difficult to determine what the reference value should be, especially as the data sets become less robust. Another reviewer commented that it is EPA's

mission to protect public health and to set reference values that will be protective for the general population and sensitive subpopulations. This first reviewer recommended that values of the RfVs be accompanied by clear quantitative statements. This reviewer suggested that it would be helpful to provide some indication of the level of protection provided by the reference value, for example, that they are protective of 95 or 99% of the population. Other reviewers acknowledged that these values are always fraught with uncertainty and the value cannot be precisely pinpointed. The definition is an attempt to openly acknowledge this uncertainty and to address possible misinterpretations. It was suggested that the wording be changed (i.e., take out the word "centered" from the definition) and use the suggested text above. It was also suggested that the parenthetical statement in the definitions that states "including susceptible subgroups," might also state "and sensitive lifestage of exposure."

One reviewer did not have as much of an objection to the wording in the definition as the other reviewers. The exact wording in the definition includes the following: "...application of these factors *is intended* to provide...". Use of the word "intended" is key in acknowledging that this may not always be the case. This reviewer felt that it is more important to describe what went into developing the reference values rather than trying to describe the imprecision as part of the definition.

A comment was made referring to the subscript labels for the reference values. If these definitions are presented on IRIS, more common language should be used. The use of the word acute could be confusing. It was suggested that labels such as 15-minute or 1-hour exposures, along with the type of exposure (e.g., inhalation), be used instead of the more general terms proposed. Some suggestions included "one day air concentration" or "Acute Inhalation Exposure Reference." Other reviewers agreed with some of these ideas. Other reviewers felt that the subscripts were too confusing and that the notations should be kept, but not as subscripts. There was also a comment that the subscripts/notations need to be defined in the text near where they are used so as to avoid confusion.

A philosophical issue was raised by one of the reviewers on having four exposure durations (acute, short-term, longer-term, and chronic). He noted that four divisions are just as arbitrary as the 3 commonly used (acute, sub-chronic, chronic). More importantly, if toxicity is defined in terms of exposure and exposure is defined in terms of dose and time, there really exists a need to look at the kinetics/dynamics of the exposure. There is also a need to find a way to use real times, not arbitrary ones. Similarly, a concern was brought up over what kind of burden on testing might be created if four exposure durations are included rather than three. Another reviewer commented that the exposure durations are meaningless if no specific time is associated with them and if a linkage to real life exposure scenarios is not provided. A suggestion was made that rather than having four definitions, it might be better to have one definition and say it is for a particular time duration so as to leave room for some flexibility. As is stands, the definitions seem to limit the time intervals to four (i.e., there are four and only four). This suggestion for a more flexible definition was supported by some of the reviewers.

Harmonization between cancer and non-cancer risk assessment was generally agreed to be a good objective for EPA. One reviewer, however, pointed out that EPA has been discussing different approaches for several years but has not yet decided on a particular approach. By eliminating reference to cancer and non-cancer endpoints, this essentially could be considered to be policy-forcing, in that the reference values would implicitly apply to cancer. EPA needs to clarify what its goal is for harmonization. Another comment was made that if cancer is an endpoint for a chemical with a particular reference value, then the definitions provided are inadequate. Also, if there is a latency effect from a less-than-lifetime exposure, the definitions again fall short.

3.3 Charge Question #3

Please comment on the recommendation that endpoint-specific reference values should not be derived.

The reviewers agreed that endpoint-specific reference values should not be developed but felt strongly that narrative needs to be added to the reference doses that explains what endpoint the RfD is based on, what other endpoints were considered, what areas were covered by testing, where deficiencies are, etc. This information needs to be explicitly laid out so that the data provided by IRIS will not be misinterpreted and/or misused. Reviewers acknowledged that such text is planned for the confidence statements associated with the various RfVs.

There was also some discussion among reviewers concerning the inclusion of developmental toxicity in derivation of reference doses. It was agreed that developmental toxicity should be considered when deriving all reference values, as one of many endpoints, but at least one reviewer felt that this inclusion needs to be made explicitly in an IRIS document. For example, a new notation could be used, such as RfD_{+DT} or RfD_{-DT} to indicate whether or not the value includes developmental toxicity as an endpoint. Most reviewers disagreed, however, and felt that this point was clear in the document and that the toxicological reviews already lay out this information. Furthermore, it was pointed out that this notation may be confusing and could be interpreted as whether or not the RfV is based on developmental toxicity.

Reviewers agreed that the intent of the reference value is to protect the most sensitive or critical, endpoints. If values are developed on the most sensitive endpoint, these values will cover anything less sensitive (including developmental toxicity). One reviewer suggested that information should be provided that details what endpoints were considered in developing a particular reference value, in addition to the critical effect. Furthermore, there may not be sufficient data for all endpoints and therefore, caveats need to be included in the narrative that explicitly state this. Other reviewers stated, however, that EPA's confidence discussion should address this point. One reviewer commented that this may not fully answer the questions raised and, therefore, a need may still exist to have discussion in the weight-of-evidence section that provide numbers for other endpoints along with the critical effect used in deriving the reference dose.

A reviewer also commented that it might be useful to have more detail on what an exceedence of the reference level might mean. Information could be included that would help a user to determine what effects might be expected if the dose /concentration is close to a reference level and whether the database is sufficient to determine the effect. It was pointed out that a database uncertainty factor has been applied for chronic exposures by EPA and that EPA will apply this factor to other exposure durations as well. Other reviewers disagreed and some indicated that this point should be further clarified in the document.

Returning to the need for endpoint-specific values, the Chair observed that EPA has new guidelines for chemical mixtures and the development of target (organ) toxicity doses. Although the reviewers do not support the development of endpoint-specific reference values, the two EPA RAF technical panels may need to discuss their apparent differences.

3.4 Charge Question #4

Please comment on the life-stage approach taken in this review, as well as the recommendations for alternative testing approaches and strategies for developing a targeted testing strategy to support setting less-than-lifetime reference values as well as chronic values. Also, please comment on a proposal from the Office of Children's Health Protection to request a study by the National Academy of Sciences to take a fresh look at toxicity testing approaches and strategies based on this and other reports (e.g., ILSI, 2001).

Overall, the reviewers felt that the Technical Panel had done a very good job of summarizing and discussing the different testing methods and the gaps that exist. Reviewers commented that the generalities that came out of the review of methods (e.g., gaps) are good, as is the general recommendation of places where endpoints might be added. Combining testing protocols is a good idea although there is some concern over the specifics of how this might be done. Mode of action research may be a way to gain more information of chemicals with limited testing. It was suggested that the overall toxicity testing topic gain the full attention of another panel, to address the deficiencies in more detail. One reviewer commented that the testing schemes were very elegant and will produce a lot of information but that more time and money will be needed to obtain that information. In considering the limitations of resources and time, it was suggested that the Agency only generate data on chemicals if there is a specific need for information that could be answered by testing.

There was general agreement among the reviewers that a request be made should be made to the NAS to study testing approaches. Reviewers cautioned that if a proposal is made, that it be broad in scope and time frame to address related issues such as latency, life stage, pharmacokinetics, and mode of action. Another reviewer suggested that only chemicals that have been well-studied should be looked at to see how extra data would really affect the reference value. Besides looking at additional data, it was further suggested that range-finding studies should be examined to understand how they are performed and how/why the doses are chosen, so that subsequent full

studies can be more useful for risk assessment. Another reviewer commented that many of the points discussed in response to this charge question should be the kinds of things the NAS study examines.

Reviewers also cautioned that some of the testing approaches discussed by EPA tend to go beyond what is realistic (e.g., in regards to reproducibility and interpretation). Also, there is little mention of mode of action, which, if discussed more fully, could open the door to many more methods. A comment was also made that before testing is performed, it would be very useful to conduct pharmacokinetic tests/modeling in order to gain more information about the timing and doses for subsequent toxicity tests. Ultimately, there will be a move towards differential testing. It was also suggested that more life stage data be collected and that it needs to be done more efficiently using existing tests. These are associated issues that an NAS study could examine. One reviewer suggested that in this move towards more testing protocols, that first, existing data for well-studied chemicals should be examined with two issues in mind: (1) information on latency to effect and (2) information on doses to neonates, based on individual mother exposures. A comment was made that there seemed to be an inconsistency in the immunotoxicity section of the document that discusses hypersensitivity and competency. There was confusion over when each of these should be used and what the goals should be. It was suggested that the authors of the document perhaps take another look at this section.

The time frame for implementing new testing is thought to be a number of years away However, in the meantime many new techniques will be developed that will provide very useful information. A suggestion was made that not every chemical needs to be tested; it would be more efficient to perhaps test classes of chemicals and extrapolate to individual chemicals. One reviewer added that the HPV testing approach is a good example where more information has been developed in this manner. It was suggested that this process be stimulated by using existing data-generating systems to understand mode of action and to create a mechanism whereby people could add in studies (e.g., mechanistic studies). The Chair of the EPA Technical Panel provided clarification that EPA was very open to ideas of how to display the data that exist and to encourage more work. One reviewer commented that a lot of planning and thought is needed before moving forward with new testing requirements. The idea of tiered testing was mentioned as well as the potential utility of strawman guidance to provide a template of how to do these studies, what to do, and what type of data EPA needs. Reviewers felt that such tools would provide more direction to researchers, help with prioritizing chemicals, and alleviate the problem of data incompatibility. There was also concern that overly complicated testing schemes might discourage people from initiating testing, except when it was absolutely needed.

Reviewers emphasized the importance of mode of action research and the anticipated advances in these techniques. Reviewers agreed that mode of action should be discussed as part of the harmonization of the risk assessment process between cancer and noncancer endpoints. This could follow approaches such as those used for endocrine disruptors or where mutagenicity is used as a predictor for carcinogenicity. Reviewers also encouraged EPA to continue to think about the type of information needed to protect the public. Research for research's sake is not

what is needed. The real need is to focus on the information gaps that exist, and to create a research environment where hypotheses are investigated in a useful manner. Ultimately, however, the goal is protecting public health and EPA's willingness to conduct risk assessments in the absence of more information, by the use of clearly articulated uncertainty factors, for example, was supported by the reviewers. There may be data gaps and uncertainty, but it is better to proceed than to not act while waiting for information to be developed on mode of action, mechanisms, or pharmacokinetics.

3.5 Charge Question #5

Please comment on the recommendation to include duration-adjustment for inhalation developmental toxicity as for other health endpoints.

Reviewers agreed with EPA that duration adjustment is important and should be attempted, recognizing the limitations in approaches. They also felt that it is important that the document stress that, in addition to the impacts to the developing fetus and children, developmental effects may be life-long concerns and some effects may not show up until later in life. The developmental toxicity definition should be expanded to include these latent effects, which may not be obvious from most testing protocols because they do not follow subjects through the aging process. Another comment from several reviewers emphasized the importance of matching the exposure duration with the data used to derive the RfD, particularly for intermittent exposure scenarios (e.g., fish consumption, hot spot/peak releases). If there are multiple peak exposures, should they be averaged out over the time period? If there is a one-time exposure, should this be treated differently? In general, the reviewers encouraged EPA to calculate exposures for an assessment by matching the way it was calculated in deriving the reference dose. As a result, several reviewers recommended that the document needs to provide the reader with adequate narrative that describes the underlying data, so the reference value is used in an appropriate manner.

One reviewer voiced the need for the additional 10x adjustment factor for added protection with respect to developmental toxicity. This reviewer felt that the document did not provide an adequate explanation as to why this factor is not included in reference values for chemicals with developmental toxicity effects. This concern is particularly evident for persistent chemicals (e.g., dioxins) where the subtle effects from very low doses may not be detectable. This reviewer later stated that it would be acceptable to base the reference value on an LOAEL/NOAEL from another endpoint, as is usually done, but the extra 10x uncertainty factor should be included to provide added protection due to developmental toxicity concerns. Other reviewers generally agreed with these statements, but noted that EPA already includes an uncertainty factor for data gaps, including those for developmental toxicity studies, or bases the assessment on NOAELs or BMDLs from developmental toxicity studies, e.g., methyl mercury. Few reviewers saw the need to use multiple uncertainty factors for the same concern (i.e., developmental toxicity).

One reviewer commented that duration adjustment presents a case that supports the use of pharmacokinetic information to understand the actual doses during critical windows. The

reviewer provided a **graph (Appendix E)** to display an approach where the actual dose during those critical time periods can be determined, instead of looking at the overall dose (total area under the curve) during a longer time period, which may result in an overestimation of the dose required to elicit a toxic effect.

Discussion of dose adjustment procedures turned back to limitations in approaches to adjust testing data to apply to other time periods of concern, particularly for those chemicals that exhibit "dramatic evidence" of developmental effects. One reviewer thought that it would be useful to have a way to identify those chemicals *a priori*, which would aid in deciding what adjustments need to be made. Another reviewer stated that future mode of action information will help in that regard and toxicologists will have better tools for predicting the effects of chemicals and groups of chemicals. Another reviewer pointed out the need to distinguish between deficiencies in testing data and limitations in how the data will be used in an assessment. This reviewer felt that there are existing data that could be used to improve these approaches, before additional testing is carried out in the future.

3.6 Charge Question #6

Please comment on the recommendations in the report for using a weight-of-evidence approach for hazard characterization (Chapter 4, section C.2.), and for expanding characterization of the extent of the database and using a narrative to describe the database rather than a confidence ranking (Section C.3.). Is the discussion of weight of the evidence clear in terms of how it would be used in characterizing the database? Also, please comment on the consideration of this information in the derivation of reference values.

The reviewers were generally supportive of the idea of using a weight-of-evidence approach for hazard characterization. There is a need to educate risk managers and to provide a transparent explanation of the quality of the data that are used in deriving reference values. It was also acknowledged that EPA is already doing something like this, as evident in the more recent IRIS toxicological review documents, but nonetheless done also in prior IRIS assessments. However, there was concern that if a fuller description of a weight-of-evidence approach will be required, there must be a good supporting infrastructure. It was suggested that EPA's guidance emphasize the goals of the process, while leaving room for scientific judgment. One reviewer provided an outline in the premeeting comments that suggested a more formalized method. Other reviewers found the examples presented in Tables 4-3 and 4-4 to be useful templates for displaying such information. Another reviewer offered to provide a paper describing a possible, published method to follow (Bayesian method).

Many reviewers agreed with the idea of using a prescriptive statement but were concerned that if the strength of the database is to be characterized, EPA will need to seriously think about what that means. Alternatively, a comment was made that there is a need to think about what is meant by weight-of-evidence and why and for whom it is included. For example, is it meant to be a way

for users to determine how concerned they should be over the RfD? Or is a weight-of-evidence description needed to help the individual determine the degree of severity or seriousness of the critical effect? The ranking of confidence that EPA now employs has some drawbacks, but the meaning behind such confidence statements as found for all RfDs and RfCs on IRIS is clear. The confidence rankings provide information on the likelihood that additional data will change the value of the RfD or RfC. One reviewer suggested that the existing narrative be enhanced and that the confidence rankings (high, medium, low) also be kept. In addition, focus should be placed on those end-users of the database who do not have an extensive knowledge of toxicology. The weight-of-evidence approach provides good information for professionals in toxicology, but others will need something like categories to help risk managers to understand the confidence in the data.

Several of the reviewers expressed concern over the use of the terms "robust" and " minimal." It was felt that these terms were too extreme and that there may be possible misinterpretations associated with them. It was, however, agreed that there is a need to characterize the degree of completeness of the database and it was suggested that perhaps, it would be a good idea to provide guidance for the inclusion of this type of information in the narrative. Characterizing the extent of the database should include a description of the quality of the information as well. One reviewer expressed concern that when the weight-of-evidence approach was applied to the case study presented in the document, this type of information was not apparent. It was also stated that it is useful to have a list of endpoints and then a decision needs to be made as to whether or not there are sufficient data to judge the confidence in the reference value that covers a particular endpoint.

The Chair posed a question to those reviewers who work for state environmental agencies: Have the confidence rankings that are provided in IRIS been a useful tool? Several reviewers stated that they have not been particularly influenced by the confidence rankings in IRIS. Rather, they are useful as a guide of where EPA stands on the value. One reviewer stated that on the low end, the rankings made sense but that on the high end, they did not. Another reviewer commented that the rankings were only really used when the hazard quotient was greater than one and the data were examined in more detail to locate the chemicals that were driving the risk.

3.7 Charge Question #7

Please comment on the recommendations for use of an exposure-response array and derivation of sample reference values to select the point of departure (POD) for the final reference value. Also, please comment on consideration of the nature of the effects for different endpoints, durations, timing and routes of exposure in selection of the POD.

In general, most of the reviewers liked the idea of an exposure array and the idea of using a visual tool to display the data, however, many had several concerns. The exposure array has two roles. One is to give a sense of the size or completeness of the database and the other is to give a sense

of the dose-response relationships for various endpoints. However, in the example presented in the document, each of the effects were treated equally, which resulted in an inability to portray to the reader the quality of the data or severity of the effect. Additionally, this example uses a minimal database; if more data points were presented, the graphic could lose resolution and not be as helpful. It was also noted that the reader must have knowledge of toxicology to understand the exposure array graphics. Several reviewers noted that it would be nice to have more information presented on the dose-response curves. It would be helpful to see dose-response curves for different endpoints in relationship to one another, as well as confidence in the severity of the effect. One reviewer expressed concern over relying solely on the use of graphics. It was stressed that they should only be used if the caveats and narrative associated with them are enhanced and clarified. Other reviewers responded by saying that the exposure array is a tool to help users decide which endpoint to choose and that it provides very useful information. Another point was brought up that the exposure array allows the user to see if individual endpoints lie on a continuum.

One reviewer was quite impressed by the presentation of information in the document in tables B-4 and B-5 and expressed a wish that this kind of information would be available for all the IRIS files. However, other reviewers noted that EPA routinely created such tables listing NOAELs, LOAELs or BMDLs, but did not add the columns for uncertainty factors and possible RfD or RfC values. At least one of these latter reviewers did not see the value of comparing multiple RfVs and then picking the lowest one, since this seemed to abridge the hard work of sorting among the NOAELs, LOAELs and BMDLs, ADME information, and mode of action understanding to determine the critical effect and weighing all the evidence.

In contrast, one reviewer expressed disappointment that the proposed approach did not end the use of the critical effect concept. Although the document seemed to be saying there was a movement away from the critical effect concept, there really was not. Others reviewers felt that stepping away from the existing critical effect concept would result in additional effort, with no value added. It was noted that the proposal presented here is different from what is done now. The proposal here suggests that the selection of the RfV occur from an array of RfVs after the calculation of the HED's or HEC's, a point of departure is selected for all relevant endpoints, and a series of reference values are derived from the use of uncertainty factors for each of the selected PODs. In this approach the lowest RfV may actually have the largest uncertainty factor. The current approach is similar to that proposed, except that the selection of the appropriate POD is done prior to the use of the uncertainty factors. In effect, the selected POD is the critical effect. One reviewer argued that, rather than picking the lowest point of departure, it is better to "run all the endpoints through to an RfD," possibly using meta analysis, which will better reflect uncertainty in the decision. Several reviewers agreed that this approach might have merit. It would diverge from the critical study/critical effect concept and carry the entire analysis through, particularly in cases where a benchmark dose is calculated, and then select the lowest reference value. Another reviewer commented that this meta analysis could be conducted at the level of the POD and not the RfV and the results would be similar and more consistent with current practice. The graphical display can be helpful with either analysis. Another reviewer commented that there

really is a need for the graphic to display two points, the POD and the reference value, for each endpoint, to determine how the uncertainty factors influence the values.

3.8 Charge Question #8

Please comment on several recommendations concerning the application of uncertainty/variability factors. Are there additional data or analyses in the literature not cited here that can be used to strengthen the recommendations? Should other factors be considered in the application of uncertainty/variability factors?

8A - *The Technical Panel recommends limiting the total UF applied to a chronic reference value for any particular chemical to 3,000. This maximum of 3,000 applies only to the UFs and does not include various duration adjustment factors. Setting a reference value when there is uncertainty in at least four areas would need to be carefully evaluated on a case-by-case basis.*

Many of the reviewers expressed concern over this recommendation, and generally felt that more clarification was needed. The wording was considered too ambiguous and could be taken to mean that users should truncate values so as to not exceed 3000 when the database was so weak as to not be able to develop an RfV. Some reviewers felt that it would be more meaningful to state that if a reference value has more than four uncertainty factors, then that should trigger concern and caution on the part of the assessor whether or not the database is sufficient to derive a value. Clarifying questions were posed among the reviewers to better understand what would be considered part of the 3000 uncertainty cut-off and what would not (e.g., human variability, probabilistic information). A reviewer felt that several points need to be noted: (1) determining the total uncertainty factor is a multiplicative process, and (2) it is difficult to put "titles" to the different sources of uncertainty because it is hard to separate the issues into categories. A reviewer responded by noting that Canada uses one factor (1 to 100) to characterize uncertainty for study duration, lack of a NOAEL and database deficiencies.

One reviewer reiterated a concern over merging uncertainty factors with adjustment factors. It was suggested by another reviewer that the rules on limiting the number or magnitude of uncertainty factors should be left flexible, so the assessor can use judgement. This is particularly evident in the Superfund program where many times there is a need for a RfD for site assessments, even if the uncertainty associated with them is high.

A question was posed by one of the reviewers as to why the uncertainty process was multiplicative and not additive. The Chair responded that each factor was considered to be independent and therefore, they would be multiplied, not added. However, if you have uncertainty in five areas, the total factor would be as high as 10,000, and not 100,000, because in brief the multiplication of multiple conservative factors yields unrealistic values. (This is explained more fully in existing EPA methods texts on RfD.) Another reviewer added that an idea has been put forth to use the squares of the uncertainty factors rather than multiplying them,

which would put the result between adding and multiplying. It was suggested that a fresh look be made as to how uncertainties are combined since there might be a better way than multiplying, perhaps using distributions. A reviewer suggested that EPA look at a reference by Swartout et al. (2001) or Baird et al (1996) that discuss distributions.

One reviewer suggested that the part of this recommendation describe uncertainty in 4 areas as a red-flag, making it clear that the use of a 3000-fold uncertainty factor or greater is meant to be a warning light and not a prohibition. Also, emphasis should be placed on providing an adequate narrative to provide a transparent description of the basis for the uncertainty factors.

8B - *The Technical Panel supports and expands the recommendation of the 10X Toxicology Working Group, i.e., that reduction of the intraspecies UF should be considered only if data are sufficient to support the conclusion that the data set on which the POD is based is representative of the exposure/dose-response data for the susceptible subpopulation(s). Given this, whether and how much the intraspecies UF may be reduced must be linked to how completely the susceptible subpopulation has been identified and its susceptibility described (e.g., versus assumed). At the other extreme, a 10-fold factor may sometimes be too small because of factors that can influence large differences in susceptibility, such as genetic polymorphisms. The Technical Panel urges the development of data to support the selection of the appropriate size of this factor, but recognizes that often there are insufficient data to support a factor other than the default.*

One reviewer noted that this is analogous to the FQPA factor which assumes a 10-fold uncertainty unless there are data that support a smaller value. The Chair responded that EPA routinely changes the 10-fold uncertainty factor, generally to a lower number, if sensitive subpopulations have already been addressed in a particular assessment (e.g., nitrate and selenium on IRIS). Several reviewers suggested that the recommendation clarify that the factor to which it refers is the usual default uncertainty factor that accounts for inter human variability and uncertainty.

Some reviewers stated that the recommendation was too prescriptive and didn't allow for use of scientific judgement to modify the usual default factor of 10. A comment was also made that for certain exposures (e.g., inhalation), that the 10-fold factor may not be a good default value. It was pointed out, however, that many times there simply are not sufficient data and in those cases, this 10-fold factor would be used.

One reviewer asked whether or not EPA has considered how far risk assessors need to go to protect individuals within sensitive sub-populations. It was acknowledged that it might be impossible to protect the entire population. Reviewers recognized the differences in the susceptibilities of individuals within different populations, considering factors such as lifestyle and genetic polymorphisms, but one reviewer questioned whether these differences would be as large as a factor of 10. Another reviewer suggested that independent guidance needs to be developed as to how far EPA will go to protect particular individuals within sensitive subpopulations. EPA

already considers sensitive populations on a case-by-case basis but the issue this recommendation addresses is should EPA make a probabilistic statement with each RfV (see also the comment above about distributions for uncertainty factors). A suggestion was made to look at animal studies to narrow the 10-fold factor and to put emphasis that this factor is more oriented at variability than uncertainty. An EPA Technical Panel representative clarified that one point to be made with this recommendation was the fact that there is not much information on population variability and there is a need to promote obtaining more such information on children as well as adults and the aging.

8C - *Sound scientific judgment should be used in the application of UFs to derive reference values which are applied to the value chosen for the POD (NOAEL, LOAEL, or BMDL) derived from the available database. IRIS documentation should contain a justification for the individual factors selected for a particular chemical because rigid application of UFs could lead to an illogical set of reference values.*

All of the reviewers generally agreed with this recommendation. They all agreed that promoting collection of data is a good idea. One reviewer suggested leaving $10^{0.5}$ as it stands, instead of describing this factor as 3-fold.

8D - *Given that there are several UFs that can be used to deal with data deficiencies as part of the current reference value process, and given that the FQPA safety factor is assumed to overlap to a large extent with these factors, the Technical Panel agrees with the 10X Task Force that the current interspecies, intraspecies, LOAEL-to-NOAEL, subchronic-to-chronic, and database deficiency UFs, if appropriately applied using the approaches recommended in this review, will be adequate in most cases to cover concerns and uncertainties about children's health risks. Any residual concerns about toxicity and/or exposure can be dealt with in risk characterization/risk management (e.g., by retention of all or part of the FQPA safety factor for pesticides).*

This recommendation appeared fairly vague to some of the reviewers. Other reviewers felt that the argument was not based on data and did not have scientific backup. It was then further clarified that the EPA Technical Panel was trying to identify the scientific issues behind the FQPA factor and then to state whether these issues were addressed in the context of other, existing uncertainty factors. Many of the reviewers agreed that adding another 10x factor on top of what already is included did not sound like a good idea because other factors already cover the underlying scientific issues and provide an adequate margin of safety. Other reviewers were pleased that EPA has taken a systematic look to see how this FQPA factor relates to the existing uncertainty factors before adding it as another default 10-fold factor.

One reviewer pointed out that two issues were really being addressed here. One is what happens when there are no data and the other is what happens if there are data on gross developmental effects. This reviewer's perspective is that developmental toxicity data are not very limited, even though they appear to be so. Another reviewer also saw a need to clarify the potential for

different types of developmental toxicity and to better clarify the its sub-categories.

There was a general agreement among the reviewers that the Agency should have the ability to deal with this on a case-by-case basis, and it was felt that the addition of another factor would not help EPA's judgment. Although it was felt that there is a need to address developmental toxicity as a unique database deficiency, the reviewers acknowledged that EPA already addresses this to some extent with its database uncertainty factor. In addition, one reviewer pointed out that the interspecies and intraspecies factor may also address this concern . Reviewers felt that more relevant data collection is needed and that if developmental toxicity data are missing and considered relevant, then that gap needs to be flagged so it can be filled.

8E - *The Panel considers the purpose of the MF to be sufficiently subsumed in the general database UF. Therefore, the Panel recommends the discontinuance in use of the MF.*

Most reviewers agreed with EPA's recommendation to discontinue use of the modifying factor because of its infrequent use and coverage by other factors. However, several comments were brought up by reviewers in reference to this recommendation. There are cases where it is fairly certain that a chemical at a low dose is a carcinogen and in these particular cases, it might be prudent to add a MF. [Chair's Note to EPA: It was also acknowledged that EPA's Office of Water does something similar to this for Class C carcinogens, not as part of the RfD derivation, but in developing drinking water criteria.] A few reviewers also noted that different state agencies address the issue by adding an extra 10x factor as part of the risk characterization, rather than as part of deriving the RfD. One reviewer suggested that this might be a good place to add in mode of action information. However, another reviewer felt that a place for mode of action exists, but that this was not it.

8F - *The Agency is encouraged to develop its own guidance, based on some of the available methodologies (e.g., IPCS), but caution should be used in that there are relatively few data available for many substances to serve as an adequate basis to replace defaults with CSAFs.*

Reviewers generally agreed with this recommendation but were a little wary of the cautionary statement. Reviewers agreed that criteria were needed on how much data are enough and what kind of data are needed to justify a chemical specific adjustment factor. The key issue is the need for hard data that are compelling, such as those available for methylmercury. Reviewers felt that the case study presented in this methods text had no hard/raw data. Another reviewer suggested that EPA look for examples in the SAB review of the RfC methods where a series of cases studies were assembled. Another reviewer recommended that EPA examine a 1998 paper by Renwick. Also, several reviewers added that the IPCS report on methods for CSAF already contains criteria that EPA might review for its purposes, and also contains appendices of chemical-specific data that could be used for this purpose. The EPA Technical Panel Chair clarified the cautionary statement by saying that the caution was not to let chemical specific data drive an assessment when data are not available on the endpoint of concern. One reviewer offered the comment that

some of this analysis is not as complex as thought and encouraged assessors to examine available data for appropriate opportunities to include chemical specific adjustment factors.

3.9 Other Discussion Topics

Following the discussion of the eight charge questions, the reviewers raised other issues that they felt would improve the document.

Reversibility - This issue had been raised briefly earlier in the workshop (in the discussion of acute effects) and was held for further discussion in this session. Because of time constraints, the Chair stated that he would provide comments as part of his individual written comments. Another reviewer noted that a study on reversibility had been published by CRC Press in the 1990's. The reviewer will submit the paper as part of his written comments.

Route-to-route extrapolation - One reviewer raised this issue and stated that his written premeeting comments adequately addressed the issue.

Benchmark Dose - One reviewer noted that the benchmark dose is very useful but it has not been defined as to what it means and how it is to be used (relative to NOAELs/LOAELs, points of departure, etc.) in the future development of reference values. While it is a useful tool, there is still a question over what it's target is and how it relates to derivation of reference doses. Another reviewer responded that EPA has been very tepid in regard to this issue. It was felt that the NOAEL should really not be taken to be equal to a background response since many times this is below the statistical power of the study to determine. The benchmark dose has a specific response value associated with it while a NOAEL is an artifact of the experimental design, so basing a reference value on it might be of concern. This reviewer finished his comment by noting that "there must be a better method available."

Uncertainty Factor for Severity - One reviewer felt that there was a serious omission in the current and proposed approach and that risk assessors should also look at the severity of effects at the chosen POD. This reviewer suggested that EPA should apply uncertainty factors and see which resulting reference value would be better, based on severity of effect. He felt that this uncertainty factor was needed because of the uncertainty surrounding the use of a NOAEL as the point of departure, particularly when frank effects are observed. To do this would require ranking effects (e.g., minimal, moderate, severe, lethal), and assigning of a default severity factor to each, which would provide added protection for more severe effects. The added benefit of using this approach is that it also promotes harmonization with the risk assessment of cancer endpoints, if these are considered as frankly toxic, or their precursors, which might be considered as minimally toxic. Now the measure of cancer and non-cancer are on different scales. A common approach based on severity would put them both on the same scale. The Chair clarified that the severity uncertainty factor should be applied to the point of departure (a LOAEL or BMD) based on severity of the effect. One reviewer noted that this is done by the Office of Water in their

assessments for chemicals with some evidence of carcinogenicity, but this factor is not used to develop the RfD per se. Some reviewers expressed concern over how the severity rankings would be created, however, especially across endpoints. These reviewers felt that there was not enough data on the biology to create a quantitative ranking system. In contrast, EPA has already established such ranking systems in its Superfund program, and uses a crude ranking system for its existing RfD and RfC methods (i.e., NOAEL, UF = 1; minimal LOAEL, UF = 3; more severe LOAEL, UF = 10; FEL, RfD/RfC not normally estimated). Other reviewers stated that adding another uncertainty factor was not a good idea and this issue of severity is more of a policy decision. Rather, it was suggested that perhaps this factor could be addressed in the database uncertainty factor, if the assessor feels the need to have an extra level of protection. At least one publication discusses the use of this factor by different health organizations (Renwick, 1995). The Chair stated that he will provide this reference as part of his written comments.

4.0 OBSERVER COMMENTS

Two periods during the workshop were dedicated for observer comments, one in the late morning and one in the late afternoon after the eight charge questions had been completed. Summaries of the observer comments and subsequent discussion are presented below.

A comment was made by Abraham Tobia of Bayer Corporation that when exposures are broken down into specific time periods, such as short-term and acute, a problem arises over terminologies for exposures that are intermittent. There is a need to capture this problem somehow and address how people should handle intermittent exposures. This also brings into account the reversibility issue.

The reviewers felt that this comment touched on an important issue. When less-than-lifetime exposures are examined for lifetime effect, there is the question over whether this means delayed or latency effects or whether the exposure occurs in bursts over a lifetime. There is a definite question over how to address bursts of exposure and to consider latent effects.

Abraham Tobia also expressed concern over the fact that there has not been any discussion over how the data are going to be used or interpreted. If there is no data evaluation method, then what statistics or calculations are going to be used? There is a need to understand how the data will be used in the risk assessment process and also before testing methods are developed.

Jennifer Seed, from EPA's Technical Panel, provided clarification that the purpose of this document is not to do what Mr. Tobia suggests but that there are a number of other efforts underway to develop such guidance with ILSI. This effort should be viewed as the first step in this process. The idea is to approach the strategies for testing from the point of view of what is needed for risk assessment.

Comments were made on how important it is to go back and look at the data that already exist to understand better what is needed and where those gaps exist.

Elizabeth Margosches, from EPA/OPPT, commented on the graphic display of the exposure array and pointed out that there are now various ways to display complex variables (e.g., multivariate displays) such as those shown by Edmund Tufte. Perhaps these could be examined for their use in the exposure arrays.

Onyemaechi Nweke posed two questions to the reviewers: (1) How are risk assessments dealt with when there are conflicts over the strength of the database? Can the definition of the database uncertainty factor be expanded to address issues that effect the outcome (e.g., positive and negative studies)? and (2) How do you handle biological versus statistical significance?

In response to the first question, reviewers commented that those types of situations should be handled in the weight-of-evidence section and in evaluating database sufficiency. This is a case-by

-case practice and is based on best professional judgement when examining the critical effect such that it is not clear if the effects are adverse or not. If this occurs in the hazard characterization, then it would not be addressed as part of the database uncertainty factor. The reviewers stated that in this case, it is important that the data and assessment be presented in a transparent way so all the issues are aired.

In response to the second question, reviewers commented that there are some very good texts (e.g., ones by Rick Hertzberg) that address that question. They stated that it is not a question of one or the other. It was also suggested that benchmark analyses could be used to deal with those types of situations.

The Chair of the EPA Technical Panel provided clarification on an earlier comment made by one of the reviewers on a method to estimate the doses in critical windows, rather than the overall dose under the curve, representing a longer time period. It was felt that this comment did not really apply to what EPA was trying to accomplish. It applies if the goal is to examine a particular endpoint but EPA has to defend against all kinds of developmental toxicity and for example, animal critical windows are much narrower than human critical windows. Therefore, there is a need to think about the area under the curve rather than the window of susceptibility.

Barbara Henry, from Bayer Crop Science, commented that she supports the idea to review the existing developmental toxicity database to see the extent of data already collected and how the proposed approaches will influence the derivation of reference values.

Appendix A

List of Peer Reviewers

Peer Review Workshop on EPA's
A Review of the Reference Dose and Reference Concentration Processes

June 19, 2002

Peer Reviewers

Name	Affiliation
Janusz Byczkowski	Independent Consultant
Eric Clegg	U.S. Army Center for Environmental Health Research
Kenny Crump	Environ Corp.
George Daston	Proctor & Gamble Company
Rodney Dietert	Cornell University
John Doull	University of Kansas School of Medicine
Michael Dourson	Toxicology Excellence for Risk Assessment
Pat McGinnis	Syracuse Research Corp.
Bonnie Ransom Stern	BR Stern Associates
Pam Shubat	Minnesota Department of Health
Ellen Silbergeld	Johns Hopkins University
Alan Stern	New Jersey Department of Environmental Protection
Lauren Zeise	California Environmental Protection Agency

Appendix B

List of Observers

Technical Peer Review Workshop
On The Draft Document Entitled:
A Review of the RfD and RfC Processes

June 19, 2002

OBSERVER ATTENDANCE	
Kulbir Bakshi	NAS/BEST
David Baylos	U.S. EPA - NCEA
Mike Beringer	U.S. EPA
Elizabeth Boa	American Chemistry Council
Michael W. Broder	U.S. EPA - ORD/NCEA
Marilyn Brower	U.S. EPA - Risk Assessment Forum
Margaret Chu	U.S. EPA - ORD/NCEA
Joyce Donohue	U.S. EPA - OW
Irene Dooley	U.S. EPA - OW
Chuck Elkins	Chuck Elkins & Associates
Michael Firestone	U.S. EPA
Michael J. Firth	ExxonMobil Biomedical Sciences, Inc.
Lynn Flowers	U.S. EPA - NCEA
Gary Foureman	U.S. EPA
H. Galal-Gorchev	U.S.EPA - OW/OST/HECD
Jane Hamblen	AMEC
Barbara Henry	Bayer Cropscience
Colette S. Hodes	U.S. EPA - OPPT/RAD
Lee Hofmann	U.S. EPA - OSWER
Leslie J. Hushka	ExxonMobil Biomedical Sciences, Inc.
Jennifer Jinot	U.S. EPA
Elizabeth Ketchum	U.S. EPA - OSWER

OBSERVER ATTENDANCE	
Carole A. Kimmel	U.S. EPA - ORD/NCEA
Gary Kimmel	U.S. EPA - NCEA
Aparna M. Koppikar	U.S. EPA - NCEA
Steven Kueberuwa	U.S. EPA - Office of Science & Technology
Kate Mahaffey	U.S. EPA - OPPTS
Susan Makris	U.S. EPA
Elizabeth Margosches	U.S. EPA
Beth Mileson	Technology Sciences Group
Amy Mills	U.S. EPA - ORD/NCEA
Jini Mohanty	U.S.EPA - OW
Onyemaechi Nweke	U.S. EPA - OPEI
Edward V. Ohanian	U.S. EPA - HECD/OST/OW
Yogi Patel	U.S. EPA - OW
Pat Phibbs	BNA, Inc.
Deborah Rice	U.S. EPA - NCEA
Susan Rieth	U.S. EPA
Amy Rispin	U.S. EPA
Jim Rowe	U.S. EPA - ORD
Jennifer Seed	U.S. EPA - OPPTS
Yvette Selby	U.S. EPA - OSWDW
Jack Snyder	Styrene Information & Research Center
Abraham Tobia	Bayer Cropscience
Paul White	U.S. EPA - NCEA
Diana Wong	U.S. EPA - Office of Water
Bill Wood	U.S. EPA - Risk Assessment Forum

Appendix C

Charge Questions

A REVIEW OF THE RFD AND RFC PROCESSES
CHARGE QUESTIONS - EXTERNAL PEER REVIEW

This document is a draft Risk Assessment Forum Technical Panel Report. This is not a guidance document but represents an analysis of the current RfD/RfC process and a series of recommendations to improve the process. The Peer Review Panel is being asked to review the recommendations of the Technical Panel and to provide comments regarding the scientific rationale for the recommendations. Final decisions on implementing the recommendations from the Technical Panel will be made by the Agency's Science Policy Council. Comments from the external peer reviewers will help inform the process.

The report of the RfD/RfC Technical Panel of the Risk Assessment Forum, *A Review of the Reference Dose and Reference Concentration Processes*, summarizes the review and deliberations of the Panel. The RfD/RfC Technical Panel was established by EPA's Risk Assessment Forum in response to a request from the Agency's 10X Task Force[1] to the Science Policy Council and the Risk Assessment Forum. The Science Policy Council and the Risk Assessment Forum agreed that the issues raised by the 10X Task Force should be examined on a broader scale than just for pesticides, with input from various program offices within the Agency and from the outside scientific/policy community. Later, the charge to the Technical Panel was expanded by the Forum to include a more in-depth review of a number of issues related to the RfD/RfC process, in part because of several other Forum activities that were underway (e.g., development of the Framework for Harmonization of Approaches to Health Risk Assessment, the Benchmark Dose Guidance document, and the Carcinogen Risk

[1]The 10X Task Force was created by the Administrator, EPA, to explore the adequacy of current testing approaches for pesticides for protecting children's health, and to recommend approaches for implementation of the additional 10X safety factor mandated by the 1996 FQPA.

Assessment Guidelines). Additionally, the RfD/RfC process as a whole had not been reviewed for some time.

The Technical Panel makes a number of recommendations for improvements in the RfD/RfC process as well as additional efforts that are needed. The document is a review, not guidance, but it does make recommendations that should be considered in the implementation of changes in the current process and/or development of needed guidance. The Agency is committed to harmonization of health risk assessment procedures, including the harmonization of approaches for noncancer and cancer endpoints and making efficient use of animal testing to achieve this goal. As noted several places in the document, all such topics have not been discussed and resolved by the agency. For instance, the differences in scaling factors used for cancer and noncancer derivations from oral exposure data is discussed as an issue that has not been resolved. Thus, there will likely be the need for revised or further guidance in the future on this and other items.

The methodology developed in the RfD document is considered generally applicable to both cancer and noncancer endpoints where dose response relationships are thought to be either nonlinear or consistent with a threshold. Although the emphasis in this document is on the calculation of RfDs and RfCs, the same processes and considerations are applicable to the Margin of Exposure, as discussed in the Draft cancer risk assessment guidelines (EPA, 1999d).

As part of its deliberations, the Technical Panel has considered the recommendations of the Toxicology Working Group of the 10X Task Force (EPA, 1999a, see detail of recommendations in Appendix A of this report). The following charge questions are posed for comment by the Peer Review Panel.

CHARGE QUESTIONS

1. Please comment on the recommendation to derive less-than-lifetime reference values in addition to the chronic RfD and RfC? Is the need for these values adequately justified in the report?

Issue: The 10X Toxicology Working Group felt that data on developmental toxicity would often be a greater factor in calculating less-than-lifetime reference values, and that exposures to children are more often of this type. The RfD/RfC Technical Panel concurred with this view and expanded the types of data to be considered for different duration reference values to include other life stages as well.

Recommendation: *The Technical Panel concurred with the recommendation of the 10X Task Force that reference values should be derived, where possible, for acute, short-term, and longer-term durations, as well as chronic exposures for oral, dermal, and inhalation routes, and that they should be included in the IRIS database for use by EPA programs, where applicable. The Technical Panel recommended that the definitions for duration should be standardized but left flexible so they can be adjusted depending on the exposure situation of concern.*

2. Please comment on the revised definitions for reference values?

Issue: Definitions were proposed for the acute, short-term, and longer-term reference values as well as revisions for the chronic RfD and RfC. The revised definitions (see below) are aimed at clarifying (a) that the reference value is intended to provide an estimate that is centered within an order of magnitude, further emphasizing that the estimate is not a bright line, but has some range of variability that may be considered by risk managers in decision making; (b) the term "deleterious" in the original definitions has been replaced with the term "adverse," because the latter is more commonly used and understood in data evaluation and selection of endpoints for

setting reference values; and (c) the term "noncancer" has been removed from the original definitions in the spirit of harmonization of risk assessment approaches for human health effects because it has been recommended that health effects no longer be categorized as "cancer" or "noncancer" for the purposes of hazard characterization and dose-response analysis. This change denotes the move toward defining approaches for low dose estimation or extrapolation based on mode of action.

With this new set of definitions, standardizing the terminology used to refer to the reference values would help clarify the scope and purpose of each reference value in terms of route and duration of exposure.

Recommendations:

A. *Use revised definitions for the reference values as follows.*

Acute [Oral, Dermal, Inhalation] Reference Value: *An estimate of an exposure for 24 hours or less to the human population that is likely to be without an appreciable risk of adverse effects for a lifetime (including susceptible subgroups[a]). It can be derived from a BMD, a NOAEL or a LOAEL, with uncertainty/variability[b] factors generally applied to reflect limitations of the data used. The application of these factors is intended to provide an estimate centered within an order of magnitude.*

Short-Term [Oral, Dermal, or Inhalation] Reference Value*: An estimate of an exposure for up to 30 days to the human population that is likely to be without an appreciable risk of adverse effects for a lifetime (including susceptible subgroups). It can be derived from a BMD, a NOAEL or a LOAEL, with uncertainty/variability factors generally applied to reflect limitations of the data used. The application of these factors is intended to provide an estimate centered within an order of magnitude.*

Longer-term [Oral, Dermal, or Inhalation] Reference Value: *An estimate of an exposure for up to approximately 7 years (10% of the average life span) to the human population that is likely to be without an appreciable risk of adverse effects for a lifetime (including susceptible subgroups). It can be derived from a BMD, a NOAEL or a LOAEL, with uncertainty/variability factors generally applied to reflect limitations of the data used. The*

application of these factors is intended to provide an estimate centered within an order of magnitude.

Chronic [Oral, Dermal, or Inhalation] Reference Value: *An estimate of an exposure for up to the average life span of the human population that is likely to be without an appreciable risk of adverse effects for a lifetime (including susceptible subgroups). It can be derived from a BMD, a NOAEL or a LOAEL, with uncertainty/variability factors generally applied to reflect limitations of the data used. The application of these factors is intended to provide an estimate centered within an order of magnitude.*

---

[a]*Susceptible subgroups may refer to life stages, e.g., children or the elderly, or to other segments of the population, e.g., asthmatics or the immune-compromised, but they are likely to be somewhat chemical-specific, and may not be consistently defined in all cases. See Chapter 4 (Section C.2.c) for further discussion.*

[b]*See discussion in Chapter 4 (Section D.5) on application of uncertainty/variability factors.*

B. *Standardize terminology as much as possible, e.g., using a generic term such as reference value (RfV) with subscripts to designate route and duration of exposure, e.g., RfV_{AO} for acute oral reference value.*

3. Please comment on the recommendation that endpoint-specific reference values should not be derived.

Issue: Given the recommendation for setting different duration reference values, the Technical Panel felt strongly that all relevant endpoints should be considered in the derivation of each duration- and route-specific reference value, thus ensuring that reference values are derived to be protective of all types of effects for that route and duration of exposure. The original recommendation for setting a RfD_{DT} (reference dose for developmental toxicity) was based on the fact that developmental toxicity did not require a chronic exposure scenario, which was the basis for the only values set at that time, the RfD and RfC. If the derivation of less-than-lifetime

reference values that account for all forms of toxicity is adopted, there will be no need for endpoint-specific reference values, although this recommendation does not preclude such derivation for certain purposes, e.g., endpoints based on a common mode of toxicity for cumulative risk assessment.

Recommendation: *Endpoint-specific reference values should not be developed, including the* RfD_{DT} *(reference dose for developmental toxicity), as originally proposed in the Guidelines for Developmental Toxicity Risk Assesment (EPA, 1991).*

4. A review of current guideline study protocols and approaches to testing was conducted to determine what information is currently developed to support the derivation of less-than-lifetime reference values. Please comment on the life-stage approach taken in this review, as well as the recommendations for alternative testing approaches and strategies for developing a targeted testing strategy to support setting less-than-lifetime reference values as well as chronic values. Also, please comment on a proposal from the Office of Children's Health Protection to request a study by the National Academy of Sciences to take a fresh look at toxicity testing approaches and strategies based on this and other reports (e.g., ILSI, 2001[2]).

Issue: The Technical Panel reviewed and evaluated current testing guidelines and approaches testing approaches as a follow-up to its recommendation concerning the derivation of less-than-lifetime reference values. This review was undertaken to determine what information is currently gathered with regard to life stage assessment, endpoint assessment, route and duration of exposure, and latency to response. A primary goal of this review was to provide a basis for recommendations for the development of innovative alternative testing approaches and the use of such data for risk assessment. The Technical Panel is not recommending additional testing for every chemical but is suggesting that alternative strategies and guidance for testing approaches be developed that incorporate information on pharmacokinetics and mode of action

[2] http://www.ilsi.org/publications/pubslist.cfm?pubentityid=8&publicationid=305

early in the process, thus allowing a more targeted testing approach. In addition, alternative protocols are discussed that are aimed at more efficient use of animals and resources in combined studies that would provide more extensive data on life stages, endpoints and other factors not well characterized in current testing approaches. Recommendations were also made about research areas that should be encouraged to aid in better study design and interpretation of data for risk assessment.

Recommendations:

A. *Develop several new OPPTS guideline study protocols and modify current ones to collect more comprehensive data across life stages, route, duration and timing of exposures, that would be useful for setting acute and short-term reference values. Develop guidance for how and when to use the guidelines.*

B. *Develop additional guideline study protocols to evaluate several potential children's health issues, e.g., developmental immunotoxicity, carcinogenesis, more detailed neurotoxicity, pharmacokinetics, including direct dosing of neonates. Develop guidance for how and when to use the guidelines.*

C. *Encourage research to evaluate latency to effect and reversibility of effect from less-than-lifetime exposures. Encourage research on mechanisms/modes of action and pharmacokinetics at different life stages.*

5. Please comment on the recommendation to include duration-adjustment for inhalation developmental toxicity as for other health endpoints.

Issue: Duration adjustment for inhalation developmental toxicity studies is a notable exception to the practice of adjusting from intermittent to continuous exposures for inhalation toxicity studies. EPA's Guidelines for Developmental Toxicity Risk Assessment (1991) recommended against dosimetric adjustment on the basis that developmental effects were more likely to depend on peak exposure concentration. However, more recent information from the literature shows the relevance of area under the curve in several studies and supports the use of some type

of adjustment, even for short half-life agents, such as ethylene oxide. Based on this information, together with the rationale used for dosimetric adjustment for other health effects (i.e., that exposure adjustment based on C × t tends to be more health protective), the Technical Panel has recommended that duration adjustment procedures from intermittent to continuous exposures be used for inhalation developmental toxicity studies as for other health effects from inhalation exposure. Of course, if specific data and/or models on pharmacokinetics or mode of action are available for determining the proper dose metric, these should be used instead of the default duration adjustment. The Panel urges continued development of data, modeling, and improved procedures for dose-duration adjustments related to developmental toxicity.

Recommendation: *Duration adjustment procedures from intermittent to continuous exposures for inhalation developmental toxicity studies should be done in the same way as for other health endpoints.*

6. Please comment on the recommendations in the report for using a weight-of-evidence approach for hazard characterization (Chapter 4, section C.2.), and for expanding characterization of the extent of the database and using a narrative to describe the database rather than a confidence ranking (Section C.3.). Is the discussion of weight of the evidence clear in terms of how it would be used in characterizing the database? Also, please comment on the consideration of this information in the derivation of reference values.

Issue: The 10X Task Force had recommended the use of a weight of evidence approach and considerations for level of concern in evaluating the data on children's health effects. These recommendations have been incorporated into the approach to hazard characterization proposed here for factors to be considered in a weight of evidence approach, and for the use of a narrative description rather than confidence rankings. The narrative approach is intended to emphasize the extent of the types of data available (both human and animal data) as well as the data gaps that could improve the derivation of reference values, and it should encourage a wider range of

information to be used in deriving reference values, taking into consideration the life stages evaluated, the issues of timing, duration and route of exposure, the types and extent of endpoint assessment (i.e., structural and function), and the potential for latent effects and/or reversibility of responses. Factors to be considered in a weight of evidence evaluation both generally, and for characterizing potentially susceptible subpopulations, are described. The extremes for the extent of the database, i.e., minimal or robust, are defined in Chapter 4, but the Technical Panel did not define additional categories between minimal and robust and had serious concerns about developing such categories because of the tendency to try to characterize a database with single word descriptors, i.e., high, medium and low confidence. Instead, a narrative description of the extent of the database, with emphasis on the strengths and limitations of the data was strongly encouraged.

Recommendation: *An expanded approach to the evaluation of studies and characterization of the extent of the database as a whole is recommended; in particular, several factors are discussed that should be considered in a weight-of-evidence approach for characterizing hazard for the population as a whole as well as for potentially susceptible subpopulations. As part of this evaluation, a narrative approach should be used in describing the extent of the database instead of using a confidence ranking of high, medium, or low.*

7. Please comment on the recommendations for use of an exposure-response array and derivation of sample reference values to select the point of departure (POD) for the final reference value. Also, please comment on consideration of the nature of the effects for different endpoints, durations, timing and routes of exposure in selection of the POD.

Issue: Currently, the "critical effect" is used as the basis for the POD, and various UFs are applied to the dose at the critical effect for derivation of the RfD or the RfC. The critical effect is defined as "the first adverse effect, or its known precursor, that occurs to the most sensitive species as the dose rate of an agent increases" (EPA, 2002a). The underlying assumption is that

if the RfD or the RfC is derived to prevent the critical effect from occurring, then no other effects of concern will occur; in addition, this approach assumes that the relationship of various health effects for a particular chemical is maintained across species. The Technical Panel is concerned that presenting only a single critical effect and the critical study from which it was derived in the IRIS summary table that appears at the beginning of each RfD or RfC file may not provide enough information to the reader who is unfamiliar with risk assessment, and thus could be misleading. Presentation of a single endpoint as a POD for a systemic effect, for example, cannot capture the nature of the dose-response curve for that particular endpoint. Nor does the presentation of a single endpoint convey the possibility that other more serious endpoints may have a dose-response character markedly different from the less serious endpoint. Most importantly, in light of the Technical Panel recommendations for deriving an expanded number of reference values for different durations and routes of exposure, the limitations of focusing only on the critical effect become apparent because the most sensitive endpoint may be different for different durations or routes of exposure.

Recommendation: *An exposure-response array should be used as a visual display of all relevant endpoints and durations of exposure, as shown in the case study. This array can be used to evaluate the range of exposure-response data for different durations of exposure in order to determine the range of numerical values available for each route and duration of exposure. The POD should be selected on the basis of an evaluation of all relevant endpoints carried through to reference value derivation with selection of the limiting value(s) as the final step rather than on a single "critical study" and "critical effect."*

8. Please comment on several recommendations concerning the application of uncertainty/variability factors. Are there additional data or analyses in the literature not cited here that can be used to strengthen the recommendations? Should other factors be considered in the application of uncertainty/variability factors?

Issue A: The Technical Panel felt strongly that if there is uncertainty in more than four areas of extrapolation (interspecies, intraspecies, LOAEL to NOAEL, subchronic to chronic, database deficiencies), it is unlikely that the database is sufficient to derive a reference value.

Recommendation A: *The Technical Panel recommends limiting the total UF applied to a chronic reference value for any particular chemical to 3,000. This maximum of 3,000 applies only to the UFs and does not include various duration adjustment factors. Setting a reference value when there is uncertainty in at least four areas would need to be carefully evaluated on a case-by-case basis.*

Issue B: The Toxicology Working Group of the 10X Task Force recommended that reduction of the intraspecies UF from a default of 10 be considered only if data are sufficient to support the conclusion that the dataset on which the POD is based is representative of the exposure/dose-response data for the susceptible subpopulation(s), including lifestages. The Technical Panel urges the development of data to support the selection of the appropriate size of this factor, but recognizes that often there are insufficient data to move away from the default. For example, identifying children as a susceptible subpopulation would not necessarily be sufficient to modify the intraspecies UF, because children of different ages or having other ethnic, racial, or SES characteristics may have differences in susceptibility. The most susceptible subpopulation would need to be thoroughly characterized to allow reduction of the UF.

Recommendation B: *The Technical Panel supports and expands the recommendation of the 10X Toxicology Working Group, i.e., that reduction of the intraspecies UF should be considered only if data are sufficient to support the conclusion that the data set on which the POD is based is representative of the exposure/dose-response data for the susceptible subpopulation(s). Given this, whether and how much the intraspecies UF may be reduced must be linked to how completely the susceptible subpopulation has been identified and its susceptibility described (e.g., versus assumed). At the other extreme, a 10-fold factor may*

sometimes be too small because of factors that can influence large differences in susceptibility, such as genetic polymorphisms. The Technical Panel urges the development of data to support the selection of the appropriate size of this factor, but recognizes that often there are insufficient data to support a factor other than the default.

Issue C: The exact value of the UF chosen should depend on the quality of the studies available, the extent of the database, and scientific judgment. Default uncertainty factors of 10 are recommended, with 3 used in place of half-power values (i.e., $10^{0.5}$) when occurring singly, when data are insufficient to support a data-derived value. The Technical Panel urges continued research and evaluation of the similarities and differences between the general population and susceptible subpopulations in their responses to particular agents, particularly children and the elderly. From such evaluations, the protectiveness of the 10-fold default factor can continue to be assessed.

Recommendation C: *Sound scientific judgment should be used in the application of UFs to derive reference values which are applied to the value chosen for the POD (NOAEL, LOAEL, or BMDL) derived from the available database. IRIS documentation should contain a justification for the individual factors selected for a particular chemical because rigid application of UFs could lead to an illogical set of reference values.*

Issue D: The Office of Pesticide Programs (OPP) recently published its guidance document on Application of the Appropriate FQPA Safety Factor(s) in Tolerance Assessment (2002b). In that document, OPP considered the FQPA factor to overlap with several of the traditional UFs, and to be in addition to the interspecies and intraspecies UFs. The traditional UFs that overlap with the FQPA factor are ones that account for data gaps (extrapolation from the LOAEL when a NOAEL is not available, extrapolation from a subchronic study to a chronic-exposure scenario when no chronic study data are available, and application of a database UF when there are gaps in the data considered essential for setting a reference value, including lack of data on children). The recommendation of the Technical Panel is in line with that of the 10X Task Force

Toxicology Working Group and indicates that the current traditional UFs will be adequate in most cases to cover concerns about children's health risks.

Recommendation D: *Given that there are several UFs that can be used to deal with data deficiencies as part of the current reference value process, and given that the FQPA safety factor is assumed to overlap to a large extent with these factors, the Technical Panel agrees with the 10X Task Force that the current interspecies, intraspecies, LOAEL-to-NOAEL, subchronic-to-chronic, and database deficiency UFs, if appropriately applied using the approaches recommended in this review, will be adequate in most cases to cover concerns and uncertainties about children's health risks. Any residual concerns about toxicity and/or exposure can be dealt with in risk characterization/risk management (e.g., by retention of all or part of the FQPA safety factor for pesticides).*

Issue E: The actual application of the MF in various IRIS files has been inconsistent and it has not been used frequently. Furthermore, most of the issues raised in the internal review for which an MF might be used, e.g., bioavailability, could be dealt with in the narrative description of the database, and/or other parts of the RfD/RfC derivation process.

Recommendation E: *The Panel considers the purpose of the MF to be sufficiently subsumed in the general database UF. Therefore, the Panel recommends the discontinuance in use of the MF.*

Issue F: The EPA has not yet established guidance for the use of specific data to replace UFs (i.e., chemical-specific adjustment factors, CSAFs), but the division of UFs into pharmacodynamic and pharmacokinetic components has been used in the RfC methodology (EPA, 1994).

Recommendation F: *The Agency is encouraged to develop its own guidance, based on some of the available methodologies (e.g., IPCS), but caution should be used in that there are relatively few data available for many substances to serve as an adequate basis to replace defaults with CSAFs.*

Appendix D

Agenda

United States
Environmental Protection Agency
Risk Assessment Forum

Peer Review Workshop on EPA's *A Review of the Reference Dose and Reference Concentration Processes*

Key Bridge Marriott
Francis Scott Key Ballroom
1401 Lee Highway
Arlington, VA 22209

Agenda

Workshop Chair: **Michael Dourson**
Toxicology Excellence for Risk Assessment (TERA)

W E D N E S D A Y , J U N E 1 9 , 2 0 0 2

8:00AM **Registration**

8:30AM **Welcome & Introductions** *David Bottimore, Versar, Inc.*

8:45AM **Chair's Introduction** *Michael Dourson, Workshop Chair*

9:00AM **Background** *Carole Kimmel, National Center for Environmental Assessment, U.S. Environmental Protection Agency*

9:15AM **Peer Reviewer Q&A Session on Background Presentation**

9:30AM **Discussion Session**

Charge Question #1 Please comment on the recommendation to derive less-than-lifetime reference values in addition to the chronic RfD and RfC? Is the need for these values adequately justified in the report?

Charge Question #3 Please comment on the recommendation that endpoint-specific reference values should not be derived.

Charge Question #5 Please comment on the recommendation to include duration-adjustment for inhalation developmental toxicity as for other health endpoints.

10:30AM B r e a k

10:45AM **Discussion Session (continued)**

Charge Question #2 Please comment on the revised definitions for reference values.

Charge Question #4 A review of current guideline study protocols and approaches to testing was conducted to determine what information is currently developed to support the derivation of less-than-lifetime reference values. Please comment on the life-stage approach taken in this review, as well as the recommendations for alternative testing approaches and strategies for developing a targeted testing strategy to support setting less-than-lifetime reference values as well as chronic values. Also, please comment on a proposal from the Office of Children's Health Protection to request a study by the National Academy of Sciences to take a fresh look at toxicity testing approaches and strategies based on this and other reports (e.g., ILSI, 2001).

12:00PM **Observer Comment Period**

12:15PM L u n c h (M i g h t B e W o r k i n g L u n c h)

1:15PM **Discussion Session (continued)**

Charge Question #6 Please comment on the recommendations in the report for using a weight-of-evidence approach for hazard characterization (Chapter 4, section C.2.), and for expanding characterization of the extent of the database and using a narrative to describe the database rather than a confidence ranking (Section C.3.). Is the discussion of weight of the evidence clear in terms of how it would be used in characterizing the database? Also, please comment on the consideration of this information in the derivation of reference values.

Charge Question #7 Please comment on the recommendations for use of an exposure-response array and derivation of sample reference values to select the point of departure (POD) for the final reference value. Also, please comment on consideration of the nature of the effects for different endpoints, durations, timing and routes of exposure in selection of the POD.

2:45PM B r e a k

3:00PM **Discussion Session (continued)**

Charge Question #8 Please comment on several recommendations concerning the application of uncertainty/variability factors. Are there additional data or analyses in the literature not cited here that can be used to strengthen the recommendations? Should other factors be considered in the application of uncertainty/variability factors?

4:00PM **Observer Comment Period**

4:15PM **Other Issues for Discussion**

4:45PM **Wrap-Up, Summary of Comments, and Next Steps**

5:00PM A d j o u r n

Appendix E

Presenter Overheads and Diagrams

Technical Peer Review Workshop On The Draft Document Entitled *A Review of the Reference Dose and Reference Concentration Processes*

June 19, 2002

David Bottimore
Versar, Inc.

Key Bridge Marriott
1401 Lee Highway
Arlington, VA

Overview of RfD/RfC Peer Review Workshop

- Welcome and Purpose
- Review of Agenda
- Introduction of Participants and Chair
- EPA Presentation Providing Background Information for Reviewers
- Chair - Mike Dourson - Groundrules, Peer Review Process
- Observer Comments
- Post Meeting Activities – Workshop Report
- Housekeeping

Technical Peer Review Workshop On The Draft Document Entitled *A Review of the Reference Dose and Reference Concentration Processes*

June 19, 2002

Michael Dourson - Workshop Chair

Chair's Opening Remarks

Goals for Meeting – Provide technical input to EPA on *A Review of the RfD and RfC Processes*, based on 8 charge questions

Peer Review Process

– Obtaining input on technical issues from experts in diverse specialties and from broad perspectives

– Focus on technical issues (not regulatory or policy)

– Not a consensus building process, individual comments

– Documentation of comments and recommendations

– Role of EPA in peer review meeting

Observer Comment Periods

Post Meeting Activities – Workshop report that summarizes comments and recommendations

Charge Questions

1. Please comment on the recommendation to derive less-than-lifetime reference values in addition to the chronic RfD and RfC? Is the need for these values adequately justified in the report?

Issue: The 10X Toxicology Working Group felt that data on developmental toxicity would often be a greater factor in calculating less-than-lifetime reference values, and that exposures to children are more often of this type. The RfD/RfC Technical Panel concurred with this view and expanded the types of data to be considered for different duration reference values to include other life stages as well.

Recommendation: *The Technical Panel concurred with the recommendation of the 10X Task Force that reference values should be derived, where possible, for acute, short-term, and longer-term durations, as well as chronic exposures for oral, dermal, and inhalation routes, and that they should be included in the IRIS database for use by EPA programs, where applicable. The Technical Panel recommended that the definitions for duration should be standardized but left flexible so they can be adjusted depending on the exposure situation of concern.*

Charge Questions (cont'd)

3. Please comment on the recommendation that endpoint-specific reference values should not be derived.

Issue: Given the recommendation for setting different duration reference values, the Technical Panel felt strongly that all relevant endpoints should be considered in the derivation of each duration- and route-specific reference value, thus ensuring that reference values are derived to be protective of all types of effects for that route and duration of exposure. The original recommendation for setting a RfD_{DT} (reference dose for developmental toxicity) was based on the fact that developmental toxicity did not require a chronic exposure scenario, which was the basis for the only values set at that time, the RfD and RfC. If the derivation of less-than-lifetime reference values that account for all forms of toxicity is adopted, there will be no need for endpoint-specific reference values, although this recommendation does not preclude such derivation for certain purposes, e.g., endpoints based on a common mode of toxicity for cumulative risk assessment.

Recommendation: *Endpoint-specific reference values should not be developed, including the RfD_{DT} (reference dose for developmental toxicity), as originally proposed in the Guidelines for Developmental Toxicity Risk Assessment (EPA, 1991).*

Charge Questions (cont'd)

5. Please comment on the recommendation to include duration-adjustment for inhalation developmental toxicity as for other health endpoints.

Issue: Duration adjustment for inhalation developmental toxicity studies is a notable exception to the practice of adjusting from intermittent to continuous exposures for inhalation toxicity studies. EPA's Guidelines for Developmental Toxicity Risk Assessment (1991) recommended against dosimetric adjustment on the basis that developmental effects were more likely to depend on peak exposure concentration. However, more recent information from the literature shows the relevance of area under the curve in several studies and supports the use of some type of adjustment, even for short half-life agents, such as ethylene oxide. Based on this information, together with the rationale used for dosimetric adjustment for other health effects (i.e., that exposure adjustment based on C × t tends to be more health protective), the Technical Panel has recommended that duration adjustment procedures from intermittent to continuous exposures be used for inhalation developmental toxicity studies as for other health effects from inhalation exposure. Of course, if specific data and/or models on pharmacokinetics or mode of action are available for determining the proper dose metric, these should be used instead of the default duration adjustment. The Panel urges continued development of data, modeling, and improved procedures for dose-duration adjustments related to developmental toxicity.

Recommendation: *Duration adjustment procedures from intermittent to continuous exposures for inhalation developmental toxicity studies should be done in the same way as for other health endpoints.*

Charge Questions (cont'd)

2. Please comment on the revised definitions for reference values?

Issue: Definitions were proposed for the acute, short-term, and longer-term reference values as well as revisions for the chronic RfD and RfC. The revised definitions (see below) are aimed at clarifying (a) that the reference value is intended to provide an estimate that is centered within an order of magnitude, further emphasizing that the estimate is not a bright line, but has some range of variability that may be considered by risk managers in decision making; (b) the term "deleterious" in the original definitions has been replaced with the term "adverse," because the latter is more commonly used and understood in data evaluation and selection of endpoints for setting reference values; and (c) the term "noncancer" has been removed from the original definitions in the spirit of harmonization of risk assessment approaches for human health effects because it has been recommended that health effects no longer be categorized as "cancer" or "noncancer" for the purposes of hazard characterization and dose-response analysis. This change denotes the move toward defining approaches for low dose estimation or extrapolation based on mode of action.

With this new set of definitions, standardizing the terminology used to refer to the reference values would help clarify the scope and purpose of each reference value in terms of route and duration of exposure.

Recommendations:

A. Use revised definitions for the reference values as follows.

Charge Questions (cont'd)

Acute [Oral, Dermal, Inhalation] Reference Value: An estimate of an exposure for 24 hours or less to the human population that is likely to be without an appreciable risk of adverse effects for a lifetime (including susceptible subgroups[a]). It can be derived from a BMD, a NOAEL or a LOAEL, with uncertainty/variability[b] factors generally applied to reflect limitations of the data used. The application of these factors is intended to provide an estimate centered within an order of magnitude.

Short-Term [Oral, Dermal, or Inhalation] Reference Value: An estimate of an exposure for up to 30 days to the human population that is likely to be without an appreciable risk of adverse effects for a lifetime (including susceptible subgroups). It can be derived from a BMD, a NOAEL or a LOAEL, with uncertainty/variability factors generally applied to reflect limitations of the data used. The application of these factors is intended to provide an estimate centered within an order of magnitude.

Longer-term [Oral, Dermal, or Inhalation] Reference Value: An estimate of an exposure for up to approximately 7 years (10% of the average life span) to the human population that is likely to be without an appreciable risk of adverse effects for a lifetime (including susceptible subgroups). It can be derived from a BMD, a NOAEL or a LOAEL, with uncertainty/variability factors generally applied to reflect limitations of the data used. The application of these factors is intended to provide an estimate centered within an order of magnitude.

Charge Questions (cont'd)

Chronic [Oral, Dermal, or Inhalation] Reference Value: An estimate of an exposure for up to the average life span of the human population that is likely to be without an appreciable risk of adverse effects for a lifetime (including susceptible subgroups). It can be derived from a BMD, a NOAEL or a LOAEL, with uncertainty/variability factors generally applied to reflect limitations of the data used. The application of these factors is intended to provide an estimate centered within an order of magnitude.

---

[a]*Susceptible subgroups may refer to life stages, e.g., children or the elderly, or to other segments of the population, e.g., asthmatics or the immune-compromised, but they are likely to be somewhat chemical-specific, and may not be consistently defined in all cases. See Chapter 4 (Section C.2.c) for further discussion.*

[b]*See discussion in Chapter 4 (Section D.5) on application of uncertainty/variability factors.*

B. Standardize terminology as much as possible, e.g., using a generic term such as reference value (RfV) with subscripts to designate route and duration of exposure, e.g., RfV_{AO} for acute oral reference value.

Charge Questions (cont'd)

4. A review of current guideline study protocols and approaches to testing was conducted to determine what information is currently developed to support the derivation of less-than-lifetime reference values. Please comment on the life-stage approach taken in this review, as well as the recommendations for alternative testing approaches and strategies for developing a targeted testing strategy to support setting less-than-lifetime reference values as well as chronic values. Also, please comment on a proposal from the Office of Children's Health Protection to request a study by the National Academy of Sciences to take a fresh look at toxicity testing approaches and strategies based on this and other reports (e.g., ILSI, 2001[3]).

Issue: The Technical Panel reviewed and evaluated current testing guidelines and approaches testing approaches as a follow-up to its recommendation concerning the derivation of less-than-lifetime reference values. This review was undertaken to determine what information is currently gathered with regard to life stage assessment, endpoint assessment, route and duration of exposure, and latency to response. A primary goal of this review was to provide a basis for recommendations for the development of innovative alternative testing approaches and the use of such data for risk assessment. The Technical Panel is not recommending additional testing for every chemical but is suggesting that alternative strategies and guidance for testing approaches be developed that incorporate information on pharmacokinetics and mode of action early in the process, thus allowing a more targeted testing approach. In addition, alternative protocols are discussed that are aimed at more efficient use of animals and resources in combined studies that would provide more extensive data on life stages, endpoints and other factors not well characterized in current testing approaches. Recommendations were also made about research areas that should be encouraged to aid in better study design and interpretation of data for risk assessment.

---

[3] http://www.ilsi.org/publications/pubslist.cfm?pubentityid=8&publicationid=305

Charge Questions (cont'd)

Recommendations:

A. Develop several new OPPTS guideline study protocols and modify current ones to collect more comprehensive data across life stages, route, duration and timing of exposures, that would be useful for setting acute and short-term reference values. Develop guidance for how and when to use the guidelines.

B. Develop additional guideline study protocols to evaluate several potential children's health issues, e.g., developmental immunotoxicity, carcinogenesis, more detailed neurotoxicity, pharmacokinetics, including direct dosing of neonates. Develop guidance for how and when to use the guidelines.

C. Encourage research to evaluate latency to effect and reversibility of effect from less-than-lifetime exposures. Encourage research on mechanisms/modes of action and pharmacokinetics at different life stages.

Charge Questions (cont'd)

6. Please comment on the recommendations in the report for using a weight-of-evidence approach for hazard characterization (Chapter 4, section C.2.), and for expanding characterization of the extent of the database and using a narrative to describe the database rather than a confidence ranking (Section C.3.). Is the discussion of weight of the evidence clear in terms of how it would be used in characterizing the database? Also, please comment on the consideration of this information in the derivation of reference values.

Issue: The 10X Task Force had recommended the use of a weight of evidence approach and considerations for level of concern in evaluating the data on children's health effects. These recommendations have been incorporated into the approach to hazard characterization proposed here for factors to be considered in a weight of evidence approach, and for the use of a narrative description rather than confidence rankings. The narrative approach is intended to emphasize the extent of the types of data available (both human and animal data) as well as the data gaps that could improve the derivation of reference values, and it should encourage a wider range of information to be used in deriving reference values, taking into consideration the life stages evaluated, the issues of timing, duration and route of exposure, the types and extent of endpoint assessment (i.e., structural and function), and the potential for latent effects and/or reversibility of responses. Factors to be considered in a weight of evidence evaluation both generally, and for characterizing potentially susceptible subpopulations, are described. The extremes for the extent of the database, i.e., minimal or robust, are defined in Chapter 4, but the Technical Panel did not define additional categories between minimal and robust and had serious concerns about developing such categories because of the tendency to try to characterize a database with single word descriptors, i.e., high, medium and low confidence. Instead, a narrative description of the extent of the database, with emphasis on the strengths and limitations of the data was strongly encouraged.

Recommendation: *An expanded approach to the evaluation of studies and characterization of the extent of the database as a whole is recommended; in particular, several factors are discussed that should be considered in a weight-of-evidence approach for characterizing hazard for the population as a whole as well as for potentially susceptible subpopulations. As part of this evaluation, a narrative approach should be used in describing the extent of the database instead of using a confidence ranking of high, medium, or low.*

Charge Questions (cont'd)

7. Please comment on the recommendations for use of an exposure-response array and derivation of sample reference values to select the point of departure (POD) for the final reference value. Also, please comment on consideration of the nature of the effects for different endpoints, durations, timing and routes of exposure in selection of the POD.

Issue: Currently, the "critical effect" is used as the basis for the POD, and various UFs are applied to the dose at the critical effect for derivation of the RfD or the RfC. The critical effect is defined as "the first adverse effect, or its known precursor, that occurs to the most sensitive species as the dose rate of an agent increases" (EPA, 2002a). The underlying assumption is that if the RfD or the RfC is derived to prevent the critical effect from occurring, then no other effects of concern will occur; in addition, this approach assumes that the relationship of various health effects for a particular chemical is maintained across species. The Technical Panel is concerned that presenting only a single critical effect and the critical study from which it was derived in the IRIS summary table that appears at the beginning of each RfD or RfC file may not provide enough information to the reader who is unfamiliar with risk assessment, and thus could be misleading. Presentation of a single endpoint as a POD for a systemic effect, for example, cannot capture the nature of the dose-response curve for that particular endpoint. Nor does the presentation of a single endpoint convey the possibility that other more serious endpoints may have a dose-response character markedly different from the less serious endpoint. Most importantly, in light of the Technical Panel recommendations for deriving an expanded number of reference values for different durations and routes of exposure, the limitations of focusing only on the critical effect become apparent because the most sensitive endpoint may be different for different durations or routes of exposure.

Recommendation: *An exposure-response array should be used as a visual display of all relevant endpoints and durations of exposure, as shown in the case study. This array can be used to evaluate the range of exposure-response data for different durations of exposure in order to determine the range of numerical values available for each route and duration of exposure. The POD should be selected on the basis of an evaluation of all relevant endpoints carried through to reference value derivation with selection of the limiting value(s) as the final step rather than on a single "critical study" and "critical effect."*

Charge Questions (cont'd)

8. Please comment on several recommendations concerning the application of uncertainty/variability factors. Are there additional data or analyses in the literature not cited here that can be used to strengthen the recommendations? Should other factors be considered in the application of uncertainty/variability factors?

Issue A: The Technical Panel felt strongly that if there is uncertainty in more than four areas of extrapolation (interspecies, intraspecies, LOAEL to NOAEL, subchronic to chronic, database deficiencies), it is unlikely that the database is sufficient to derive a reference value.

Recommendation A: *The Technical Panel recommends limiting the total UF applied to a chronic reference value for any particular chemical to 3,000. This maximum of 3,000 applies only to the UFs and does not include various duration adjustment factors. Setting a reference value when there is uncertainty in at least four areas would need to be carefully evaluated on a case-by-case basis.*

Charge Questions (cont'd)

Issue B: The Toxicology Working Group of the 10X Task Force recommended that reduction of the intraspecies UF from a default of 10 be considered only if data are sufficient to support the conclusion that the dataset on which the POD is based is representative of the exposure/dose-response data for the susceptible subpopulation(s), including lifestages. The Technical Panel urges the development of data to support the selection of the appropriate size of this factor, but recognizes that often there are insufficient data to move away from the default. For example, identifying children as a susceptible subpopulation would not necessarily be sufficient to modify the intraspecies UF, because children of different ages or having other ethnic, racial, or SES characteristics may have differences in susceptibility. The most susceptible subpopulation would need to be thoroughly characterized to allow reduction of the UF.

Recommendation B: *The Technical Panel supports and expands the recommendation of the 10X Toxicology Working Group, i.e., that reduction of the intraspecies UF should be considered only if data are sufficient to support the conclusion that the data set on which the POD is based is representative of the exposure/dose-response data for the susceptible subpopulation(s). Given this, whether and how much the intraspecies UF may be reduced must be linked to how completely the susceptible subpopulation has been identified and its susceptibility described (e.g., versus assumed). At the other extreme, a 10-fold factor may sometimes be too small because of factors that can influence large differences in susceptibility, such as genetic polymorphisms. The Technical Panel urges the development of data to support the selection of the appropriate size of this factor, but recognizes that often there are insufficient data to support a factor other than the default.*

Charge Questions (cont'd)

Issue C: The exact value of the UF chosen should depend on the quality of the studies available, the extent of the database, and scientific judgment. Default uncertainty factors of 10 are recommended, with 3 used in place of half-power values (i.e., $10^{0.5}$) when occurring singly, when data are insufficient to support a data-derived value. The Technical Panel urges continued research and evaluation of the similarities and differences between the general population and susceptible subpopulations in their responses to particular agents, particularly children and the elderly. From such evaluations, the protectiveness of the 10-fold default factor can continue to be assessed.

Recommendation C: *Sound scientific judgment should be used in the application of UFs to derive reference values which are applied to the value chosen for the POD (NOAEL, LOAEL, or BMDL) derived from the available database. IRIS documentation should contain a justification for the individual factors selected for a particular chemical because rigid application of UFs could lead to an illogical set of reference values.*

Charge Questions (cont'd)

Issue D: The Office of Pesticide Programs (OPP) recently published its guidance document on Application of the Appropriate FQPA Safety Factor(s) in Tolerance Assessment (2002b). In that document, OPP considered the FQPA factor to overlap with several of the traditional UFs, and to be in addition to the interspecies and intraspecies UFs. The traditional UFs that overlap with the FQPA factor are ones that account for data gaps (extrapolation from the LOAEL when a NOAEL is not available, extrapolation from a subchronic study to a chronic-exposure scenario when no chronic study data are available, and application of a database UF when there are gaps in the data considered essential for setting a reference value, including lack of data on children). The recommendation of the Technical Panel is in line with that of the 10X Task Force Toxicology Working Group and indicates that the current traditional UFs will be adequate in most cases to cover concerns about children's health risks.

Recommendation D: *Given that there are several UFs that can be used to deal with data deficiencies as part of the current reference value process, and given that the FQPA safety factor is assumed to overlap to a large extent with these factors, the Technical Panel agrees with the 10X Task Force that the current interspecies, intraspecies, LOAEL-to-NOAEL, subchronic-to-chronic, and database deficiency UFs, if appropriately applied using the approaches recommended in this review, will be adequate in most cases to cover concerns and uncertainties about children's health risks. Any residual concerns about toxicity and/or exposure can be dealt with in risk characterization/risk management (e.g., by retention of all or part of the FQPA safety factor for pesticides).*

Charge Questions (cont'd)

Issue E: The actual application of the MF in various IRIS files has been inconsistent and it has not been used frequently. Furthermore, most of the issues raised in the internal review for which an MF might be used, e.g., bioavailability, could be dealt with in the narrative description of the database, and/or other parts of the RfD/RfC derivation process.

Recommendation E: *The Panel considers the purpose of the MF to be sufficiently subsumed in the general database UF. Therefore, the Panel recommends the discontinuance in use of the MF.*

Issue F: The EPA has not yet established guidance for the use of specific data to replace UFs (i.e., chemical-specific adjustment factors, CSAFs), but the division of UFs into pharmacodynamic and pharmacokinetic components has been used in the RfC methodology (EPA, 1994).

Recommendation F: *The Agency is encouraged to develop its own guidance, based on some of the available methodologies (e.g., IPCS), but caution should be used in that there are relatively few data available for many substances to serve as an adequate basis to replace defaults with CSAFs.*

Review of the RfD/RfC Process

Carole Kimmel, Chair
RfD/RfC Technical Panel
External Peer Review
June 19, 2002

RfD/RfC Review - Background

- Initiated in response to a charge from the EPA 10X Task Force that dealt with implementation of an additional safety factor to protect children's health under FQPA
- Original charge was to examine the RfD/RfC process with regard to protecting children's health
- Charge was expanded to include a more in-depth review of the process for setting reference values

RfD/RfC Review -Purpose

- Evaluate the current state-of-the-art for hazard and dose-response assessment with a focus on protection of potentially sensitive subpopulations
- Summarize what additional scientific issues can bring to the process
- Raise issues that should be further explored or developed for consideration in the process
- Recognize that the process should not be static, but continually evolving and that new information should be incorporated as new RfDs/RfCs are set, or as they are re-evaluated

RfD/RfC Review - Purpose

- Review – NOT a guidance document
- Recommendations for changes in current process and/or need for new guidance
- Two underlying themes
  - Harmonization of health risk assessment procedures
  - Making efficient use of resources and animal testing

RfD/RfC Review Recommendation 1

- Derive reference values for acute, short-term, longer-term, and chronic durations for oral, dermal, and inhalation exposures, where appropriate
- Make available through IRIS
  - Less-than-lifetime values already set by several program offices and other agencies

RfD/RfC Review Recommendation 1, con't

- Standardize durations, leave flexible for adjustment to the exposure scenario of concern
  - Acute – 24 hrs or less
  - Short-term – up to 30 days
  - Longer-term – up to approximately 7 years (10% of the average lifespan)
  - Chronic – up to the average lifespan

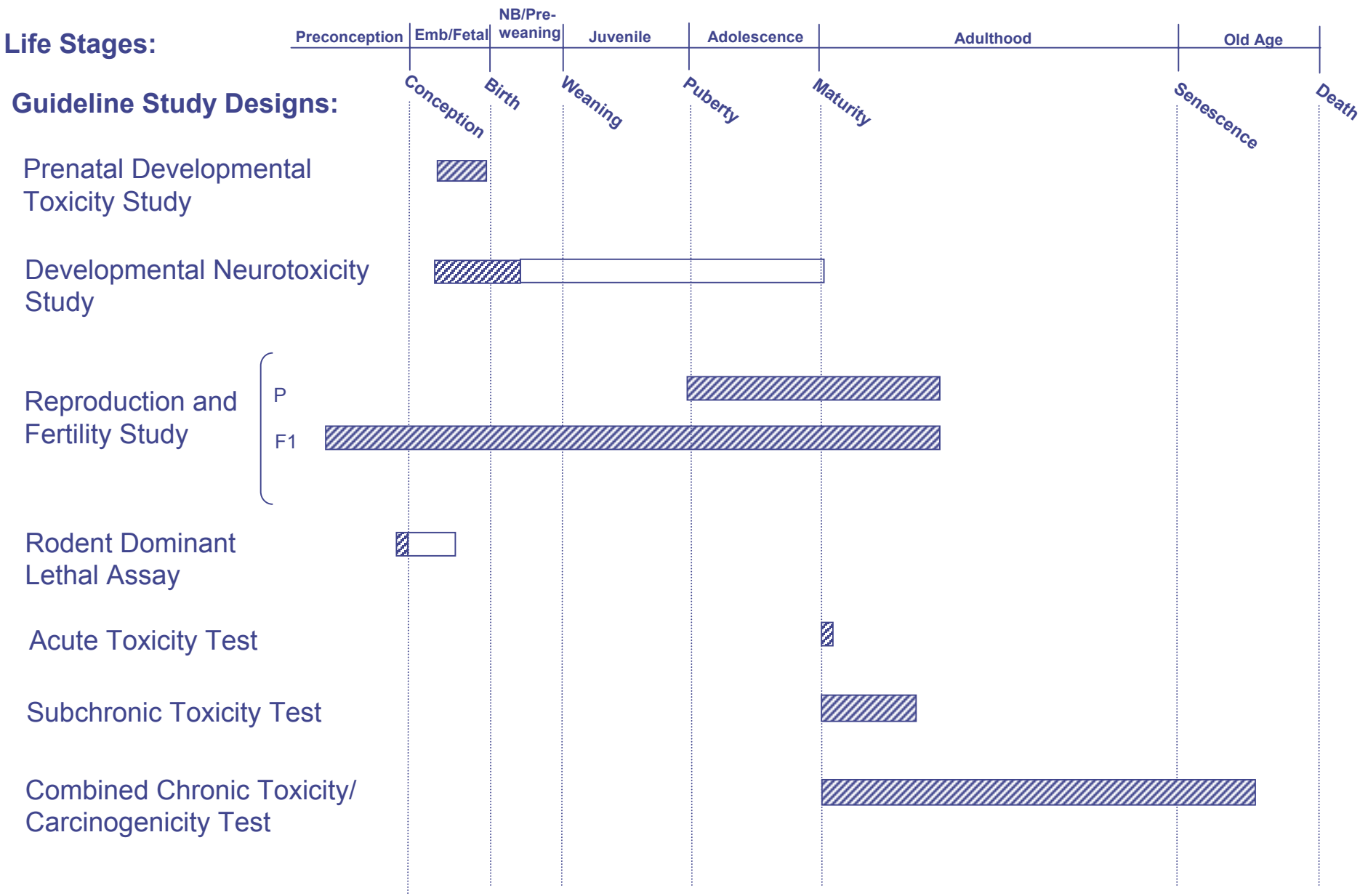
RfD/RfC Review Recommendation 2

- Use revised definitions for reference values
  - An estimate of an exposure for (duration) to the human population that is likely to be without an appreciable risk of adverse effects for a lifetime (including susceptible subgroups). It can be derived from a BMD, a NOAEL or a LOAEL, with uncertainty/variability factors generally applied to reflect limitations of the data used. The application of these factors is intended to provide an estimate centered within an order of magnitude.
- Standardize terminology, e.g., RfV_{AO} or RfD_A = acute oral reference value or dose

RfD/RfC Review Recommendation 3

- Endpoint-specific reference values should not be developed
  - E.g., Reference dose for developmental toxicity (RfD_{DT})
  - Reference values should cover all relevant endpoints

A Life Stage View of Timing and Duration of Exposure in Standard Toxicity Testing Protocols



RfD/RfC Review Recommendation 4

- Develop several new OPPTS guideline study protocols to guide collection of more comprehensive data across life stages, routes, durations, and timing of exposures
  - Acute and short-term durations
  - Developmental issues (immunotoxicity, carcinogenicity, neurotoxicity, direct dosing of neonates, endocrine disruptors)
  - Aging and toxicity issues
  - Pharmacokinetics at several life stages
  - Latency and reversibility
- Develop guidance for how and when to use
- OCHP – Consider request to NAS for review of testing strategies using current technologies

RfD/RfC Review Recommendation 5

- Duration adjust for inhalation developmental toxicity data as for other health endpoints
  - Notable exception to current practice
  - Some data available to support the need for adjustment
  - Use specific data and/or models, if available

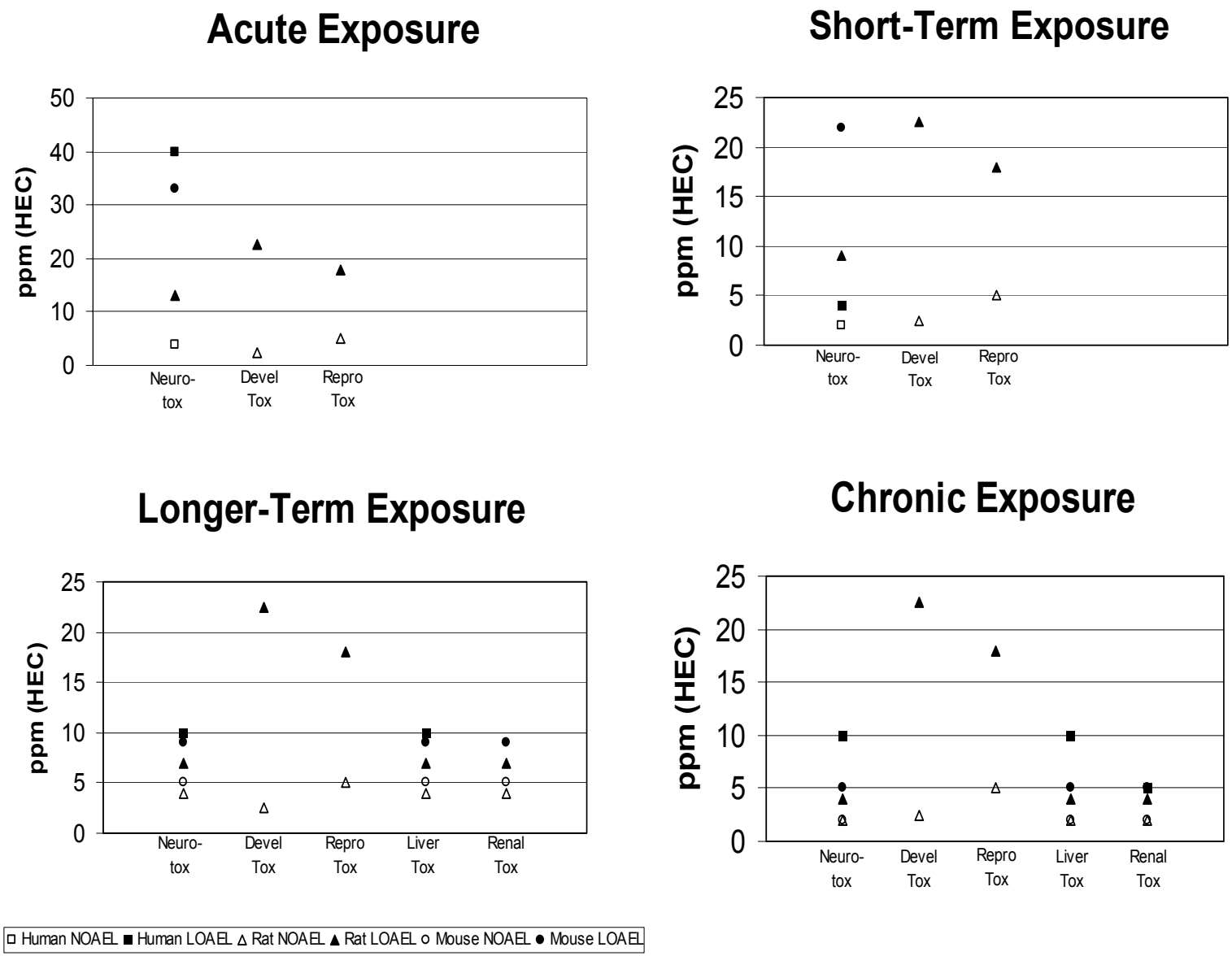
RfD/RfC Review Recommendation 6

- Expand characterization of the database using WOE characterization
  - Encourages consideration of a wider range of information, including extent of data on life stages, timing, duration, and route of exposure, types & extent of endpoint assessment (structure and function), latency & reversibility of responses
- Use a narrative to describe the extent of the database; extremes, i.e., minimal and robust, are described in the document

RfD/RfC Review Recommendation 7

- Use exposure-response array and derivation of sample reference values to select POD
  - Encourages consideration of all relevant endpoints, durations and routes of exposure, rather than a single "critical study" and "critical effect"
  - Does not preclude (enables) consideration of specific endpoints for certain purposes, e.g., cumulative risk based on same mode of action

Example: Exposure-Response Array for Inhalation Exposure to Chemical X



Example: Derivation of Sample and Final Longer-Term Inhalation Reference Value for Chemical X

Exposure Duration	HEC (ppm)	Species	Type of Effect	Uncertainty Factors						Reference Value (ppm)	
				Total	A	H	L	S	D	Sample	Final
Longer-term	20L*	Human	NT	300	1	10	10	1	3	0.07	0.03
	4	Rat	NT	100	3	10	1	1		0.04	
	5	Mouse	NT	100	3	10	1	1		0.05	
	2.5	Rat	DT	100	3	10	1	1**		0.03	
	5	Rat	RT	100	3	10	1	1		0.05	
	4	Rat	LT	100	3	10	1	1		0.04	
	5	Mouse	LT	100	3	10	1	1		0.05	
	4	Rat	KT	100	3	10	1	1		0.04	
	5	Mouse	KT	100	3	10	1	1		0.05	

*LOAEL. **Duration uncertainty factor not applied; should be considered further.

RfD/RfC Review Recommendation 8

- Application of Uncertainty/Variability Factors
  - A. Limit the total UF to 3000 (not including adjustment factors)
  - B. Consider reduction of the intraspecies UF from a default of 10 only if data are sufficiently representative of the exposure/dose-response data for the most susceptible subpopulations; acknowledge that a 10-fold factor may sometimes be too small
  - C. Base determination of the size of UFs applied on available scientific data in deriving reference values (e.g., size of intraspecies UF); include justification of UFs applied

RfD/RfC Review Recommendation 8, con't

- Application of Uncertainty/Variability Factors
  - D. The current interspecies, intraspecies, LOAEL-NOAEL, subchronic-chronic, and database deficiency UFs, if appropriately applied, will be adequate in most cases to cover concerns and uncertainties about children's health risks
  - E. Discontinue use of the modifying factor (MF)
  - F. Develop Agency guidance on the use of Chemical-Specific Adjustment Factors (CSAF)

Diagram Provided by Dr. Byczkowski
During Discussion of Charge Question #5

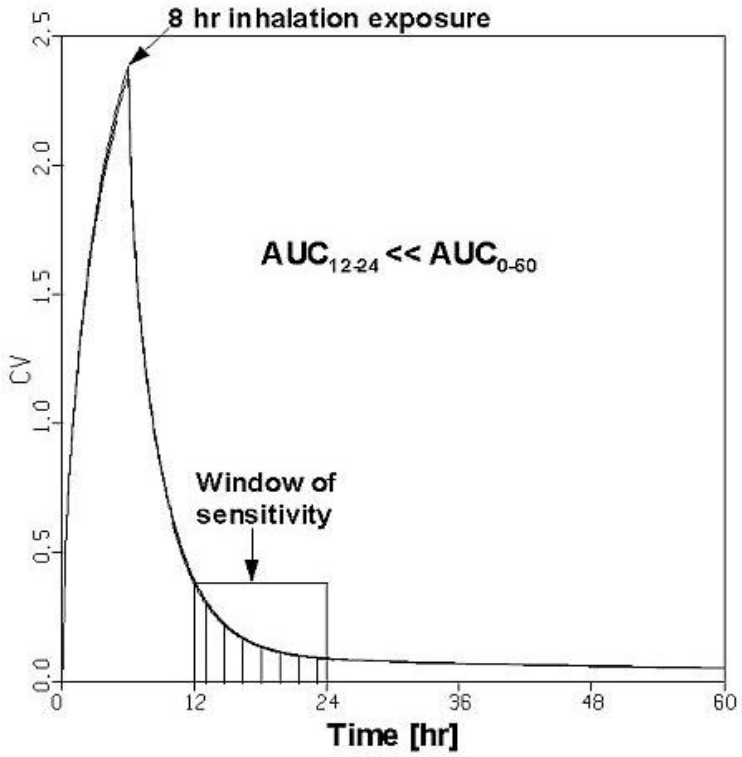


Fig. 1. The graph representing concentration of the volatile chemical in maternal blood (CV in mg/L) over time (hr) after a typical inhalation exposure for 8 hr. The reference value for developing fetus, extrapolated from the inhalation exposure, may be not protective when a large dose corresponding to the total area under the curve (AUC0-60) is used in calculations, instead of the much lower dose corresponding to a short window of sensitivity (AUC12-24) during organogenesis.

Appendix F

Written Comments from Peer Reviewers

Written Comments from Peer Reviewers

Janusz Byczkowski
Independent Consultant

Kenny Crump
Environ Corp.

George Daston
Proctor & Gamble Company

Rodney Dietert
Cornell University

John Doull
University of Kansas School of Medicine

Michael Dourson
Toxicology Excellence for Risk Assessment

Pat McGinnis
Syracuse Research Corp.

Bonnie Ransom Stern
BR Stern Associates

Pam Shubat
Minnesota Department of Health

Ellen Silbergeld
Johns Hopkins University

Alan Stern
New Jersey Department of Environmental Protection

Lauren Zeise
California Environmental Protection Agency

Review by
Janusz Byczkowski
Independent Consultant

General Impressions

The reviewed document is a clear and relatively well written summary of the problems and the recommended possible solutions to pitfalls in the current process of toxicity values derivation, including some implications of the Food Quality Protection Act (FQPA).

This reviewer has somewhat ambivalent feelings about practical implications of some recommendations in this document. On one hand, recommendations for the science-based and more realistic than before approach to the derivation of reference values (RfD and RfC) are laudable. Especially, the focus on mode of action, shape of dose-response relationship, and temporal considerations may add a much needed realism to the risk assessment process. On the other hand, under the existing practice of chemical risk assessment, adding into this process an array of route- and time-dependent reference values, or even more troublesome - their probabilistic distributions, may produce a further confusion among some of the practitioners involved in both risk assessment and management.

One of the documents' recommendations is to include the acute, short-term, longer-term and chronic reference values in the Integrated Risk Information System (IRIS) data base. Considering separate values for each route of exposure, this may add a significant amount of numerical data to the data base, which probably would be welcomed by the experts but hated by those without extensive training in toxicology. As stated on the IRIS Web site (http://www.epa.gov/iris/index.html): "*...The information in IRIS is intended for those without extensive training in toxicology, but with some knowledge of health sciences...*" Thus, without re-defining the audience targeted by this additional amount of numerical values, it may increase the level of misunderstanding and aggravate further confusion. One possible solution to this problem may be the addition of a next layer to the IRIS data base, designed for risk assessors and managers appropriately trained in the toxicology. This additional IRIS data base layer could contain also a detailed information on dose-response assessments, including the life stages.

Response to Charge Questions

1. *Comments on the recommendation to derive less-than-lifetime reference values in addition to the chronic RfD and RfC. Is the need for these values adequately justified in the report?*

The need for less-then-lifetime reference values, in addition to existing RfD/RfC, is adequately justified and discussed appropriately in the report.

It may be worth to advise users of the expanded toxicity values data base (risk assessors and managers) how these less-than-lifetime reference values may be applicable to the typical exposure scenarios (e.g., residential, recreational, trespasser, commercial worker,

construction worker, etc).

2. *Comments on the revised definitions for reference values.*

A. The revised definitions of Reference Values are clear and contain all necessary elements.

Emphasizing variability "*within an order of magnitude*" seems to be very appropriate.

The usage of the term "*adverse*" instead of "*deleterious*" really does not change the meaning of the reference value and both terms seem equally appropriate.

Deemphasizing "*cancer vs*. *noncancer*" endpoints seems justifiable, and probably it is long overdue.

B. The standardization of terminology is always advisable, however, addition of so many subscript characters to the "*RfV*" symbol may produce some confusion, unless the sub-scripted modifiers are always defined in a footnote or a glossary in any written document quoting them.

3. *Comments on the recommendation that endpoint-specific reference values should not be derived*.

While it seems reasonable to consider all relevant endpoints in derivation of time- and route- specific reference values, so the most sensitive health effect could be selected, the information about this specific selected endpoint should always accompany the reference value. This will help to prevent an inappropriate application of a wrong reference value to some irrelevant exposure scenario. For example, application of a short term reference value for developmental toxicity to the short term construction worker exposure scenario.

The information about the selected one and all other relevant endpoints would be crucial in cumulative risk assessment.

4. *Comments on the life-stage approach taken in this review, as well as the recommendation for alternative testing approaches and strategies for developing a targeted testing strategy to support setting less-than-lifetime reference values as well as chronic values.*

The recommendations on the life-stage approach in the testing strategy, seems very appropriate and reasonable. Especially, incorporation of pharmacokinetics and mode of action, relevant to the specific sensitive period, will be crucial, right at the study design step.

Comments on a proposal from the Office of Children's Health Protection to request a study by the NAS to take a fresh look at toxicity testing approaches and strategies based on this and other reports.

It seems that the review by the National Academy of Science of the current toxicity testing approaches and their strategy would be helpful.

5. *Comments on the recommendation to include duration-adjustment for inhalation developmental toxicity as for other health endpoints.*

For most therapeutic effects an area under the curve (of the internal concentration over time; AUC) is typically used to determine the internal dose equivalency by the clinical pharmacologists. The same method should be applicable, in general, to the toxic effects. The obvious exception to this general recommendation are those fast cleared developmental toxicants that affect a specific narrow window of susceptibility, for example, a teratogenic effect of a chemical applied during the sensitive period of fetal organogenesis.

It seems that for the chemicals with unknown pharmacokinetics and/or mode of action, no inhalation developmental toxicity values should be calculated, pending the development of the appropriate and sufficient data. Thus, the duration-adjustment could be unnecessary.

6. *Comments on the recommendations in the report for using a weight-of-evidence approach for hazard characterization (Chapter 4, section C.2.), and for expanding characterization of the extent of the database and using a narrative to describe the database rather than a confidence ranking (Section C.3.). Is the discussion of weight of the evidence clear in terms of how it would be used in characterizing the database? Consideration of this information in the derivation of reference values.*

While the confidence ranking is a nightmare for those who have to perform it, it is a blessing for those who can use it. Under the current practice of hazard characterization and risk assessment, in general, the confidence ranking involves a lot of subjective judgement regarding the quality of experimental or epidemiological evidence. At the other end of the risk assessment process, the end-user often working under the narrow time constrains, receives a ready-made grade (high, medium or low confidence) without need to analyze the evidence. This approach has been highly enhanced by the current managerial trend to perform “streamlined risk assessments” and to provide a “quick documentation turnover”.

The weight-of-evidence approach, which seems to be a very appropriate from the point of view of scientific objectivity as discussed in the report, may be not well received by the risk assessment practitioners as it shifts at least in part the need to read, to understand and to analyze the information from the data base developer to the end-user (often “*without extensive training in toxicology*”).

7. *Comments on the recommendation for use of an exposure -response array and derivation of sample reference values to select the point of departure (POD) for the final reference value. Comments on consideration of the nature of the effects for different endpoints, durations, timing, and routes of exposure in selection of the POD.*

As with the previous issue (see above, point 6) the use of the exposure-response array, which seems to be a very appropriate from the point of view of scientific objectivity and will be much more realistic than the selection of a single "*critical effect*" from the "*critical study*", but it also would imply that the end-user of the data base must have an extensive training in toxicology.

The idea of providing a visual display of all relevant endpoints and durations of exposure used is a very good idea (analogous to the ATSDR displays in the toxicity profiles) as it would aid in comprehending the relationship between different toxic values.

8. *Comments on several recommendations concerning the application of uncertainty/variability factors. Are there additional data or analyses in the literature not cited here that can be used to strengthen the recommendations? Should other factors be considered in the application of uncertainty/variability factors?*

A. This reviewer agrees that in the cases where uncertainty covers too many areas of extrapolation, no reference value should be calculated, pending the development of the appropriate and sufficient data. A case-by-case evaluation is a correct approach.

B. To link the intraspecies uncertainty factor (UF) reduction from the default with the identification of the appropriate susceptible subpopulation is a reasonable approach. In any case, there should be enough data developed to support the selection of the UF value.

C. The recommendation to apply a "*sound scientific judgement*" in the selection of the UF value seems very appropriate. This reviewer agrees that a specific justification should be provided for each factor involved in the selection of the UF value for a particular chemical.

D. Applying additional x10 safety factor as a default on the top of existing UFs does not sound like a reasonable idea. This reviewer agrees, that the UFs applied appropriately, in accordance to the recommendations in this report, should be sufficiently protective to children and would cover most of the data gaps and known deficiencies.

E. This reviewer agrees that the modifying factors (MF) can be eliminated.

F. This reviewer agrees that chemical-specific adjustment factor (CSAF) replacement of the default UFs may be possible only when sufficient data exist to support such a replacement.

Specific Observations

Change of IRIS definition of "Toxicology" (page G-10, line 18).

Traditionally, "Toxicology" has been defined as a "*science of poisons and poisonings*" and deals with chemical hazards.

This reviewer does not see a compelling reason why physical and some biological agents should be added to the domain of toxicology, as they are already covered by such disciplines as Health Physics, Hygiene, Clinical Microbiology, etc. It would be ridiculous to call, for example, a physical agent known to produce adverse health effects - noise - "*a poison*". Similarly, an act of listening to the heavy metal band cannot be called "*a poisoning*".

It seems that the current IRIS definition of Toxicology is sufficient as it covers harmful (or adverse) effects of chemical hazards on biological systems and does not trespass into the domains already covered by other disciplines.

Review by
Kenny Crump
Environ Corp.

Comments on "A Review of the Reference Dose and Reference Concentration Processes"
EPA/630/P-02/002A

Kenny Crump
Environ Corporation
602 East Georgia Avenue
Ruston, LA 71270
kcrump@environcorp.com

June 2002

Charge question #1
Having standardized procedures for computing less than lifetime values and to report such values in IRIS would seem to be a worthy goal. The document describes various procedures used presently by different offices to derive similar values, but does not provide convincing evidence that these values are serving a useful need, or that the recommended less-than-lifetime reference values are needed. To more fully support the need for such factors, some discussion of their use should be provided.

Charge question #2
The specific exposure durations used in the definitions appear reasonable. Their usefulness will depend on how well they mimic exposure situations occurring in practice.

I think it is ill-advised to include in the definition the statement "The application of these factors is intended to provide an estimate centered within an order of magnitude." First of all, it is not clear what this statement means. It will probably be interpreted somewhat like a confidence interval as a measure of the uncertainty of an estimated value. But it makes no sense to assign a generic uncertainty bound around a value that takes no account at all of the data that went into its calculation. Furthermore, given the non-specific nature of the definition of the RfV ("likely to be without appreciable risk"), there would be no way even with data to determine a standard uncertainty bound for the RfV, such as a statistical confidence bound. If it is felt that some sort of statement regarding the uncertainty in RfV is needed, how about something like this: "In most instances, due to uncertainty in the underlying data and the mechanisms by which exposure causes disease, it will not be possible to say with certainty that exposures at or below the RfV are without risk, or that exposures moderately above the RfV pose any appreciable risk."

The proposal to standardized terminology, using subscripts to indicate route and duration of exposure, seems fine.

Charge question #3
I agree that should the proposed reference values for different exposure durations be developed, there should be no need for endpoint-specific reference values.

Charge question #4

There is probably a need for a thorough overall review of testing guidelines that considers data needs to support specific RfVs. The Technical Panel has taken a useful first step in evaluating current testing guidelines and pointing out potential gaps, and suggesting some modifications that could help to close these gaps. However, this is a very complicated issue that needs a great deal of planning and thought. Expanding testing guidelines without giving specific guidance as to when various tests should be conducted and how the results are to be used would likely add confusion to the RfV process as well as being wasteful of resources. It must be kept in mind that the tests are not being conducted in the species of interest, and the value of data collected on different life stages, exposure patterns, etc. is lessened by the uncertainties in species to species extrapolation. Before embarking on guidelines for a highly expanded testing protocol, it would be helpful to see some evaluation of data from well-studied chemicals that already have much of the data that might be specified in such guidelines. Such an evaluation could determine how the additional data would affect RfVs, which data was most useful, and whether RfVs based on current guidelines would not have been health protective.

Charge question #5

I see no conceptual reason for not adjusting for duration in developmental effects as proposed by the Technical Panel, as is done for other endpoints or routes. This is not meant to imply that this should be done rote, without consideration of the special nature of developmental toxicity.

Charge question #6

In general I support the idea of a narrative, weight-of-evidence approach to hazard characterization and characterization of the data base. Moreover, I see no particular value in the descriptor "robust" since it has no operational significance that I am aware of (in contrast to the "minimal" descriptor, which is proposed for determining whether a data base is suitable for supporting an RfV). It seems unlikely that a chemical would achieve this standard. And what if it did? Would this mean that no additional study would be useful?

More thought needs to be given to the specifics in this section. Several of the questions on page 4-19 for guiding the extent of the data base are inappropriate. E.g., "Are the metabolism and pharmacokinetics in the animal species similar to those of humans?" is perhaps useful for hazard characterization, but not for evaluating the extent of the data base. The same is true for other questions.

Charge question #7

I agree that with carrying calculations for a number of endpoints and durations of exposure through down to the final step, for all the reasons cited in the report. In addition, it is not always apparent what the "critical effect" is when derived from a modeling exercise such as benchmark modeling, before the analysis is performed.

Charge question #8

In addition to the factors discussed in the report, I recommend consideration of a factor to reflect seriousness of health endpoint upon which the RfV is based. We never know for sure what the dose threshold is, or even if there is one. Therefore, we must view seriously the possibility of some residual risk at the RfV. As a result, it is appropriate to provide greater margins of safety for more severe effects. This recommendation is in keeping with the recommendation presently in the document to carry all potential POD to the last step in the calculation. At the last stage we could then apply a "severity factor" to each potential POD. It could well be that larger potential POD derived from a more severe effect would become the POD after the severity was accounted for. This approach would also be helpful in the harmonization of cancer (a severe endpoint) with non-cancer.

Recommendation A: A maximum of 3,000 for uncertainty factors seems reasonable, with the understanding that it does not apply to other areas. E.g., if the severity factor recommended above were adopted, I would recommend that it not be subject to the 3,000 limit.

Recommendation B: I think more latitude should be given for reducing the intraspecies uncertainty factor than is presently proposed. Consider the example presented of children having different racial, ethnic or SES characteristics that *may* correspond to differences in susceptibility. We will never have data to cover all the *potential* nuances in susceptibility. Suppose, e. g., there is a well-conducted study among children of a particular SES or racial group. An analyst needs to have the latitude to consider the plausibility that a different SES or racial group could be ten times more susceptible, i.e., needs to have greater latitude for modifying the default 10-fold factor.

Recommendation C: Seems OK, although I recommend making $10^{1/2} = 3.2$, rather than 3. The extra precision can be important in some cases, and it seems more appropriate to slightly overstate, rather than understate, the factor.

Recommendation D: I agree with the Technical Panel that the current interspecies, intraspecies, and database deficiency UFs, if appropriately applied using the approaches recommended in this review, will be adequate in most cases to cover concerns and uncertainties about children's health risks.

Recommendation E: I agree with the discontinuation of the MF.

Recommendation F: I agree.

Additional Comments

General comment: There are many recommendations in the document for using scientific judgment. More attention needs to be given in the document to ensuring there is consistency in application of scientific judgment, and consequently in RfVs for different substances.

Page 2.2 "For any approach used, the preferred adjustment procedure is to use a pharmacokinetic model, if available."

"Use a pharmacokinetic model" is not an "adjustment procedure". There are many ways to use a pharmacokinetic model to adjust AREs to different durations.

Page 4-3 Any exposure duration longer than 10% of lifespan is considered "chronic"? Does this mean that a 91-day exposure in rats would be considered chronic?

Page 4-9 Recommend deleting the phrase "Precedence is given to biological significance" and continuing with the remainder of the sentence. This phrase could be interpreted to mean that if an effect is biologically significant, but not statistically significant (e.g., not clearly dose-related), that it would still be considered as an adverse effect of exposure.

Page 4-7 Somewhere in the discussion of the power of a study, insert the sentences: "Once a study has been completed, the concept of "power" can often be better quantified in terms of the statistical confidence intervals for the response, rather than through a formal power calculation. E.g., if an epidemiological study found no significant increase in relative risk of lung cancer, but the 95% upper confidence bound on the relative risk was 10, then this study obviously had little power to detect an effect. "

Page 4-7 More thought perhaps needs to be given to the list of questions to consider in evaluating studies. I suspect that epidemiologists have somewhere come up with a more specific and useful list of things to look for that what is listed here. Although it would be useful to have a clearly delineated hypothesis before embarking upon a study, I don't see why that should be an issue in evaluating the usefulness of a study for risk assessment once it is completed; the data then speak for themselves. "Biological plausibility" actually encompasses several other questions. A BMD lower confidence bound would seldom be established in a published study. Perhaps a more pertinent question would be concerning whether the necessary data for quantitative modeling, including BMD calculations, are provided.

Page 4-10 Suggest changing "nor does it imply an averaging of the doses or exposures identified in individual studies that may be suitable as PODs for risk assessment." to "nor does it imply an averaging of the candidate PODs derived from different studies."

Page 4-24 The validity of the C x t paradigm will depend upon the exposure level in some cases. Although I am sure the Technical Panel is aware that twenty-four times a safe daily exposure may not be a safe one-hour exposure, this needs more careful discussion in the report.

Page 4-27 "The optimal approach for extrapolating from one dose-duration response situation to another is with the use of a physiologically-based pharmacokinetic model (PBPK) model." is incomplete. The availability of a valid PBPK model doesn't obviate the need for additional assumptions (concerning averaging time, critical metabolite, etc.)

to extrapolate across durations.

Page 4-35 It would be inconsistent to apply a factor to a BMD to "reflect a more appropriate level of risk", without applying a similar factor to NOELs. E.g., the Allen *et al.* study indicated that BMDs corresponding to a 10% increased risk are generally smaller than NOELs.

It defeats one of the advantages of the BMD over the NOEL to use different levels of risk to define the BMD depending upon the strength of the study. E.g., if a smaller risk level is used for a very large study, because of the study's greater sensitivity to detect effects at lower doses, unless some counteracting adjustment is made, a larger study will tend to provide a lower RfV. This will occur for a reason that has nothing to do with the toxicity of the chemical, but only because a larger study happened to be available. Default risk levels for calculating BMDs should be developed by the Agency. Further, it should be recommended that any deviations from the default should be accounted for, e.g., in the selection of uncertainty factors.

Page 4-42 This discussion should include comments on BMD calculations. A BMD can be calculated even when a NOEL does not exist.

Post Meeting Additions

I want to modify and expand on my pre-meeting comments in a couple of areas.

First, after hearing the discussion about the potential value to users of the confidence classifications now in IRIS, I feel less comfortable with my agreement with the recommendation that these be omitted in favor of a narrative presentation. I gathered from the discussion that they are helpful to users who have minimal toxicological expertise, so it now appears to me that both a narrative and a simple classification scheme might be a better approach.

Second, I wanted to expand upon my idea for a modifying factor to account for the severity of the endpoint. I spent a few minutes looking through IRIS today to see what endpoints RfVs are based upon. The second chemical I looked at (Acifluorfen, sodium (CASRN 62476-59-9)) reported "compound-related mortalities" at the LOEL of 500 ppm. So certainly not all RfDs are based on very mild effects, as members of the Review Panel seemed to believe. My proposal is to use a modifying factor to account for varying severities of different endpoints. To do this would require grouping effects into bins (e.g., minimal, moderate, severe, lethal), and assigning of a default severity factor to each bin. If a potential POD was based on an effect more severe than minimal, an additional factor would be applied (assuming that the default factor for minimal severity would be 1). The RfV process would otherwise be the same. Note that it could happen that if severe effects did not occur at the NOEL but at doses only moderately higher than

the NOEL, then the resulting RfD might be based on the more severe effect, although the resulting RfV would be lower because of the modifying factor. Regarding the binning of effects, there are some results in the literature that could be a starting point. Categorical regression, which has been endorsed by EPA, requires such binning of endpoints.

An added value of this approach is that it would conceptually provide a way for cancer and non-cancer to be harmonized, which is one of the stated goals of the Technical Panel. All that would be needed would be to assign a modifying factor for cancer -- possibly different modifying factors to account for different cancer mechanisms (e.g., linear, non-linear).

Review by
George Daston
Proctor & Gamble Company

General Impressions:

My overall impression of the review is very positive. The RfD/RfC Technical Panel is made up of some of EPA's most experienced (and wisest) risk assessors. They understand the Agency's risk assessment procedures intimately; they clearly know what works well and what hasn't. They have nicely addressed the request from EPA's 10X Task Force to evaluate the adequacy of the process for protecting children's health. In fact, they (rightly) identified that the charge was too narrow and have taken a comprehensive look at the ability of the RfD/RfC process to protect all life stages, not just childhood. The result is an extensive, critical review of the process, with a number of pragmatic, actionable recommendations for improvement.

The four most significant contributions that the Review makes are 1) paving the way for a new way of thinking about risk assessment that explicitly addresses differential sensitivity of different life stages; 2) proposing a number of less-than-lifetime reference values; 3) continuing the effort to harmonize and improve cancer and non-cancer risk assessment processes; and 4) recognizing the progress of the science of toxicology, and applying relevant advances to the risk assessment process.

It's obvious that people of different ages may have different responses to their environment: we understand this from personal experience and, more objectively, from scientific and medical observation. The most dramatic differences occur early in life, particularly in the embryo, fetus, infant and child. Limited, even single, exposures during embryonic stages can lead to irreversible alterations in structure and function that are unique to this life stage. Short-term exposures in the fetus or child may also produce persistent alterations in function that are also unique to the life stage. Failure to recognize the uniqueness of these life stages has led to some of our greatest public health tragedies, including the thalidomide episode. It's no wonder that developmental toxicology has been the bellwether for advances in regulatory toxicology and risk assessment, or why childhood (defined broadly) has been identified as the life stage that is clearly different from the rest and in need of special attention in the risk assessment process.

However, embryos and children are not the only life stages that may differ in their response from 20-45-year-old adults. For example, women in and after menopause are at much greater risk of a variety of diseases associated with aging, many of which are strongly linked to decreased levels of estrogen (including heart disease, osteoporosis, and even some dementia). The degree to which the physiological changes that occur with aging change one's sensitivity to toxicants is not known, at least not on a general level, but there are specific examples in which aged subpopulations appear to be at greatest risk (e.g., itai-itai disease was principally, if not exclusively, observed in post-menopausal women).

The RfC/RfD Technical Panel has made the appropriate choice to expand its focus such that risk assessment practices will now explicitly consider life stage-specific risk, which, of course, includes children. This same conclusion has been arrived at independently by an ILSI-Risk Science Institute-sponsored workshop that was convened in July 2001 to develop a framework for children's risk assessment. Many of the recommendations made by the Technical Panel were also made by the workshop. (NB: The workshop report is available from Dr. S. Olin or Dr. P. Fenner-Crisp of ILSI-RSI, and should be appearing in print sometime in 2002.) ILSI-RSI's work is now engaged in work on case studies that will be used to flesh out the concepts that were introduced in the workshop. The RfD/RfC Panel review and ILSI-RSI's continuing efforts will provide mutual support and synergism as children's and life stage-specific risk assessment procedures are developed in detail.

The Review is also praiseworthy in its recommendation of less-than-lifetime RfVs. While the current RfCs and RfDs are certainly protective, there are numerous situations in which short-term exposures occur that may be higher than the RfD/RfC. RfVs based solely on lifetime exposures provide little guidance to risk managers, emergency management officials, health care practitioners, or the lay public on the degree of concern that should be placed on incursions above the chronic RfD. A variety of RfVs that cover more real-world alternative exposure scenarios will be of considerable help to all of these groups.

The recommendations made in the Review are consistent with EPA's goal of harmonizing cancer and non-cancer risk assessment. The divergence of risk assessment methods has made risk assessment a cumbersome process, the rationale for which is far from transparent. The recommendations made in the Review advance the goal of harmonization and clearly identify steps in the process that can be improved by acknowledging that different manifestations of toxicity may emanate from the same mechanism of action. For the purpose of children's (or life stage-specific) risk assessment, these recommendations will facilitate the use of data from all studies (not just those using immature animals) to interpret the nature of adverse outcomes and predict human risk.

The Review's recommendations incorporate advances in science that will improve the risk assessment process. Risk assessment and testing practices tend to be out of step with the pace of science. There are good reasons for being conservative in making changes, but often this conservatism crosses the boundary to bureaucratic inertia. Many of the recommendations made by the RfD/RfC Panel serve to update the scientific foundations of the risk assessment process, and should also make it easier to incorporate advances in scientific understanding into the process without the need of formally modifying standard operating procedures. This includes a consideration of what we have learned retrospectively about the validity of our assumptions of what the magnitude of each uncertainty factor should be; explicit incorporation of mode of action data into the process; inclusion of exposure/effect-response arrays in the narrative description of the RfV; and recommendations on the use of PBPK and pharmacodynamic data to better dimension the extent of difference between test animals and humans.

While my overall impression of the Review is very positive, there are a number of statements and recommendations that I don't agree with. I will deal with most of these in my Specific Comments and Responses to Charge Questions (vide infra). A few that bear highlighting are 1) the recommendation to discontinue the RfD for developmental toxicity; 2) lack of a recommendation for how to do route-to-route extrapolations, to obviate the need for additional animal toxicity studies simply for the purpose of deriving NOAELs for three different exposure routes: 3) a concern that the recommendations for new testing protocols are not optimal; and 4) the short shrift that is given to benchmark dose as a point of departure for risk assessment.

Response to Charge Questions:

1. Less-than-Lifetime Reference Values?
I agree with the recommendation to derive RfVs for less-than-lifetime exposure situations. There are two important justifications for having RfVs for shorter-term exposures. In the context of children's (life stage-specific) risk assessment, less-than-lifetime RfVs are necessary for expressing the level of risk for specific life stages (none of which, by definition, last a lifetime). Secondly, these less-than-lifetime RfVs will be useful to risk managers, emergency response officials, health care practitioners and the lay public in making decisions about the degree of concern from short-term exposures. Situations occur all the time in which people are exposed to levels of a chemical that are above the lifetime RfD/RfC. EPA has the opportunity to provide much more guidance, in those situations where the science is robust enough to support it, that will be helpful in these situations.

2. Revised Definition for Reference Values?
I agree with the revised definitions, with one exception. It's clear that the RfD/RfC Technical Panel has a great deal of experience within the Agency and the individuals on the Panel have encountered every conceivable misinterpretation of the current definition. The revised definitions appear to address the aspects of the definition most prone to misinterpretation.

The definition that I'm not entirely satisfied with is the one for short-term RfV. The RfVs that are based on developmental toxicity are likely to be subsumed under this RfV, since exposure to a developmental toxicant during a relatively short period of pregnancy has the potential to produce adverse effects. However, only a subset of these RfVs will be based on developmental toxicity; not every compound is a developmental toxicant. The basis for setting the short-term RfV might be included as part of the narrative statement supporting the RfV, but I fear that this might be missed by some readers of the IRIS website, particularly pregnant women who believe they have been exposed to a chemical and are concerned about the health of their fetuses.. It would be of benefit to such readers to provide more information about the possible nature of the risk. One way to do this is to indicate in a single sentence the toxicological basis for the RfV (e.g., renal toxicity in rats, reproductive toxicity in male rabbits, etc.)

I agree with the recommendation to standardize terminology.

3. Endpoint-Specific Reference Values Should Not Be Derived?
I disagree with this recommendation. For the reason noted in the paragraph above, reference values for developmental toxicity may have some utility. It may not be necessary to derive endpoint-specific RfVs as long as the generic RfVs include information about the nature of the effects on which the RfV is based. However, if this cannot be done, there is good reason to continue the RfD for developmental toxicity. This endpoint is of particular concern because of the possibility for pregnant women or their health provider to misinterpret more generic information, leading them to make poor decisions about the need for extensive prenatal diagnosis (with its own attendant risks) or even pregnancy termination.

4. Targeted testing strategies to support less-than-lifetime RfVs?
The Panel has identified a number of gaps and possible deficiencies in current testing guidelines, and recommends the development of a number of new study protocols and/or the modification of existing protocols. I agree that some protocols should be modified, particularly those that are used in determining risks to children or risks from exposure during early life. However, I disagree with the specifics that have been proposed. The idea of developing more extensive acute tests is of concern, as is the implication that more comprehensive data needs to be developed for a variety of routes of exposure.

Regarding acute toxicity tests, it needs to be recognized that these tests provide limited information, irrespective of how well they are designed. The currently available up-down procedure uses as few animals as possible and provides adequate data about lethality and severe health effects. It is possible to add measurements to this test to provide more data on less pronounced effects. This, along with other information from subchronic tests, including target organs, nature of effect, and potential for reversibility can be analyzed in combination to understand acute hazard potential. Pharmacokinetic considerations can be used to adjust the subchronic point of departure, or at least to determine whether the subchronic POD is a credible surrogate for the acute POD.

Regarding routes of exposure, it should be made clear that there are acceptable empirical and computational methods to extrapolate from route-to-route which are preferable from an economic and animal welfare perspective to conducting separate animal tests for each relevant route of exposure. This is in keeping with the stated goals of making the testing process as animal-efficient as possible, and in using the best science in risk assessment. Pharmacokinetic models for absorption by different routes are available for many, many compounds and can often be generalized to related materials. These can be supplemented with in vitro data (e.g., dermal penetration through cadaver skin). Often, repeated testing by different routes provides little or no additional information, and can occasionally be counterproductive. For example, it is widely recognized that there are a number of technical challenges in conducting dermal developmental toxicity studies, including that it is more difficult to exaggerate dosages to the point of eliciting a developmental response and significant dermal irritation at the site of administration can cloud the interpretation of pregnancy outcome (Kimmel and Francis, 1990, Fund Appl.

Toxicol. 14: 386-98).

I support the idea of combining existing study designs to be more efficient in animal use and to fill gaps in the kinds of information that are obtained from guideline studies. However, there is a limit to what can be included in a single test method; the "unified screening study" exceeds that limit. Such a study would be extremely labor intensive and could only be carried out in a few labs. Furthermore, a large number of trade-offs would need to be made in terms of which strains of animal to use; optimizing for the reproductive aspects of the study would probably compromise sensitivity for detecting carcinogens for example, as the F344 strain, which is commonly used in chronic bioassays, does not breed as well or produce litters as large as outbred strains.

My recommendation is to retain the recommendation to add endpoints (particularly functional endpoints) to increase sensitivity, and to combine protocols to fill information gaps (especially about long-term consequences of early exposure) and be more animal-efficient. However, the recommendations for specific new assays should be deleted. Instead, the Agency should be encouraged to work on these via a process that is dedicated to test method development.

I concur with the recommendation for further research on life-stage specific pharmacokinetics and modes of action, and on latency and reversibility. One pending publication that the Panel should be aware of is one that was commissioned by the ACC entitled "Evaluation of physiologically-based pharmacokinetic and pharmacodynamic models of pregnancy and lactation for assessing dosimetry in the embryo, fetus and newborn". This work has been submitted for publication, but a version is available from ACC and may be useful as a starting point in developing research directions.

5. Duration-adjustments for Developmental Toxicity Studies?
I agree with the Panel that there is no good reason to have a different default for developmental toxicity than for all other endpoints. I also agree with the Panel that the most appropriate course of action is to make it a high priority to develop models and procedures for determining what dose-duration adjustments need to be made, if any, for inhalation studies. Clearly, peak concentration is more important for some compounds, and simply dividing the NOAEC by three or four to account for the 16-18 hours/day that the animals aren't in the exposure chambers is not scientifically justified. It may be possible to develop some heuristic rules/ defaults based on other information about the chemical, particularly its mode of action, that could be used to determine whether peak or AUC is the more likely determinant of toxicity for a given compound. For example,the toxicity of an agent that acts via receptor interaction is more likely to be related to peak concentration, whereas the toxicity of an agent that acts by inducing repetitive damage (e.g., accumulation of mutations) is more likely to be related to AUC.

6. Weight-of-evidence/Narrative Statement?
I fully agree with this recommendation. Given what we now know about biological plausibility and the variety of manifestations that can be produced by a single mode of action, no other approach than a comprehensive, weight-of-evidence-based analysis of the data should be acceptable. While it is true that this will require additional care and the collective expertise of a number of people, particularly for data-rich chemicals, I am sure that EPA is up to the task.

The inclusion of a narrative statement is important for two reasons. First, experience tells us that it is impossible to convey enough information about risk simply with an estimate of a safe dose. The narrative statement is a succinct way of communicating the nature of the effect, the certainty behind the estimate, the anticipated variability in response, etc. Second, because a weight-of-evidence risk assessment requires the expert judgment of one or more risk assessors, the narrative statement is necessary to maintain transparency of the process.

7. Exposure-response arrays, etc.?
I support the use of exposure-response arrays but believe that the concept should be expanded to include even more data. The example array that was presented (Fig. 4-3, p. 4-49) includes only general information (NOAEL and LOAEL) about which aspects of toxicity have been evaluated. It would be nice to have additional information that conveys dose-response information, nature of effect, whether the effect increases in severity with increasing dose, as well as information that addresses weight of evidence, especially the number of studies and what they showed. Bob Kavlock developed a very nice, two dimensional graphic presentation for developmental toxicity data that conveyed most of that information (Kavlock et al., 1991, Teratology. 1991 Feb;43(2):159-85).

I also agree with the recommendation that POD be selected after an evaluation of all the data rather than relying on a single critical study.

8. Uncertainty/variability factors?
I agree with the recommendation that the total UF be limited to 3000. There comes a point where one just has to conclude that the data are too uncertain to permit the calculation of an RfV.

I agree with the recommendation regarding the need to explicitly consider sensitive subpopulations before moving from the default factor of 10 for human variability. I agree with the recommendation that more data are needed to support the selection of the appropriate default value for his factor.

I agree that the selection of uncertainty factors be based on god scientific judgment, and be explicitly justified in the narrative statement accompanying the RfV.

I agree with the conclusion that UFs already exist that deal with database deficiencies. I agree with the Panel's conclusion that the current UFs are adequate to cover the uncertainties in children's risk assessment. The FQPA safety factor provides no

additional value to a scientifically based assessment. I have been parts of discussions in several forums about this factor. Even its advocates admit that it adds no scientific value, but defend it on the basis that it is a negative incentive for pesticide registrants to run more studies. Therefore, this is clearly a policy issue that has no place in the risk assessment process.

I agree with the recommendation to scrap the modifying factor.

I agree with the recommendation that EPA should develop more guidance for moving from default values for uncertainty factors. I believe that the IPCS documents that are cited are a good starting point for the development of that guidance.

Specific Comments:

Executive Summary: I did not find this to be very informative. The Charge Questions to the external peer review actually provided a better summary: it teed up the issues and presented the recommendations. I would suggest that the Issues and Recommendations from the Charge Questions be cut-and-pasted into the appropriate places of the draft Executive Summary.

Chapter 2: This is a nice summary of existing procedures. I didn't know that much about AEGLs before reading this; I was favorably impressed with the three levels of AEGLs that indicate the levels expected to elicit non-adverse effects, reversible adverse effects and irreversible adverse effects. This kind of guidance might also be helpful if provided as part of the less-than-lifetime RfVs.

Chapter 3: Page 3-31: The alternative acute toxicity protocol that is described adds little value beyond what a subchronic study could provide. It seems to me that with a little thought it would be possible to modify the existing up-down method to get additional information about organ toxicity, and then rely on subchronic study data and PK considerations to arrive at acute RfVs.

Pp-3-34-40: These chronic/unified assays are comprehensive but are too ambitious and require that compromises be made such that the study design will not be optimal for all types of toxicity to be assessed. One of the big issues is choice of strain and species, which the Panel does address on p.3-39. Perhaps a more serious concern is that studies this large will be logistically complicated and difficult.

I recommend that a separate group look at test design alternatives, with that as their only objective. One possible way to get the information needed is to think a little more modestly. In the context of children's health, AIHC convened a group of experts about two years ago to see if a special protocol could be developed. That group concluded that a study that combined a one-generation and subchronic design (in which the F1 generation is evaluated for subchronic endpoints), along with neurotoxicity and immunotoxicity endpoints in the F1, would be a potentially powerful evaluation of possible effects on children's health. It would obviate the need for several studies. One

could also envision that a subgroup of animals from this study could be maintained on a chronic basis to look for latent effects. This would have most of the advantages of the unified protocols in the RfD/RfC reviw, without as many downsides.

Chapter 4:
P. 4-11, fourth paragraph: I disagree that U-shaped or inverted U-shaped dose-response curves " are not uncommon" in toxicology. In fact, they are uncommon. In most circumstances, they represent less than comprehensive analysis of the study data. The classic example is that of rate of developmental toxicity after teratogen treatment. The malformation rate does, in fact, have an inverted U-shape, because the affected fetuses die as dose increases. Therefore, when one does the appropriate thing and plots the data for malformations plus in utero death, the curve is the typical sigmoid shape. The Review should be changed to indicate that dose-response relationships are straightforward, and that one should look for explanations for unusual shapes. If the explanation is not found, then skepticism should be the appropriate prescription unless the weird shape can be recapitulated in a separate study (preferably by different investigators).

P. 4-14, second paragraph: This statement is true, but exposure assessment is not part of the RfV determination.

P. 4-24, lines 8-12: The C x t procedure does indeed provide an automatic margin of conservatisim, as is noted in this passage. However, if it is not described in the narrative summary associated with the RfV then it takes away a measure of transparency in the process.

P. 4-33-34 (section D.3.d): The use of allometric scaling for cross-species PK extrapolation is worth considering. However, before the recommendation is accepted it will be important to do a retrospective study comparing the responses of two species with rich databases (e.g., rat and mouse) to determine for several endpoints whether the $3/4^{th}$ power scaling accurately reflects the species differences in response.

P. 4-35, second paragraph: This paragraph represents a giant step backwards in risk assessment practice. The paragraph above this one explains that the benchmark response level is selected to approximate the limit of detection for the study design in question; i.e., it is equivalent to the highest possible NOAEL or lowest possible LOAEL for the study. The BMD is then further adjusted so that a lower confidence limit is used as a POD, which will be substantially lower than the highest possible NOAEL. Yet, in the second paragraph on this page, the BMD is criticized for not being protective of the human population. This is risible. It needs to be corrected.

The third paragraph on this page (continuing to p. 4-36) acknowledges that NOAEL is a moving target that depends on study design, study conduct (degree of variability), sample size, etc. Yet there is only a tepid recommendation that these factors be considered in

study acceptance. The Panel needs to rethink the value of BMD and the limitations of NOAEL as a POD.

P. 4-37, third paragraph, item 4, and p. 4-42, section D.5.d: There should no longer be a need for an uncertainty factor to adjust for a study without a NOAEL. I think that there is universal agreement that BMD be used for this purpose. If for some reason the BMD cannot be calculated, then the default value should be 3 bsaed on the data presented in section D.5.d.

P. 4-43, last paragraph: The Panel should recommend that the subchronic-to-chronic UF default value be reconsidered on the basis of the retrospective comparisons that have been done (and which are cited in this paragraph).

Review by
Rodney Dietert
Cornell University

Revised Comments (post meeting)—Rodney R. Dietert, Cornell University
June 22, 2002

EPA Draft Document: **Review of Reference Dose and Reference Concentration Processes**

Peer Review Workshop---June 2002

General Impressions

The recommendations, in general, appear to improve both the presentation of data and the derivation of RfDs through a consideration of the changes needed to enhance the protection of children's health. The proposals, in this draft report, recognize the diversity of new data which will collected on life-stages, the disparities in not only does-response but also most sensitive endpoints that may exist across life stages, and the need to represent and utilize the diverse data. It should be recognized that while the application of multiple (more than 4) factors-of-10 may be overly protective for many chemicals, the scarcity of developmental toxicity data for many physiological systems (e.g. immune and neurological) is a true deficit that needs to be addressed in terms of effective health protection. This draft document appears to achieve a good balance among difficult choices in this regard.

Response to Charge Questions

1. **The recommendation for deriving less-than lifetime reference values**.

I would strongly support the need for deriving less-than lifetime reference values as an addition to chronic RfDc and RfCs. The justification can be found in the existing data (although more limited than desirable) regarding exposure of non-adults including embryonic and fetal exposures. It is clear that even brief exposures occurring during critical windows of development may pose a disproportionate health risk compared with chronic exposures of adults to similar concentrations. The current report does cite this justification in addition to commenting on the general lack of acute juvenile exposure information. Any additional information that could be added to the IRIS database that would capture these problematic life-stage-based acute exposures would greatly aid the protection of the most vulnerable populations.

2. **The revised definitions for reference values**.

The revised definitions for reference values are necessary to present standardize data across route and duration of exposure and to accommodate the need to consider risk relative to mode of action across both cancer- and non-cancer-related endpoints. By citing reference value relative to route and duration, the context of each value will be more fully appreciated and the eventual application of Ufs more understood by a broader segment of

the scientific and lay public.

3. **The recommendation that end-point specific reference values should not be derived**.

The recommendation is effectively is married to recommendation Number 2. If duration-based reference vales are derived, then there is little need to derive endpoint-specific values as well. If fact derivation of developmental toxicity-specific reference values would be linked inherently to less than lifetime exposures. Therefore, it would be important to decide whether life stage of exposure or duration (and route) of exposure should be the basis for establishing a broader array of reference values. Under the latter, life stage-based differences in sensitivities are accommodated in considering the most sensitive populations. However, a RfD (DT) approach could not accommodate all considerations of sensitive sub-populations. Therefore, it is logical to base RfD on duration and route and use life-stage-specific data and/or appropriate UFs to ensure the most sensitive populations are adequately protected. However, it should be recognized that the database for sensitivities of different life-stages is grossly inadequate and needs to be addressed through data collection whenever possible.

4. **The recommendation to alter testing approaches given the potential importance of life-stage variables and the need for less than lifetime reference values.**

I strongly support the need for a life-stage driven approach to future testing. It is clear that new testing strategies should be adopted which encourage additional data to be obtained covering less-than-lifetime exposures over potential critical windows of development. Some of the models presented in the draft reports are quite helpful to that end. While it is true that every chemical need not necessarily be re-tested to this end, we do need to ensure that potential life-stage differences in sensitivity are addressed and hopefully with highly relevant data in place much of the time. While it is extremely helpful to acquire quality data on mode of action and such data do enhance predictability of exposure outcome, I do not believe that clear mode of action data alone can predict most life-stage–specific effects. Developmental PK data cannot address different sensitivities based on life-stage timed organogenesis process (T lymphocyte "education" in the thymus). In many cases, life-stage related differences in target organ susceptibility are not readily predicted even when a chemical is well known. Developmental stage changes in TCDD-gene interactions are one example that highlights this point. Therefore, the recommendation to developing and new testing approach is key.

The proposal to develop additional guidelines is very helpful for enhancing protection of children's health. I strongly support this. As indicated, the issue of when to trigger such testing is potential as important as the guideline development. Having new guidelines that can adequately assess neurotoxicity and developmental immunotoxicity is helpful, but if these are never utilized or are utilized unnecessarily, then the anticipated benefits will not achieved. Likewise latency and timing of endpoint assessment are important factors. The prior committee should be applauded for identifying the concern over of duration of

assessment following early exposures. Long term health outcomes resulting for early acute exposures have rarely been tracked into geriatric-equivalent populations.

5. The recommendation to include duration-adjustment for inhalation developmental toxicity as for other health endpoints.

I endorse this recommendation that duration adjustment processes from intermittent to continuous exposures for inhalation developmental toxicity studies should be done the same manner as for other health endpoints.

6. The recommendation for using a weight-of evidence approach for hazard characterization and for substituting a narrative to describe the database rather than a confidence ranking.

The narrative approach is a particularly positive revision for the database. It will enable scientists to examine not only the nature of data available and the limitation to those data but also facilitate the filling of potential gaps in data sets particularly as pertains to issue of early exposure and children's health. It will also provide descriptive analysis of the application of UFs and situations where UFs may be considered as overlapping or duplicative such that PODs are considered.

7. The recommendation to use an exposure response array and derivation of sample reference values to select the point of departure for the final reference value.

In general I support this recommendation. It is consistent with the proposal to derive multiple RfDs. In this case, single most sensitive endpoints would still be considered but would not eliminate consideration of other dose response-derived effects. Dose response curves would take on a greater significance. One reason this is needed is that a critical effect at one life stage is unlikely to be the critical effect covering all life stages of exposure. Hence by focusing on an adult-exposure associated critical effect, additional investigations might actually miss a sensitive effect on another system or more sensitive endpoint within the same physiological system that is associated with earlier life stage exposures. Additionally, over time our knowledge of the significance of particular biological endpoint improves. So called "critical effects" can take on greater or lesser significance as our biological knowledge base expands. As a result, it is helpful to have a more comprehensive approach to exposure/response included in the database as well as in the derivation of the final RfD.

8. **The recommendations concerning the application of uncertainty/variability factors**.

A. I agree that if extrapolations required in more than four areas that would mean that uncertainly within the database would not permit a reference value to be generated that would have any real utility. The suggestion that UFs for a particular chemical be capped at a maximum of 3,000 seems reasonable. However, I anticipate this point should garner significant discussion within the panel. The capping of uncertainty factors at 3,000 should not be a very palatable substitute for quality intra-species and developmental data. While I agree that using multiples reaching 10,000 may be ludicrous and a huge and unnecessary burden on our society, I also am concerned that a cap of 3,000 in some case may not protect across all life-stages combined with major ethnic sub-populations. There while I support this proposal, the opportunity for collection of appropriate developmental data is clearly the desirable alternative for protection children's health with confidence.

B. I concur with the proposal to consider reduction of the intra-species UF under circumstances where the data set is deemed to have included the most susceptible population within the species. In practice this may be more difficult than anticipated as the only way to ensure that the more susceptible population is included is going to be to have exposures relevant to several different life-stages and to have data from several different genotypes within the species. I suspect the percentage of chemical-specific data sets in which both of these extra data are collected is quite small. Nevertheless the opportunity for this UF reduction to occur may encourage the collection of much needed data sets covering multiple life-stages and multiple genotypes.

C. The recommendation to consider alternatives to the 10 factor depending upon the quality of the date is logical. There is nothing inherently magical about the magnitude factor. However, it should be recognized that one or two example pertaining to the quantitative relationship among sensitive sub-populations vs. the general population may not provide much assurance that the same quantitative relationship hold for other chemicals. Departure from the magnitude factor should be based on the existence of extensive and relevant examples and that is likely to require more life-stage-specific head to head comparisons than exist at present.

D. While the argument that these UFs tend to overlap has definite merit, the idea that children are protected by the application of existing UFs seems very uncertain. My concern is that to routinely apply an extra factor of 10 for children on top of all existing ones may not be prudent. However, given the lack of direct data for children vs. adult sensitivities for most chemicals, it is naïve to assume that protection has been optimized. If magnitude overlaps are adopted routinely, then it seems incumbent to facilitate much great data collection relevant to the children's health issue.

E. I concur completely with the recommendation of the prior panels and have no problem with the elimination of future MF use.

F. I agree that it is desirable to replace UFs with data-derived factors. However, there are not sufficient data available at this time for that to realistically be put into place. Therefore, the best the EPSA can do is to begin to accumulate data that will eventually allow appropriate modifications to default UFs.

Specific Observations

The document makes a meritorious attempt to improve the procedure for data descriptions, the applications of UF and PODs, and the standardized derivation of RfDs. The move to use a detailed narrative in which the extent, nature and deficiencies of the database are clearly indicated as well as the logic for UF application will be very helpful. It is clear that more direct data among sensitive (e.g. particular life-stages) *vs*. the general populations is critically needed before the utility of combined application of Ufs and/or departures from certain UFs is known. The proposal to facility collection of early life-stage exposure

Review by
John Doull
University of Kansas School of Medicine

Comments on the EPA document "A Review of the Reference Dose and Reference Concentration Process". John Doull, Univ. of Kansas Med. Center, Kansas City KS, 66160

As stated in the preface and in several of the chapters of this document, harmonization is one of the primary goals of the risk assessment forum and is a stated goal of the agency. This document makes several recommendations that would promote harmonization (use adverse rather than deleterious and in place of cancer vs non-cancer, use RfV rather than RfD and RfC, harmonize HEC with HED, RfD with RfC and BMC with BMD) although details of how this might be accomplished are sketchy. Harmonizing inhalation with oral exposure data requires that both the inhalation (concentration times duration) and the ingested dose have the same fundamental biological units (molecules or moles and time). Following single exposures the kinetic and/or the dynamic half life (whichever is rate-limiting) can be used to compare exposures and effects. With repeated exposures, the comparison in more complex since it involves a frequency factor in addition to the kinetic and dynamic time factors but the use of continuous exposure will simplify the comparisons (K. Rozman, J. Doull and W. J. Hayes, Chapter 1 in Handbook of Pesticide Toxicology, Ed. R. Krieger, Academic Press, 2001, Rozman and Doull, Dose and Time as Variables of Toxicity, Toxicology, 144, 169-178 2001, W. A. Waddell, Thresholds of Carcinogenicity in Flavors, Tox. Sci. in press 2002).

It is clear from reading this document that the risk assessment protocols of different workgroups and agency offices lack consistency (10X task force, ARE methodology from NCEA, IRIS, 1996 Cancer guidelines, OPPS cumulative risk and mixture guidelines, NAAQS etc.) and the recommendations contained in this document could provide a reasonable place to begin harmonizing not only EPA but other agencies (ATSDR, NIOSH, OECD, IARC). An ultimate goal might be protocols that are so consistent, clear and unambiguous that we no longer need a glossary to read and understand the agency documents.

The EHC of the SAB met on Jume 9-10, 1998 to critique an SAB report on: Review of RFC Methods, Case Studies. This was a very"hands-on" session using four groups of agents: Particle case studies (MDI, antimony trioxide and phosphoric acid, Category 1 gas studies (methylmethacrylate, acetaldehyde, chloride dioxide, and 1,3-dichloropropene), Category 3 gas studies (CS_2, vinyl chloride, n-hexane and 2-ethoxyethanol) and non-verifiable case studies (BCME, methoxychlor, caprolactam and acrylamide). Since our findings and recommendations from this exercise involved dose responses of actual agents, they might be helpful in the current review of the RfD and RfC processes.

1. Comments on Charge #1 (derive less than life-time reference values): Such values are needed in order to protect against short term exposure effects since these may differ from long term effects.

2. Comments on Charge #2 (new reference value definitions): It is not clear to me why we should replace the three traditional exposure time values (acute, sub-chronic and chronic) with four values (acute, short term, longer term and chronic). The current values

are already arbitrary and the new values will be also. Why not stop using arbitrary values and use the actual exposure times? To do this we will need data on the half-life of the rate limiting process (kinetic or dynamic). Short term values for the regulation of exposure for effects which are limited by long half lives (dioxin, asbestos = kinetic, methanol = dynamic) would be of less value than regulation based on chronic exposure. Conversely regulation of agents where the limiting effect has a short half life (solvents = kinetic, nerve gases = dynamic) should be based on these short term effects.

3. Comments on Charge #3 (derivation of end-point values): contrary to the goal of harmonization of approach.

4. Comments on Charge #4 (alternative approach for life stage): Conservation of resources (animals, time funding) should be considered before adding any new requirements to the testing protocols. The current factors for intra-species variation should be adjusted to include sensitive populations (harmonization) rather than arbitrarily adding new factors. Results of new studies (immuno, neuro, neonate tox, etc) should be incorporated into existing factors. Using latency as a factor is appropriate since both exposure and effects have dose and time as the key relevant factors.

5. Comments on Charge #5 (duration adjustment for inhalation): The AEGL approach has been validated by NAS/NRC/COT and should be used by other risk assessors. Bliss suggests exponents on both c and t rather than only on c as recommended by ten Berge and Druckrey's studies with nitrosamines clearly demonstrated the advantage of the exponent on t.

6. Comments on Charge #6 (weight of evidence and narrative description): Use of the weight of evidence rather than any single effect is more likely to result in a good regulatory result and the use of a narrative description is basically just a logical extension of this approach. However risk managers prefer bright lines and seem to be less receptive to ranges and narrative descriptions of adverse effects. Solution? Educate risk managers or present both types of information.

7. Comments on Charge #7 (ATSDR-like array of data for POD): I like the approach used by ATSDR in their tox profiles to summarize adverse effects and think it would enhance the IRIS documents but am not sure how we could use it to generate a POD that would be different from the current benchmark or NOAEL method. The idea of using a weight of evidence approach for the POD also seems interesting but am not sure how this might improve what we do now.

8. Comment on Charge #8 (uncertainty and variability): It seems to me that the agency is using the recommendations from Science and Judgement in this area and I am not aware of any new/novel methodology that would markedly alter or improve the approach. Recommendation A (limit total UF to 3000): Good idea since unrealistically high UF values damage the credibility of the process. Recommendation B (reduce default from 10 to 3): this decision should be based on the data but think most of that supports the reduction. Recommendation C

(Document science basis of factors): The new data quality bill that Congress has passed will probable require this and agency policy clearly supports this. Recommendation D (use existing uncertainty factors to cover sensitive groups): This is consistent with harmonization goals and makes scientific sense to me. Recommendation E (eliminate MF): Agree that it should be subsumed in the general database uncertainty factor. Recommendation F (use dynamic/kinetic approach which is specific for each chemical): Would greatly improve process.

Additional unsolicited recommendation: The agency should encourage or even require submitters to provide kinetic information on every chemical and should use both the kinetic and dynamic (MOA) information in regulating exposures to the chemical.

Review by
Michael Dourson
Toxicology Excellence for Risk Assessment

EPA/630/P-02/002A

May 2002

External Review Draft

A REVIEW OF THE REFERENCE DOSE AND REFERENCE CONCENTRATION PROCESS

Michael L. Dourson, Ph.D., DABT

Responses to the Charges

#1. I agree with the recommendation. But the definitions are not flexible. For example, a 2-day exposure is more like acute exposure than it is short-term exposure. Furthermore, defining a 2-day exposure as short term, or a 91-day study as chronic will generally not be followed because EPA and other risk assessment scientists routinely make judgments in the use of data for different RfVs, based on the underlying biology. Early versions of Casarett and Doull can be used for standard definitions, such as a 90-day as subchronic or a 14-day study as short term. EPA can then characterize a 30-day exposure as short-term to subchronic, or a 1-year dog study as subchronic to lifetime (or chronic), etc. This is EPA's intent anyway, and making itself clear will avoid unnecessary arguments.

#2 A. I agree with the intent of these new definitions. For example, I wholeheartedly agree with removing the word "noncancer." I like the use of the word "adverse." Unlike most folks at the meeting, I like the language on centering of the estimate. It might have been clearer if you had spoken about precision of the estimate, however. As EPA knows, RfVs are imprecise, some more so than others based on the number of uncertainty factors used in their development. For each definition, EPA may also wish to add the word "generally" after the word "estimate" in the phrase on "centering." EPA may also wish to consider adding definitions and description of "provisional" reference values, for situations where the data are insufficient for an IRIS-quality reference value, but where a regulatory body needs some estimate of toxicity for cleanup or other regulation.

#2 B. Although I understand the value of changing terms to introduce new subjects, my preference is for EPA to stick with the Reference Dose (RfD) in mg/kg-day and Reference Concentration (RfC) in mg/m^3. Both terms are well understood to represent the oral and inhalation routes, respectively, and these terms are used throughout the world. This use has promoted the distinction between risk assessment and risk management, entirely evident in this country, but less so elsewhere. What EPA really needs is a new term for the dermal reference dose, since issues associated with this concept are not yet worked out. For example, what are its units: mg/m^2 of surface area? In fact, EPA may wish to use the term RfV for the dermal route.

Furthermore, I like the use of word descriptors instead of subscripts. Thus, I favor saying an acute RfD, or longer-term RfC, rather than RfV_{AO}, or RfV_{LI}, respectively. It is easier to type and does not break up the flow of a reader's understanding of the text. Remember that risk managers are going to have to read and understand these concepts. Using English makes this easier.

#3. I agree with the recommendation, but how does this jive with EPA's recently published mixtures guidelines that recommend the development of target organ toxicity values?

#4 A. I agree that EPA should revisit current testing protocols to see if they are now answering questions that we pose. However, simple fixes to existing studies might be very helpful. For example, asking folks to monitor standard toxicity tests on the F1 and F2 generations in the current reproductive toxicity study would be most helpful in gauging the systemic toxicity of a chemical to the young animal, and would not overtly impact the conduct of this study. We did this for perchlorate and it was most useful for EPA in its deliberations of critical effect.

#4 B. I have the same comment as above for #4 A.

#4 C. I agree that research to evaluate latency to effect and on mechanisms/models of action and pharmacokinetics at different life stages should be encouraged. I am not sure why we need to further study the reversibility of an effect, however, unless we are planning to consider adverse, but reversible, effects as not critical. If EPA was proposing this, many folks would disagree.
Alternatively, if EPA and others are thinking to use an uncertainty factor for severity of effect, as suggested by one of the peer reviewers, then understanding reversibility will be important.

#5. I agree, and have felt this way since 1985.

#6. I agree with the first part of the recommendation. Although EPA's current system uses a narrative with low and high confidence defined with medium in between, the existing statements on IRIS could be vastly improved. However, further consideration should be given to the definition of the minimal database. The proposed definition excludes some situations when route-to-route extrapolation may be appropriate (for example, when there is adequate toxicokinetic data for conducting a route-to-route extrapolation, and there is information from other durations that shows that the portal of entry is not a target). Moreover, the proposed definition of a minimal database allows derivation of a short-term value from a single developmental toxicity study that reported effects only at high doses. However, EPA's own experience has shown this to not be health protective in general, if systemic toxicity has not been adequately evaluated.

The term "dose response information applicable to the duration in question" needs to be

clarified. On first reading, it appeared that subchronic to chronic extrapolation would no longer be allowed, but the discussion of UFs indicate that a subchronic study would still be used for deriving a chronic RfD. Furthermore, while the use of a more narrative description of the database quality is reasonable, the current definition of the "high confidence" database aids in the determination of the database UF, while the proposed definition of the robust database makes it much harder for assessors to apply consistent principles in the choice of the database UF.

Lastly, the text is unclear as to whether the database is characterized for each route/duration, or as a whole. (Box on page 4-17 implies specific to route/duration, but text on p. 4-17 says extensive data on one route is not sufficient for the database to be considered robust. Route-specific characterizations seem more useful, since the absence of data for one route may not affect the confidence in the reference value derived for another route. What information is the characterization intending to convey?

#7. The graphic approach is nice, but the accompanying text is very important. Not all effects will be of equal quality or import, and graphing data without explication may mask this fact.

However, I am having trouble with this logic on in this section which seems to imply that concept of critical effect is no longer needed. Are EPA risk assessment practitioners actually recommending this? Page 4-21 lists multiple reasons for the lack of need for the critical effect concept.

> For example, in Line 2: Presentation of a single endpoint… This is true, but this is obvious. This is well known to occur – so what?
>
> Or in Line 3: .Nor does the presentation… - This is true, but so what? Do you really want to know the dose response curve for mortality when developing an RfD?
>
> Or in Lines 10-12: …Focusing on a single...levels of exposure This is wrong! EPA focuses on all effects in the area of critical effect. Look to IRIS for many chemicals where co-critical effects are stated. This is also the practice of other world health organizations.
>
> Or in Line 13: Most importantly…or routes of exposure. Again, this is expected. So what? It's not a reason to abandon the concept of critical effect.
>
> The only reason that has merit in a revision of the concept of critical effect is found in Line 13…exceeds the RfD or the RfC. Unfortunately, the RfV approach cannot address it. You need categorical regression or BMD here (or something else).

Paragraph beginning with Layered upon this complex… This is already addressed by current concept of critical effect. Your citation of the EPA methods for RfCs is one example, but RfDs also use these adjustments when data are available. I am happy to talk with EPA staff about this issue in further detail if needed.

#8. A. Line 2: after "extrapolation" add "in a group of 5 areas" (interspecies,…). I agree with the recommendation, but only if BW ratio to the 3/4's power is used for oral extrapolation with an interspecies UF for toxicodynamics of 3. I can provide specific text to EPA if needed.

#8. B. I agree with the recommendation, but only with the stipulation that folks understand the appropriate interpretive use of the intraspecies UF_H. See for example, Dourson et al. (2002) for an explanation of this. Again, I can provide specific text to EPA if needed.

#8. C. I agree.

#8. D. I agree with the recommendation, but am perplexed as to the phrase "in most cases." Risk assessors are not going to propose an RfV that is not protective of sensitive populations, including children. Residue concerns about toxicity will be addressed in the development of an RfD. The FQPA was most helpful in re-emphasizing EPA's obligation in its use of UF_D, and of course exposure considerations might still be meaningful, but to state that the RfD does not protect a sensitive group because of residual concerns about toxicity, is to state that the RfD is not protective of sensitive individuals. In such cases, the RfD needs to be revisited.

Page 13 of the charge
Line 1: delete "all or" and add "the exposure"

#8. E. I agree with the recommendation, but not for the reasons stated in the text. Please see my specific comments found below.

#8. F. I agree with the recommendation, but do not understand the caution. Relatively few data exist because risk assessors have never been clear how such data would be used. More data will be forthcoming.

General comments

No need exists to use a UF_S for a critical effect derived from a developmental toxicity study, if chronic bioassays are available in the database. This is base on well-understood biological principals and long standing practice within EPA and elsewhere. I am happy to discuss this in detail with EPA folks if needed.

On the other hand, the use of a developmental toxicity study as a basis of an RfV will be problematic without at least some other standard toxicity bioassays. This is based on analysis by EPA, which shows that the frequency of developmental toxicity as the critical effect is small compared with standard bioassays. Carole Kimmel of EPA has pointed out good reasons why this prior EPA analysis is not definitive. (In brief, the comparisons of systemic toxicity were made with older studies of developmental toxicity; newer studies may not show the same disparity---good point!). However, the analysis still stands until another one is complete; besides, systemic toxicity studies were also older in this evaluation.

EPA needs to read and possibly cite the following literature in the appropriate text location (shown as page numbers behind the citations). Copies of these citations have been provided.

Baird et a. 1996 (p. 4-46)
Barnes and Dourson, 1988 in the introduction and elsewhere as appropriate.
Dourson et al. 2002 (p. 4-37)
Dourson et al., 1985 (p. 4-35)
Dourson, 1994 in the introduction and elsewhere as appropriate.
Dourson & Stara, 1983 (p.4-42)
Dourson, 1980 (p. 4-25)
Druckrey, 1967 (p. 4-25)
EPA, 1980 (p. 4-25)
Felter & Dourson, 1998 (p. 4.4)
IRIS Background Document 1, 2002 in the introduction and elsewhere as appropriate.
Scheuplein et al. 2002 (p. 4-41)
Swartout et al. 2001 (p. 4-46)

Specific Comments

Page 1-1

INTRODUCTION: I find it odd that EPA's existing RfD methods texts are not cited in the introduction, specifically the IRIS Background 1, Barnes & Dourson, 1988; or Dourson, 1994. These references are provided as attachments.

Page 2-2
3rd Paragraph, equation: for carcinogens, this equation is $DR \times T^n = k$. Why the reversal here?

Page 2-3
Although at the time of the SAB review, EPA's CatReg required parallelism across severity categories, this is no longer the case.

Page 2-5
Please provide the rationale or supporting data and reference a to why these three study types (acute neurotoxicity, developmental toxicity, developmental neurotoxicity) are the most useful for the acute RD. It would seem that a standard bioassay that looks at all endpoints would at be at least as helpful, if not more so.

Page 2-8
Last sentence: EPA's Office of Water refers to this factor as a safety factor, and not a "policy" factor.

Page 2-12
C. Recommendation – I agree with this recommendation.

Page 3-1
Line 10: Armamentarium should be armament.

Page 3-2
Line 5: "It should be noted that testing guidelines…". I do not understand how this statement could be made since standard tests included in the database for high confidence RfDs and RfCs include different life stages.

Page 3-4
Under "Human" column for Puberty 12-21 yrs: : Puberty really lasts until 21?

Page 3-5
Line 20: Delete "some." After "carried out" please add "but not necessarily in detail for all evaluations."

Line 25: Delete "any significant."

Line 26: add "complete" before "detail."

Page 3-6
Table 3-2: Very nice table.

Page 3-7
Line 1: Delete "somewhat." Also, this sentence has 2 verbs: "evaluate" and "include."

Line 2: What is the purpose of the phrase "include cage side observations…designs."

Line 6: What do you mean by "screening level"? Do you need histopathology at the e-m level when parafin shows no effects for "detailed?" Define "screening" – many endpoints that might be thought to be sensitive, seem to be described as screening.

Line 7: Delete "and."
Line 8: Add after analysis "of function" these analyses are "often" limited in regard…
Add period at end of sentence.

Line 12: The sentence "The guideline studies should be changed to read: The guideline studies shown are those from which generally occurs the derivation of a chronic oral RfD.

Line 13: Change similar studies are "required" to "often used."

Line 15: After "chronic studies shown" add "from which the RfD or RfC is estimated with the possible use of an additional uncertainty factor."

Line 25: Add "in" young adult animals… I don't understand the meaning of the sentence "Other than mortality, the to select exposure ranges for longer-term studies. Nor the next sentence "Acute guideline studies…"
Footnote: last line: "testing of pesticidal… - should this be "pesticides?"

Page 3-8

Line 11: Add "and' after exposure; add "called" after "is"

Line 13: Change to: …chemicals. "This study" includes…

Line 21: After "No direct information is " add "generally"

Line 22: After "young animals" add ", mid-life adults"

Page 3-9

Line 9: After 12-month, add "or longer"

Line 16: After No information is, add "generally." After chronic studies, delete "in" and add "started with pre- or "early" postnatal….

Line 17: End sentence at "development. Delete "although" and add "However, the…

Line 20: Add "Generally" before sentence beginning "No subchronic….

Line 21: Sentence beginning "No chronic…"– you already said this in sentence 2 of same paragraph.

Line 27: Add "generally" before "completely lacking…"

Line 30: Add "generally" before "lacking." Also add "of 1 year" after "…study in dogs.

Page 3-12
Line 1: Add "generally" before "no follow-up…."

Page 3-16
Line 15: Delete the word "screening". I have a problem with the continued use of the adjective "screening" with many, if not all, of these bioassays. Here is yet another example. Do you really mean to say that the chronic/carcinogenicity or two-generation studies are only "screening"?

Page 3-28
Line 2: Sentence beginning "Derivation of a reference…". Why? An effect can be reversible and still adverse as in its definition (p. G1). Is the thought here that some reversible adverse effects should not be counted as critical? If so, please state this clearly so that all of us can disagree with EPA.

Line 15: Delete "were reviewed as examples of systems that."

Line 22: Evaluating latency to response is a good addition to the thought behind the development of RfVs, since it is an important aspect of any value less than lifetime. However, I do not understand why reversibility gets the same attention, unless somehow you not consider reversible adverse effect as not important.

Page 3-29
B.2. (1) Line 3: Change to: "…. when to use these data.
Line 4: Delete "guidelines."

Page 3-30

Line 2: Change "….when to use the …." To "when to use these new life stage data."

Point #3: I agree with the need for latency effect; but do not necessarily agree with the need for assessing reversibility. See some comments under the charge question 4C above.

Line 21: Delete "the guidelines" and replace with "information from these studies"

Page 4-2
Line 20: The sentence "The definitions are not intended to be ……". Unfortunately, your definitions are rigid. Please see my comments under charge number 1.

Page 4-4

Box 4-2.

At the end of Acute [Oral, Dermal…] section, add "generally"

At the end of Short-Term [Oral, Dermal…] section, add "generally"

At the end of Long-Term [Oral, Dermal…] section, add "generally"

At the end of Chronic [Oral, Dermal…] section, add "generally"

Line 2 (after Box): After "..order of magnitude" add (Felter and Dourson, 1998).

Line 4: After …order of magnitude, add "generally"

Line 5: After …range of variability, add "depending on the data"

Line 6: Regarding changing the term "deleterious" to "adverse:" Great!

Line 9: Regarding dropping the term "noncancer:" Thank you!

Page 4-9
Line 22: After "adverse has not" add "always"

Line 23: Delete "Consequently" and add "In addition," the "adversity of an" endpoint…

Page 4-16
C.3. Line 6: Add additional references after (EPA, 1994, 2001B, Barnes & Dourson, 1988; Dourson, 1994; IRIS Background Document 1 (EPA, 2002) ……

Line 7: After "….single subchronic study, add "that does not define a NOAEL"

Page 4-17
Box 4-3. I guess that I am not convinced that current low, medium and high confidence ratings are any different than what minimal and robust will work out to be in practice. The emphasis on more descriptive text is good, however. The existing descriptors on IRIS are pretty minimal.

Box 4-3, Line 20: …issue of reversibility of effects and…. – Why again the focus on reversibility. Such effects are considered in the choice of critical effect currently. Are you thinking to go away from such judgment?

Page 4-20
D.1. Line 3: …that occurs… delete "to the most sensitive species" It was not originally defined this way in EPA RfD methods texts. The inclusion of this phrase suggests that EPA may choose the most sensitive species first, and not the most relevant first, in its extrapolation to humans. Whereas the current default position followed by EPA and other risk assessment groups is the choice of the most relevant species first, and if knowledge about relevance is not known, then the most sensitive species.

Line 4: Delete "of an agent". Delete the reference (EPA, 2002a.) and replace with Barnes & Dourson, 1988; Dourson, 1994; IRIS Background Document 1 (EPA, 2002).

Line 7: chemical is maintained… delete "across species" and add "between the experimental species and humans, or among humans.

Page 4-21

Please see comments regarding critical effect under response to charge question #7.

Page 4-23

Line 7: … (ARE) derivations, whereas…. As specified in existing RfD methods tests, duration adjustment is routinely done for oral RfDs. What am I missing here?

Page 4-24

I agree with applying duration adjustment to developmental toxicity.

Lines 9-14: What do you mean by this? Is C ok to use if data suggest that toxicity is due to peak exposure?

Page 4-25

Second paragraph, line 2: …C^n x T = K,… Note here that for carcinogens the appropriate equation appears to be C x T^n = K (Druckrey, 1967; Dourson, 1980). EPA historically used an adjustment base on this relationship for cancer risk extrapolation (EPA, 1980). Has EPA considered harmonization of these equations?

Page 4-27

Third paragraph, Line 3: …"For adjustment to shorter durations, the ARE….". This is reasonable.

Page 4-28

The guidance should acknowledge (for transparency) that we consider the use of the HEC to account for animal to human extrapolation in kinetics, but that this adjustment does not take into account differences in metabolism or other kinetic aspects after a chemical has been absorbed from the lung. This is particularly important as the methodology moves towards considering dosimetric adjustment factors values for category 3 gases of >1. I applaud the move toward more data-derived values, but this move should be made carefully when it removes a layer of conservatism, with particular attention to areas of variability that were previously overlooked, but may become important if the adjustment factor based on the modified approach is >1.

Page 4-33

Section D.3.d.

The discussion of oral HED should reference the work of Clewell and colleagues on how the default varies with whether the toxic agent is the parent, a reactive metabolite, or a

stable metabolite:
Clewell HJ 3rd, Andersen ME, Barton HA. A consistent approach for the application of pharmacokinetic modeling in cancer and noncancer risk assessment. Environ Health Perspect:110(1): 85-93. Jan. 2002.

I would recommend that EPA place a priority on the development of risk assessment guidelines for immunotoxicity, liver and cardiovascular effects. Currently, test guidelines are available for at the least the first area, but it is sometimes difficult to interpret such guidelines for risk assessment.

Page 4-34
Table 4-3. So after all of this nice discussion, I was expecting EPA to recommend dividing by body weight ratio to the 3/4s power and harmonize with the cancer guidelines, instead of a default value of 10. Be courageous! We will support you.

Page 4-35
D.4. This is a nice discussion of the extra factor for response at the BMDL/NOAEL. Missing from this discussion, however, is the suggestion to use a different UF at the BMDL depending on the severity of the effects (e.g., Dourson et al. 1985). Furthermore, a critical aspect of data interpretation is the severity of endpoints measured at NOAEL et al. Effects of low severity in few animals signifies lack of more severe effects in larger groups of animals. The judgment of NOAEL et al takes both the concept of severity and "n" into account. Your text focuses only on "n".

Page 4-36

Box 4-5. To what does the footnote "a" refer in the title?

Page 4-37
D.5.a The recommendation to not derive a value if there are 4 or more full areas of uncertainty is not consistent with the definition of a minimum database. A chemical for which there is a LOAEL in a single chronic study, supported by acute studies, would meet the minimal database requirements for a chronic value, but would require 4 full areas of uncertainty.

Lines 5-11: At least 2 concepts in here are incorrect: (1) $UF_L$and UF_H are variability factors in that they both account for going down a dose response curve, which for quantal effects is simply a reflection of sensitivity (or variability in response) among individuals. UF_A, UF_S and UF_D are uncertainty factors in that they simply go from a known dose response curve to a hypothesized one. The same risk (or its lack) is found after the use of each of these latter factors. (2) EPA does not use a factor for intra-human variability; nor does anyone else. This is reasonable if the RfV is averaged over some period of time. (3) Presumption that the UF_H is no more than 10 is incorrect. Please see a recently accepted publication (Dourson et al., 2002) that explains this in more detail.

Lines 16-17: Delete "less than lifetime exposure to lifetime exposure, i.e." and add "one

duration of exposure to another."

Line 18: Delete "uncertainty" and replace with "variability"

Last line on page: after …IRIS documentation – add "continue to"

Page 4-38
Line 6: After "is inappropriate:" you may wish to add a sentence: This is in keeping with current EPA practice (Dourson, 1994).

Line 9: Delete "would" after "UF" . Reword the sentence to read: In the case of the RfC, the maximum UF is generally considered to be 3000, whereas the maximum is generally considered to be 10,000 for the RfD.

Line 10: The sentence beginning "The RfC" should be reworded as follows: The RfC methodology (EPA, 1994) includes a DAF to account for……."

Line 18: Add "more than" to the beginning of the line so it reads "…more than the full 10-fold UF in four areas of extrapolation." Delete the "or more"

Page 4-39
Lines 7-8: ….adequate in most cases….. this phrase is not needed. If someone can find an RfD that is not protecting children, then the RfD needs to be revised. If managers want to add an additional FQPA factor for risk management reasons, or exposure concerns, this should be acceptable. If the manager wants to add an FQPA factor for toxicity concerns, however, he needs to go back to the RfD and revise it. The RfD is the dose that protects every sensitive subgroup. It does not need to be further lowered for toxicity reasons, although it should be revised if needed.

Page 4-40
Line 5: After "applying a factor greater" add "or less". Begin next sentence with However, unless…..

D.5.c. Intraspecies UF – This is a very nice write up!

Page 4-41
Line 9: See two new papers: Dourson et al. 2002; Scheuplein et al. 2002 regarding this paragraph.

Page 4-42
Line 5: Insert "also" after "It is"

Line 8: Add also the references "Dourson & Stara, 1983 or "Dourson et al. 1996."

Line 16: Add "and severity" before response and delete "and the NOAEL"

Line 17: Delete the word "reducing" and use "changing". The default value for this factor is 10-fold. But you really do not tell us how to do this. What if the slope is 60 probit units per 10-fold decrease in dose? What if slope is 1 probit unit? What if the magnitude of effect is 100%? What if it is 10%? What if severity of response is maximal? What if it is reversible? You need to give a little guidance here for folks to fully understand what this is all about.

D.5.e. The description of the importance of different study types in identifying the critical effect (and the implications of missing these studies on the choice of uncertainty factors) is confusing, and needs to put greater emphasis on the importance of the bioassay in a second species.
For example, the second paragraph, first line: Should read: In respects to toxicity uncertainty, the additional 10-fold…

Second paragraph, line 2: After "1996 FQPA is " – add "the same as": and delete "similar to." This paragraph is almost true. For example, a factor of 10 is currently used only when both the developmental toxicity and reproductive toxicity studies are missing along with a missing second species bioassay. In the case of the prior two types of bioassays missing but with two toxicity bioassays in different species, a factor of 3-fold is normally used. This is based on EPA analysis described in Dourson et al. (1992) that you reference in this paragraph.

Page 4-43
First new paragraph
Line 1: Delete "raise suspicions of"

Line 2: Delete "developmental toxicity and". Delete "DNT" and add: developmental, neurotoxicity, immunotoxicity, cardiovascular

Line 3: Delete "studies developmental" immunotoxicity, delete "studies, developmental" carcinogenesis delete "studies".

Line 4: Delete "developmental." Give other examples here! Developmental studies are just one of several that might be needed.

Line 7: After…"in the database and on" add "professional judgment as to:

At end of paragraph add: The default value for this factor is 10-fold.

Section D.5.f. The text needs to be clearer here and in the discussion regarding how this extrapolation relates the composite UF and the choice of critical study. For example, what is the approach if a subchronic study identifies a lower LOAEL than the lowest NOAEL/LOAEL boundary identified in a chronic study? Consideration also needs to be given to overlap of uncertainty factors. Under current practice, four full factors of 10 result in a composite UF of 3000. This is a reasonable approach that has theoretical support by EPA (Swartout et al., 2001) and should be retained.

Line 3: Add "generally" after…"UF is"

2nd paragraph. The specific use of a UF……The use of all these UFs is reasonable. What is your point here?
2nd paragraph, Lines 3 & 4: Guidance for… This is true for UF_H & UF_A. Who is doing this for other UFs? Also, see Dourson & Stara, 1983 and Dourson et al. 1996.

2nd paragraph, line 7: After "reference value is" add "generally"

Page 4-44
Section D.5.g.
Second sentence: Not true. See Barnes & Dourson (1988) where an example is given for the use of a modifying factor, or Dourson 1994.

Second paragraph, last sentence: I am not sure that I agree with this. This statement is certainly not true for nitrite.

Fourth paragraph: I agree to drop the use of the MF, but the principal reasons for my agreement are its infrequent use and potential for abuse.

Page 4-45
First sentence: After "…for the use of" add "Compound Specific Adjustment Factors (CSAFs)"

Page 4-46
Section D.6.b. Cite Baird et al. (1996) and Swartout et al. (2001) here.

Page 4-49
Figure 4-3. What about other effects? Or were these plots only for the critical effects?

Page 4-50
Section D.7.c.i. line 11: "no adequate DNT study" state why the lack of this study is of concern. Is there something about this chemical that would lead you to believe that this test is needed?

Page 4-52
Last sentence: What is this? EPA or others do not do this for RfDs nor RfCs. Let's chat if needed for clarification.

Page 4-53
Table 4-4: Column "Type of Effect". Spell out NT, DT, RT. No need to say toxicity after each one.

Page 4-54
Table 4-5. Column "Type of Effect". Spell out NT, DT, LT, KT. No need to say toxicity after each one.

Appendix A

Page A-1
A.1. Define "screening" – many endpoints that might be thought to be sensitive seem to be described as screening.

Section 1. Please see two recent publications on this very topic (Scheuplein et al, 2002; and Dourson et al. 2002).

Page A-2
Last two sentences of #3. I am not sure what this means. Would you please use a few more words to describe the thoughts here.

Page B-1
Chemical X Case Study: More information needs to be provided regarding why a database factor of 10 was chosen. What are the specific limitations that lead to choice of 10, and what data would be needed to reduce the factor to 3 (or less)? In particular, the use of a full factor of 10 when 2 bioassays are available from this route and reproductive toxicity data are available from the inhalation route (for which one could do a rough estimate of internal dose) is contrary to current practice. It appears that the factor of 10 is based on the absence of functional neurotoxicity testing, but given the specialized nature of the missing data, much more information for the rationale (or a reconsideration of the choice of the database UF) is needed.

How are the chronic oral UFs consistent with the guidance indicating that a database that requires 4 full UFs is inadequate for the derivation of a reference value?

In discussing the array, the guidance should address looking across durations. For example, if the critical effect is identified in a subchronic study and that endpoint is not adequately evaluated in a chronic toxicity study, the subchronic study should be considered as the basis for the development of the chronic RfD.

Page B-8
First sentence: "is limited." You are calling this a limited database? I suggest a random re-reading of IRIS files by EPA to remind itself on what a limited database really is.

Glossary

Page G2, Critical Effect. An older definition of this term on IRIS did not have the phrase "most sensitive species." See a previous comment as to why EPA may wish to drop this phrase.

Page G2, Chronic exposure. One does not have to define each day of the experimental duration as either acute, short term, subchronic (i.e., longer term), or chronic, as EPA is now proposing. One can define these terms approximately, as has been done by EPA in the past, and then refer to exposures between definitions in such fashion. Thus, one can judge a one year exposure in dogs as "between subchronic and chronic, or a 3 day exposure in rats as between acute and short term. EPA thus avoids needless arguments as to the characterization of a particular study, e.g., is a 9-month study in rats chronic? Is a 45-day study in dog long term? Since either characterization will always be superceded by a risk assessor's judgment of the appropriate uncertainty factor for study length, avoiding the argument seems very sensible.

Page G4, Hazard Characterization. I do not see the practical difference between this term and the older one on "Hazard Identification."

Page G5, LOAEL: What are the reasons for dropping the "statistically… significant" phrase from this definition, and that for the NOAEL?

Page G6, non-linear dose response: So does this imply that you are not confident in supra linear responses?

Page G7, RfV: Suggested insertion somewhere in the definition…order of magnitude, although depending on the underlying data, this precision may be greater or less.

Page G11, Uncertainty/Variability Factors (UFs): suggested revision follows

One of several, generally 10-fold default factors, used in operationally deriving the RfV from experimental or epidemiological data. The factors are intended to account for

1. the variation in sensitivity among the members of the human population, i.e., inter-individual variability;
2. the uncertainty in extrapolating animal data to humans, i.e., interspecies uncertainty;
3. the uncertainty in extrapolating from data obtained in a study of a given duration to that of a longer duration, e.g., extrapolating from subchronic to chronic exposure;
4. the uncertainty in extrapolating from a LOAEL (or $LOAEL_{HEC}$) rather than from a NOAEL (or $NOAEL_{HEC}$), or from a BMDL (or BMCL) of significant toxicity; and
5. the uncertainty associated with extrapolation when the database is incomplete and

the critical effect is uncertain.

Page G11, Variability: This definition refers to both exposure and response variability as inter-individual and intra-individual variability. However, EPA uses the same phrase, inter-individual variability, in its definition of the uncertainty/variability factors. In this latter case, EPA's intent is to refer to response variability. Obviously, EPA needs to avoid confusion by changing one of these definitions. I suggest that EPA avoid confusion by removing the exposure text from the definition of variability.

Review by
Pat McGinnis
Syracuse Research Corp.

Patricia M. McGinnis, Ph.D., DABT
Syracuse Research Corporation
June 14, 2002
Pre-Meeting Comments

A REVIEW OF THE RFD AND RFC PROCESSES

General Comment

The EPA is to be commended for their time and effort in identifying and tackling some very tough issues in toxicology and risk assessment. The document covers a lot of ground, is generally well-written, acknowledges that the agency has not discussed or resolved all topics, proposes some new approaches and alternatives, and identifies lots of future directions for research. Reference to more of the literature on the development of the RfD (e.g., Dourson and Stara, 1983; Barnes and Dourson, 1988) would help the reader put some of the issues in historical context. The case study for Chemical x could be improved by better reflecting more in-depth weight of evidence analysis as described in Chapter 4.

Charge Questions

1. Please comment on the recommendation to derive less-than-lifetime reference values in addition to the chronic RfD and RfC? Is the need for these values adequately justified in the report?

There is clearly a need in various EPA program offices to develop reference values for different duration exposures and routes. This agency-wide effort will standardize the definitions and approaches used. Putting this information in the IRIS database will make the assessments publicly available.

The four duration categories recommended correspond fairly well to the typical durations used in laboratory studies. While the definitions should be standardized, some flexibility needs to remain, such that a 33-day study (for example) could be used for derivation of a short-term reference value. Under the chronic definition, it seems that a 4-month study in rodents would be considered a chronic duration study and could be used to estimate effects of lifetime exposure (presumably without an associated UF). This doesn't seem appropriate; could EPA further elaborate on the basis or data supporting this? Under current practice, such a study would be considered subchronic and a UF applied for extrapolation to chronic duration.

It is not clear how the acute inhalation reference value would dove-tail with the three AEGLs; perhaps this could be explained more. AEGLs are set to identify levels above which there may be a nondisabling injury, a disabling injury, or lethality and which are set for very short exposures (as short as 10 minutes). The acute inhalation reference value may apply to the AEGL-1, but the AEGL-2 and AEGL-3 are concentrations above the threshold region.

2. Please comment on the revised definitions for reference values.

The definitions seem reasonable, with the exception of the last sentence of each. The sentence, "The application of these factors is intended to provide an estimate centered within an order of magnitude." needs to be clarified. I agree the reference dose/concentration is not a bright line and there is variability and uncertainty associated with it, but I am not so sure it is *centered within* an order of magnitude. Several papers by Dourson and colleagues (e.g., Dourson and Stara, 1983) have discussed the historical and biological basis for the 10-fold UFs and imprecision of the RfD; perhaps Dr. Dourson can review this at the meeting. My understanding is that the precision of a reference dose/concentration is related to its UFs (e.g., 10). The UFs reflect variability or uncertainty in extrapolating within a population, across species, down dose-response curves (LOAEL to NOAEL), across durations, etc. For some extrapolations, such as the animal to human, one is generally extrapolating from the average animal to the average human (hence, the idea of centered?). For other extrapolations using human data (interhuman variability), one may be extrapolating from the dose producing a median or mean response human to the dose that may produce a response in susceptible members of the population (in the tail of the response distribution) and there might be more or less than10-fold variability. I think this is a more complex concept and needs to be discussed at the meeting. It also may need further discussion in the document and reference to earlier literature.

Division of exposure into these four durations seems reasonable and would be consistent and useful to agency programs, such as development of drinking water health advisories.

I agree the terminology should be standardized but am somewhat uncomfortable with use of the RfV and with proposed subscripts. The "reference" value concept is recognized by the scientific community as applying to chronic (lifetime) exposures. Retaining the "reference value" with qualifications as to route and durations may lead to confusion. Perhaps changing the terminology to exposure estimate, exposure reference, or another term with the route and duration qualifier would be less confusing and signal a change in definition (e.g., acute oral exposure reference, AOER, or chronic inhalation exposure reference, CIER). If the agency retains "reference value" then they may want to consider not subscripting the duration and route qualifiers. Subscripts tend to get less attention, get dropped-off, and are prone to editorial errors. The meaning of these subscripts are very important and needs to standout.

3. Please comment on the recommendation that endpoint-specific reference values should not be derived.

I agree that endpoint-specific references values should *not* be derived and that developmental toxicity should be treated as one endpoint in the suite of endpoints evaluated and weighed for derivation of a reference value to protect the whole individual throughout his/her lifetime. The reference value should be sufficiently protective for all endpoints. However, Table 4-4 may give some less informed risk assessors the appearance that endpoint-specific reference values are derived.

4. A review of current guideline study protocols and approaches to testing was conducted to determine what information is currently developed to support the derivation of less-than-lifetime reference values. Please comment on the life-stage approach taken in this review, as well as the recommendations for alternative testing approaches and strategies for developing a targeted testing strategy to support setting less-than-lifetime reference values as well as chronic values. Also, please comment on a proposal from the Office of Children's Health Protection to request a study by the National Academy of Sciences to take a fresh look at toxicity testing approaches and strategies based on this and other reports (e.g., ILSI, 2001[1]).

As a general toxicologist/risk assessor, I would like to have more data for endpoints and life stages and more mode/mechanism of action data. Chapter 3 appears to do a thorough job reviewing the gaps in the current testing guidelines and proposing possible alternatives. I agree with the document that the complexity of some of the proposed protocols will require significant skill to conduct and manage. There will also be a substantial resource impact in conducting more complex and labor-intensive studies. More complex protocols will require significant time and skill for the risk assessor to interpret.

In addition to limitations in evaluation of various endpoints and lifestages, another limitation in application of some studies to risk assessment is dose/concentration spacing. Studies with free-standing NOAELs or with multiple NOAELs and a FEL are of limited usefulness. In revision of guidelines, more consideration may want to be given to range-finding studies and setting of appropriate experimental dose levels.

5. Please comment on the recommendation to include duration-adjustment for inhalation developmental toxicity as for other health endpoints.

I am not so sure there is sufficient scientific basis to support this approach for all chemicals or all types of developmental effects. I will defer to my colleagues with expertise in developmental/reproductive toxicology and pharmacokinetics, but would like to offer a few thoughts. Clearly, the biological basis for continuous exposure adjustments

[1] http://www.ilsi.org/publications/pubslist.cfm?pubentityid=8&publicationid=305

is an area that needs additional research and goes beyond the developmental endpoint. There may be more, recent literature that could be discussed in Section D.2 to help the reader. Peak concentration may be a better dose metric for some outcomes than others. For example, Boyes et al. (2000) examined three measures of neurotoxicity in adult male rats and found that, for acute effects of TCE, the functional changes observed correlated with arterial concentrations at the time of testing. As the document points out, in other cases, AUC may be the better dose metric.

On the one hand, it seems unusual that developmental studies should be treated differently for concentration adjustment than other endpoints. However, this is a period of rapid growth and differentiation. A basic principle in developmental biology is that there are critical periods or stages of development (e.g., limb bud formation). Some of these stages probably take a few hours and some (general growth) take all of gestation. Use of C x t does not seem appropriate for processes with short critical periods or for chemicals with short half-lives. It might be appropriately applied for chemicals with long half-lives or for those that produce a generalized, non-specific outcome such as growth retardation. In order to make the call as to what is the most appropriate dose metric, however, one needs information on pharmacokinetics and mode of action of the chemical.

In cases were there are insufficient data, I think it is appropriate, as a matter of policy, to apply the duration adjustment to developmental studies as a default procedure, so to be conservative and protective of the public health. However, in cases where there may be data to support other than the default duration adjustments, risk assessors should be encouraged to synthesize and present this information (not just for developmental outcomes).

6. Please comment on the recommendations in the report for using a weight-of-evidence approach for hazard characterization (Chapter 4, section C.2.), and for expanding characterization of the extent of the database and using a narrative to describe the database rather than a confidence ranking (Section C.3.). Is the discussion of weight of the evidence clear in terms of how it would be used in characterizing the database? Also, please comment on the consideration of this information in the derivation of reference values.

Good hazard characterization critically weighs all the information in the database, considers what information is lacking, and clearly and transparently lays this out. The discussion in Section C.2 does a good job reiterating current weight of evidence approaches (e.g., RfC methodology, 1994) and reminding the risk assessor of several additional points that should be considered (life stages, susceptible populations, human relevance, etc.). The characterization of susceptible subpopulations section is particularly well presented.

A narrative approach for the extent of the database, with emphasis on strengths and limitations, the types of data available as well as data gaps is a good idea and will be an improvement in many cases over the confidence statements. A narrative description of databases as minimal, robust or as somewhere along the minimal-robust continuum has advantages over specifying specific types of studies comprising a minimal database or using a one-word descriptor.

The narrative description of the database for the example Chemical x in Appendix B (page B-8) does not incorporate some of the points for database evaluation presented on pages 4-19 and 4-20 nor some of the information in the summary of health effects information (starting on page B-1). In addition to delineating what studies were conducted by what route, what duration and for what life stages, the narrative description could be improved by including statements about the overall adequacy and quality of the studies/database, consistency (or inconsistency) of effect (e.g., neurotoxicity for Chemical x) across studies and species, similarity of effects across exposure routes, relevance to humans and extent of the mode of action data.

The weight of the evidence and extent of the database should be considered in identifying and supporting the biological endpoints most relevant to humans for reference dose derivation. Consideration of the extent of the database also factors into UF selection.

The endpoint by endpoint discussion in the *Summary of Health Effects Information* in Appendix B does not sufficiently pull together or weigh the database. When presented in this format, all snippets of information tend to be treated equally even though some data may be much more reliable than other data. The current IRIS toxicological review approach does a better job of presenting a detailed assessment and critical evaluation of studies and a weight-of-evidence discussion that draws conclusions about sensitive endpoints and effects levels based on the information presented in the study descriptions (as is currently the practice), and should be retained.

7. Please comment on the recommendations for use of an exposure-response array and derivation of sample reference values to select the point of departure (POD) for the final reference value. Also, please comment on consideration of the nature of the effects for different endpoints, durations, timing and routes of exposure in selection of the POD.

A visual presentation of the data is a good idea. However, the exposure-response arrays such as those on page B-9 do not clearly represent the careful weighing of the data or consideration of human relevance. In the figure, each point (study) appears equal. The reader cannot distinguish studies of greater or lesser quality or studies examining more or less severe endpoints. ATSDR has tried to separate effects into "less serious" and "serious" on their graph, so that the reader has some indication of the type of the effect and dose-response, but this adds a layer of complexity to the graph and a set of rules as to what is a serious effect and what isn't. While Chemical x has a relatively small database, the reader needs to consult Table B-1 to aid in interpretation of the array. Arrays for chemicals with more robust databases will be difficult to interpret as there will be

multiple sets of data points for a particular species and endpoint (e.g., liver effects in mice) with overlapping NOAELs and LOAELs that will have to be keyed to the table to be understandable. Inclusion of only the most relevant, adequate studies may alleviate some of the clutter. EPA may want to consider generating some exposure-arrays for larger database chemicals to see if these will really be helpful to the reader. In the course of evaluating a database, the toxicologist finds themselves lining-up studies and NOAELs/LOAELs in a list. Perhaps a table presenting the most suitable capturing the basics of the study design, endpoints, limitations, NOAELs, LOAELs, BMDLs would also serve the objective of showing the range of numerical values available for each route and duration of exposure. I have seen some good tables in recent IRIS toxicological reviews and DWCDs. Another possibility EPA may want to explore is presenting a graph of dose-response curves of only the most relevant studies/endpoints (the RfC methodology on page 4-72 presents an example).

The POD should be selected on the basis of an evaluation of all relevant endpoints. However, the current approach appears as if (Table 4-4; B-4) every potentially suitable study/species/endpoint is carried through to reference value derivation and then the lowest one for each duration is selected. The careful, critical evaluation of study and data quality, and weighing of the evidence in the database seems to have been disconnected in this step. I agree that there should be less emphasis on identifying *the* critical effect in *one* study and more emphasis on synthesizing, weighing, and weaving together of all biological information. For example, Table 4-4 (and the exposure array on page B-9 and write-up on page 4-50) loses the information presented on page B-2 that long-term and chronic neurotoxic effects have been *consistently* reported in *several* occupational studies. Consideration of only the NOAELs also leaves out information that needs to be considered at least qualitatively in picking appropriate data for the POD. Consideration needs to be given to the fact that (Tables B1 and B-2), for acute exposure, the rat neurotoxicity $LOAEL_{HEC}$ (dose-related hyperactivity, ataxia, hypoactivity, narcosis) is less than the rat developmental (decreased performance for neuromuscular ability in pups) or reproductive (reduced litter size and survival of offspring) $LOAEL_{HECs}$ or that, for short-term duration, the human LOAEL for neurotoxicity (headache, dizziness, incoordination) is less than the the $NOAEL_{HEC}$ for rat reproductive effects (reduced litter size and survival). These data inform what endpoint(s) is most appropriate for dose-response assessment and may have been considered in this example. A more extensive narrative than that on pages B-10 to B-13 giving consideration of the nature of the effects for different endpoints, durations, timing and routes of exposure for the POD would present a more-biologically based rationale for the POD and be helpful to the reader.

In the example for Chemical X, it seems awkward to choose the RfV for rat developmental toxicity over the human neurotoxicity data because it generates a lower value due to greater uncertainty. A few questions come to mind that EPA may want to consider. Is the lowest RfC (or limiting value) in the case of Chemical x really more protective? Most of the values derived vary by less than an order of magnitude yet UFs range from 30 to 300. Isn't basing values on the lowest $NOAEL_{HEC}$ consistent with the critical effect concept of Barnes and Dourson (1988), "The critical endpoint used in the dose-response assessment is the effect exhibiting the lowest NOAEL"? Does basing

values on the lowest $NOAEL_{HEC}$ overlap to a certain extent with the database UF? For example with Chemical x, the acute, short-term and longer-term RfVs are based on the rat developmental study because it has the lowest duration-adjusted $NOAEL_{HEC}$ and is more protective of the developing individual as well as the adult. However, a 3-fold UF is added for database deficiencies because there is no adequate prenatal developmental toxicity studies in two species and no adequate developmental neurotoxicity studies.

An interesting exercise that EPA might want to consider is to take several recent toxicological review chemicals and apply the limiting RfV approach to the data (at least for chronic duration) to see what difference there is in the values derived. The limiting value approach appears to require significantly more effort and may not change underlying basis or numerical value for the RfC/D.

8. Please comment on several recommendations concerning the application of uncertainty/variability factors. Are there additional data or analyses in the literature not cited here that can be used to strengthen the recommendations? Should other factors be considered in the application of uncertainty/variability factors?

These are all good recommendations. I agree with the recommendation to limit the UF to 3000. Research should be encouraged for chemicals with insufficient databases. In order to be protective of public health, susceptible subpopulations must be considered; however, I think it may be difficult to obtain sufficient data to depart from the default UF. (Should "data set" be "database" in Recommendation B?) I agree with the Panel's Recommendations C and D. Further, the choice of UFs needs to be clear and transparent. The MF should be discontinued. The intent of the MF to capture database completeness is subsumed in the UFs. I agree with Recommendation F that EPA should review current practices and encourage research for CSAFs (see also comment #5). The assumptions underlying the current default duration adjustments need to be articulated in risk assessment documents.

Review by
Bonnie Ransom Stern
BR Stern Associates

General Impressions:

The EPA report, *A Review of the Reference Dose and Reference Concentration Processes*, is a very well-written, comprehensive document. A review of the reference dose and reference concentration processes is much needed, and I commend the Agency for their thorough and thoughtful efforts on this herculean task. The Agency is also to be lauded for its adjunct efforts, including exploration of uncertainty and variability, exploration of aging and toxic response, and exploration of perinatal pharmacokinetic – all reports prepared by Versar, Inc. during the past two years – as well as exploration of the issues raised by the EPA's 10X Task Force.

Many of the Technical Panel's recommendations are appropriate; however, some of the recommendations require reality checks. Some of the expanded testing recommendations and cautionary statements regarding uncertainty factors and compound-specific adjustment factors appear to be based on the Agency's concern for protection of children, and to a lesser extent, on protection of the elderly. This is understandable, given the widespread Congressional and public health concern about these susceptible subpopulations. However, expanded testing can be both costly and impractical without the development of a tiered strategy, with triggers, for identifying the chemicals or classes of chemicals that might be associated with life-stage susceptibility. Similarly, the development of data-derived adjustment factors for reducing uncertainty has been ongoing for many years, and the Agency should encourage this type of data acquisition, rather than approaching this area with caution and conservatism.

Response to Charge Questions:

1. Please comment on the recommendation to derive less-than-lifetime values in addition to the chronic RfD and RfC. Is the need for these values adequately justified in the report?

The recommendation to derive less-than-lifetime values, in addition to the chronic RfD and RfC values is appropriate, and is well justified in Chapter 2 of the report. There is a need for duration-adjusted reference values, to deal with variation in exposure duration and concerns about adverse health effects occurring as a result of less-than-lifetime exposures. As noted in the report, the development of less-than-lifetime values is currently being done by individual program offices, as well as ATSDR These include the AREs for air pollutants; AEGLs for accidental air releases; acute, intermediate and short-term RfDs for pesticides; and drinking water short-term, longer-term, and lifetime health advisories. The availability of Agency-wide, less-than-lifetime values will encourage program offices to use these values where applicable, resulting in a more consistent and thorough approach to protection of public health.

2. Please comment on the revised definitions for reference values.

I think that the revised definitions for the reference values are clear and concise, with the exception of the last sentence: *"The application of these factors is intended to provide an estimate centered within an order of magnitude."* It is not clear whether "centered within one order of magnitude" means that the entire range spans one order of magnitude (i.e., +/- one-half an order of magnitude on either side of the centered estimate), or two (i.e., +/- one order of magnitude). This statement should be clarified.

3. Please comment on the recommendation that endpoint-specific reference values should not be derived.

The Technical Panel felt strongly that all relevant endpoints should be considered in the derivation of each duration- and route-specific reference value, thus ensuring that reference values are derived to be protective of all types of effects for that route and duration of exposure. In my judgment, this is not warranted. All relevant endpoints are considered in the hazard characterization, which gives a broad characterization of the toxicokinetics, mode(s) of action, target organ toxicity, and other information. The critical effect is defined as "the first adverse effect, or its known precursor, that occurs to the most sensitive species as the dose rate of an agent increases" (USEPA 2000). The critical effect is traditionally selected for dose-response assessment based on the rationale that it will be protective of all adverse effects observed at higher doses. Modeling all relevant endpoints is a cost- and labor-intensive activity, not only for registrants and industry but also for IRIS and program office chemical managers, and it is not clear that this extra effort will change the bottom line, (i.e., the final reference value). It did not change this value for Chemical X , as described in Table 4-4. Whether this approach is scientifically supported by adequate examination of existing data bases for "strawman" chemicals is not discussed in the report.

I do not agree that all relevant endpoints should be modeled or analyzed statistically for dose-response, using the Technical Panel's definition of relevance. There are clearly effects that occur only at higher or the highest dose tested, and these will be protected against by the selection of more sensitive endpoints (or the most sensitive endpoint) at lower doses. In my judgment, several endpoints should be modeled separately and the results compared only when there is a question of which is(are) the most sensitive endpoint(s), or the most relevant human endpoints, based on available data.

4. A review of current guideline study protocols and approaches to testing was conducted to determine what information is currently developed to support the derivation of less-than-lifetime reference values. Please comment on the life-stage approach taken in this review, as well as the recommendations for alternative testing approaches, and strategies for developing a targeted testing strategy to support settling less-than-lifetime reference values as well as chronic values. Also, please comment on a proposal from the Office of Children's Health Protection to request a study by the National Academy of Sciences to take a fresh look at toxicity testing approaches and strategies based on this and other reports.

The Technical Panel is to be commended for their cogent and thorough review of life-stage assessment and current testing protocols. The recommendation to develop protocols for acute and short-term studies that provide more comprehensive data for setting reference values. is a good one. The next four recommendations (2-5) are also good but vague. Guidance will first have to be developed to decide how and when. It seems sensible to develop a tiered strategy of determining what information is needed to predict which specific studies might provide useful data to assess whether a particular chemical or class of chemicals is likely to exhibit life-stage or genetic susceptibility; demonstrate reversibility of effect following discontinuation of exposure; or exhibit latency to effect. This could be developed by designing a set of triggers, similar to those used in the PMN Program in OTS that determine whether additional studies, and which ones, are needed for chemical registration. SAR, pharmacokinetics, hormonal, *in vitro* studies, and a basic battery of screening assays might be useful in initially identifying potential targets.

It should be noted that most studies designed to expand the collection of toxicity information, including the alternative toxicity testing protocols discussed in the report, will involve a large outlay of resources – they are both costly and labor-intensive to conduct, review, and evaluate. It is important to limit the universe of tests on a chemical-specific basis by using existing data, as well as the results of screening assays, to inform decision- making.

Alternative testing protocols:

1. Alternative acute toxicity testing protocol: This protocol will adequately provide the data needed for the development of an acute reference value. Similar protocols have been used with a number of compounds, although the studies published to date usually target specific organ systems; examples are the extensive set of studies conducted by Linder and colleagues on the male reproductive tract toxicity of haloacetic acids, and the neurotoxicity studies of Moser and colleagues, also on the haloacetic acids.

2. *Expanded chronic/carcinogenicity study*: This study protocol would involve an enormous amount of planning, as well as resources, without focused information that has some predictive power to reduce the additional work load – i.e., the number of additional groups required, additional sacrifice times, additional statistical analysis, data interpretation, reporting, and evaluation. The logistics are overwhelming, and the practicality of this type of study is questionable.

The effects of exposure during puberty has not been identified as a specific exposure period that should be evaluated as a stand-alone. This is a critical period of rapid growth and development and the test substance may exhibit a latency period, with adverse effects being observed only later in life. For example, an increase in the incidence of breast cancer has been found in maturity in women who had been exposed to ionizing radiation at Hiroshima and Nagasaki during their adolescence. Some chemicals may act similarly. It should be noted that very few environmental substances have been identified as *in*

utero or early life-stage (infancy, childhood, adolescent) carcinogens in humans when exposure has been confined to these periods. Carcinogenicity developing as a result of continuous exposure to a test substance, from *in utero* through maturity, may be due to the longer exposure duration, or exposure during one or more crucial stages of development, and/or both; it would be difficult to distinguish causally between these scenarios.

I do not see the logic or value in extending the study duration to a period of 3 years. The carcinogenicity studies of Maltoni and colleagues clearly demonstrate the difficulties involved in extending the studies for an additional year. The duration of these studies is typically extended to correspond to the natural life span of individual animals, resulting in animals dying at different times during the third year. Often there is differential survival among groups, due to large individual variability at this life-stage, and comparisons among groups are confounded by an increase in the diseases of old age, such as cancer. Reproductive senescence also has high individual variability among animals. Extension of the study duration from 2 to 3 years would seriously compromise the use of historical control values on tumor rates or other disease rates. . The use of feed restriction to maximize the number of animals available for *in vivo* and post mortem assessment of aged animals introduces a confounder that will limit the interpretation of findings. Insufficient data currently exist to adequately describe the relationship between feed restriction and tumor incidence reduction in animals or humans, although the body of information is growing.

3. *Unified screening study*:

The logistics of this study are overwhelming. The expansion of end points is needed after criteria have been developed, but it would be necessary to develop criteria to be used to evaluate the functional and toxicologic significance of alterations in these end points as compared with controls. Some of the reproductive end points are highly variable among individuals, and would prove difficult to assess statistically (e.g., anogenital distance, vaginal opening, and preputial separation during puberty; estrus cyclicity, ovarian follicular counts and atrophy in aging female rats). Errors inadvertently made early in the study would have compounding effects over generations. Further, this study would not yield a reliable or valid estimate of exposure (e.g., daily dose) in all but parental males and females. One cannot determine actual doses to any of the F generations in the absence of knowledge of internal doses occurring *in utero* and during lactation. Therefore, this type of study has limited relevance to dose-response assessment. One also cannot distinguish between the effects of longer exposure duration verus the effects of exposure occurring only during one or more critical stages of development without a significant number of additional groups being tested and sacrificed at numerous intervals during the study period.

The use of the Sprague-Dawley rat to examine reproductive and developmental end points, including reproductive senescence, is not advisable. The normal reproductive aging processes in female Sprague-Dawley rats differ markedly from that of both F344 rats and humans. With age, female Sprague-Dawley rats undergo a process of transition from regular estrus cycling to constant estrus, which appears to be maintained by elevated levels of circulating estrogen in blood serum (that inhibits the preovulatory luteinizing hormone surges that are normally released from the pituitary gland and function to induce ovulation). In contrast, aging F344 rats transition from regular estrus cycling to persistent diestrus during which estrogen levels are reduced and progesterone and prolactin levels are elevated. Endocrine and neuroendocrine control of many reproductive functions, including those associated with reproductive aging, also differ markedly in the Sprague-Dawley rat and the human. These differences, and others, are well described in Agency documents on atrazine.

The currently- used reproductive and developmental toxicity screening assays are designed to be hypothesis-generating, not hypothesis-testing.

It should also be noted that the selection of test doses and dose spacing has a definitive effect on the identification of LOAELs, NOAELs, PODs, and BMDs, as well as the shape of the dose-response curve. Therefore, it seems to me that it would useful for the Agency to invest in developing improved protocols for range-finding studies in order to maximize the toxicity information obtained from standard studies.

5. Please comment on the recommendation to include duration-adjustment for inhalation developmental toxicity as for other health endpoints.

Duration adjustment for inhalation developmental toxicity will produce a more conservative estimate of the LOAEL/NOAEL and BMD. Even though C x t tends to be more health protective, the timing of peak exposure concentration may be more critical mechanistically.

6. Please comment on the recommendations in the report for using a weight-of-evidence approach for hazard characterization, and for expanding the characterization of the extent of the database and using a narrative to describe the database rather than a confidence ranking. Is the discussion of weight of the evidence clear in terms of how it would be used in characterizing the database? Also, please comment on consideration of this information in the derivation of reference values.

As a toxicologist and risk assessor, I would like nothing better than seeing an expansion of the characterization of the database, accompanied by a weight-of-evidence narrative. However, in order to be useful and comprehensive, it is important that these sections are consistent among chemicals with regard to presentation and integration of information. I have externally reviewed a reasonable number of IRIS toxicological reviews and

summaries during the past few years, and have extensively used these documents for a range of chemical risk assessments. There is an enormous amount of variability in the way these documents are written. Some chemicals are comprehensively reviewed, and integrated summaries of biological and toxicological information are presented in a clear and concise manner. Other chemicals are haphazardly and unevenly presented, and the documents consist of a compilation of data on numerous studies in various sections that is repeated in summary form in the weight-of-evidence characterization section, so that when one gets to the end one has no comprehensive sense of what the chemical does, where it goes, and how it exerts, or might exert, its toxic effects (i.e., its potential human, health hazards). One is then forced to skip back and forth between sections, make notes, compile one's own integration of data. Therefore, it is recommended that guidance for the structure of this section be developed to assure consistency and coverage of all relevant information in an integrated and easy-to-assimilate manner.

The Technical Panel has done an excellent job in identifying and summarizing the factors for evaluation of the weight of evidence regarding the likelihood of effects in humans (Table 4-1), factors for evaluation of evidence regarding identification and characterization of susceptible subpopulations (Table 4-2), and in developing a series of questions to help guide the assessment process (p. 4-19). I would recommend that these be used as guidance for developing a formal outline of the kinds of information to be included in the IRIS assessments. In this context, one suggestion is to divide the questions presented on p. 4-19 into two sections, one concerning data availability and the other evaluating how well the data informs human health hazard potential.
Data availability provides information on the completeness and adequacy of the data base:

1. Have adequate studies been conducted to establish the target organs/endpoints? [Are there sufficient studies?]
2. Have the effects been characterized for both sexes and all life stages?
3. Are data pertaining to potentially susceptible subpopulations available?
4. Is the route and matrix of exposure relevant to the specific reference value being derived? [This question could be answered in a table]
5. Is the duration of exposure appropriate for the specific reference value being derived?
6. Are pharmacokinetic data available? For both sexes, for relevant stages, for susceptible subpopulations?

Data informing refers to what the data tells us and whether the data are useful for health risk assessment (e.g., in terms of consistency, concordance, and human relevance):

1. Are the responses consistent across species? Are the results of the studies biologically plausible?
2. Is the animal species and strain appropriate for extrapolation to humans?
3. Is the shape of the dose-response curve consistent with the known pharmacokinetics of the test compound?
4. Are the metabolism and pharmacokinetics [I would add likely mode of action] in the animal species similar to those of humans?

5. Has the dose-response curve been replicated by or is it consistent with data from other laboratories and other test species?
6. ?????? Has the data for all relevant endpoints been adequately modeled by the BMD or other appropriate quantitative analysis to determine the most sensitive endpoints(s)? [As noted elsewhere in my review, I do not agree that all relevant endpoints should be modeled or analyzed statistically for dose-response, using the Technical Panel's definition of relevance.]
7. How well is the toxicity characterized? [This identifies data gaps.]

Responses to many of these questions require best professional judgment and transparency, and it is suggested that the questions be answered with reasonable conciseness. In this context, tables are useful. Similarly, Tables 4-1 and 4-2 provide very useful summary information. These approaches enable the reader to get a comprehensive sense of the adequacy and robustness of the data base, and can be used by toxicologists, risk assessors, research scientists, and risk managers for informed decision-making.

The guidelines in the document provide a systematic, structural logical approach to data evaluation and it is suggested that they be formalized. In my judgment, this approach would be similar in concept and intent (although more comprehensive and complex) to the epidemiologic guidelines developed by Hill (1965) and Rothman (1986) and followed by competent epidemiologists to assess evidence of causality, in the context of other available scientific information. In brief, epidemiologists ask the following questions and use the following guidelines:
1. Is the evidence biologically plausible? When the association is supported by evidence from clinical research or toxicology about biological behaviour or mechanisms, an inference of causality is strengthened.
2. Is there a temporal association? Exposure must precede the disease, and in most epidemiological studies this can be inferred. When exposure and disease are measured simultaneously, it is possible that exposure has been modified by the presence of disease.
3. What is the study precision and validity? Individual studies that provide evidence of an association are well designed with an adequate number of study participants (good precision) and well conducted with valid results (i.e., the association is not likely due to systematic bias).
4. What is the strength of association? The larger the relative risk or odds ratio, the less likely the association is to be spurious or due to unidentified confounding. However, a causal association cannot be ruled out simply because a weak association is observed.
5. What is the consistency among studies? Repeated observation of an association under different study conditions supports an inference of causality, but the absence of consistency does not rule out causality.
6. What is the specificity? A putative cause or exposure leads to a specific effect. The presence of specificity argues for causality, but its absence does not rule it out.
7. Is there a dose-response relationship? A causal interpretation is more plausible when an epidemiological gradient is found (e.g., higher risk is associated with larger exposures).
8. Is there reversibility or preventability? An observed association leads to some

preventive action, and removal of the possible cause leads to a reduction of disease or risk of disease.

7. Please comment on the recommendations for use of an exposure-response array and derivation of sample reference values to select the point of departure (POD) for the final reference value. Also, please comment on consideration of the nature of the effects for different endpoints, durations, timing, and routes of exposure in selection of the POD.

The exposure arrays presented in Appendix B is a useful visual representation of the data, which lends itself to visual interpretation of chemical effects, combining available data sets. However, ATSDR-type exposure-response arrays should not be used. ATSDR provides too much information that does not give the reader a clear visual sense of the major toxic effects, or related end points.

In the derivation of sample reference values, it does not appear that the Technical Panel's recommendations for (1) using a weight-of-evidence approach for hazard characterization, and (2) for expanding the characterization of the extent of the database and using a narrative to describe the database, are being applied to the derivation of reference values for Chemical X (Table 4-4). There appears to be a disconnect between the hazard characterization and the dose-response assessment; i.e., the hazard characterization is not being used to inform the dose-response.

It does not appear to me that biological judgment, as advocated by the Technical Panel recommendation to expand the weight-of-evidence hazard characterization and summary narrative, has been applied in the development of the reference values, or that the data have been used in a biologic context. LOAELS and NOAELs are simply identified, uncertainty factors are applied to the range of species-specific endpoints to derive a reference value, and the lowest reference value is selected as the final one for each exposure-duration study. Not all available data sets are used to select the POD. Further, inspection of column 2 (HEC) of Table 4-4, shows that for each exposure-duration category, the lowest dose identified among the relevant studies is the one used to derive the final reference value. This shows clearly that evaluation of all relevant endpoints does not need to be carried through to reference value derivation, and that the lowest NOAEL is protective of all other effects occurring at higher doses. This examples argues against the use of considering all relevant endpoints, as discussed in response to Charge Question #3..
It should also be noted that the range of sample reference values for each exposure-category is very narrow, less than an order of magnitude.

The weight-of-evidence hazard characterization does not appear to have influenced, or even been considered, for any reference value outcome. For example, neurotoxicity was reported in humans, as well as rats and mice, indicating human relevance and therefore, in my judgment, biological consideration as being a relevant, and possibly critical effect. Additionally, the LOAEL/NOAEL boundaries for each study do not appear to have been

evaluated.

8. Please comment on several recommendations concerning the application of uncertainty/variability factors. Are there additional data or analyses in the literature not cited here than can be used to strengthen the recommendations? Should other factors be considered in the application of uncertainty/variability factors?

Issue A: I agree that if there is uncertainty in more than four areas of extrapolation, the database is insufficient to derive a reference value. However, this is different than setting an upper limit of 3000 on the total uncertainty factor. An upper limit of 3000 can be achieved with four areas of extrapolation by reducing the UF in each area, or combining areas (e.g., giving a composite uncertainty factor of 10 to the use of a LOAEL instead of a NOAEL, and data base insufficiencies), a practice commonly used by U.S. and European regulatory agencies (e.g., EPA, WHO). This is often done in the absence of transparency, and indeed, scientific justification for this practice is not usually specifically described; it is often based on best professional judgment, or practical considerations.

Issue B: It is not clear to me from the description in the external peer review draft how this recommendation is supported by scientific data.

Issue C: The discussion of this issue in the report does not suggest anything new. Indeed, the exact value of the UF chosen should depend on the quality of the studies available, the extent of the database, and scientific judgment; and default uncertainty factors of 10 are recommended in the absence of sufficient data to support a data-derived value. However, Recommendation C, following on the heels of Recommendation B, suggests a very conservative approach by the Agency to the use of data-derived values, instead of encouraging the development of additional data to target reduction of uncertainty factors. It should be noted that a large amount of work has been conducted on the development of approaches and types of data that can be used to support a change in the default value, as well of supporting rationale as to why specific data support a different uncertainty factor, how uncertainty can be reduced, and what assumptions are satisfied or replaced with the use of data-derived values (e.g., work by Dourson, Renwick, and their colleagues in the published literature; Meek and her colleagues at IPCS). [See Issue E below].

Issue D: I agree with the recommendation developed for this issue. The current uncertainty factors, if appropriately applied, will be adequate in most cases to cover concerns and uncertainties about children's health risks.

Issue E: I agree with the recommendation that the use of the MF be discontinued as it rarely deviates from 1.0. I only recall seeing the MF deviate from 1.0 in cases where the data suggested, or strongly indicated, that humans were more sensitive than the rodent test model. In these cases, the MF was set at < 1.0 to account for greater human

susceptibility. It is not clear to me that this application of the MF is adequately or sufficiently subsumed in the general database UF.

Issue F: Again, this recommendation is conservative: i.e., "caution should be used in that there are relatively few data available for many substances to serve as an adequate basis to replace defaults with CSAFs", rather than encouraging. I call the Panel's attention to the guidance being developed through international efforts, led by the International Programme for Chemical Safety (IPCS), to harmonize approaches to the assessment of risk from exposure to chemicals by encouraging the development and use of compound-specific adjustment factors, using toxicokinetic and toxicodynamic data, in order to move away from the application of default uncertainty factors in risk characterization. The objective of this effort is to develop guidance, with examples, in order to encourage the development of adequate compound-specific data for replacement of default values, and to generate consistency in judgment about what constitutes adequate data. A tremendous amount of excellent cohesive and rigorous work has been conducted on this effort. The guidance document is posted on the IPCS (www.ipcsharmonize.org) and Health Canada (www.hc-sc.gc.ca/hecs-sesc/exsd) websites. At this time, initial consensus on the adequacy of data to replace default uncertainty values has been achieved; however, the guidance document will be revised, and there is a need for communication and engagement with other risk assessors and scientists. I encourage EPA to participate fully in this process.

Review by
Pam Shubat
Minnesota Department of Health

A Review of the RfD and RfC Processes
Charge Questions—External Peer Review
Pamela Shubat

1. Please comment on the recommendation to derive less-than-lifetime reference values in addition to the chronic RfD and RfC. Is the need for these values adequately justified in the report?

There is ample and obvious justification for developing and applying less-than-lifetime values if data show that high exposures for short time periods do not cause harm over a lifetime so that the less-than-lifetime values are as protective as chronic reference values. If the technical panel believes that this is a true statement, it would serve as ample justification for developing less-than-lifetime reference values. The only place where this appears to be written is in the proposed definitions ("exposure…to the human population that is likely to be without an appreciate risk of adverse effects for a lifetime"). An explicit statement of this is desirable.

It is obvious (while not stated) why risk managers would like to find ways to justify allowing exposures that exceed chronic reference values. Risk managers may want to apply higher exposure values in the short term when the costs of interim risk management are high, or technology is lacking and delay in implementing more stringent control levels is desirable. Since programs have developed higher reference values for special applications, there is obviously a need for such values. However, it is not clear whether risk managers believe that a correctly developed chronic value (i.e., one that takes into account all effects, all sensitive time periods) will always be protective of short-term exposures. Public health will be protected and will likely be overprotected as long as there is no adjustment (e.g., duration adjustment) in applying a chronic value to an acute exposure. This practice will be health protective regarding the exposure to a particular chemical (e.g., mercury in contaminated fish), but it may not be optimal in terms of weighing and managing alternative risks and benefits (e.g., health benefits of fish oil and protein) when difficult choices have to be made.

One potential justification, which did not appear to be explicitly stated, is that the EPA finds it desirable to standardize the existing less-than-lifetime reference values developed by different programs for specific applications. It was not clear whether this was the compelling reason for establishing a single strategy throughout the agency.

The EPA could strengthen the review by writing a clear statement of whether or not the chronic RfV would be protective if applied, without adjustment, to any exposure. In addition, this statement should explicitly state the reasons that programs believe a chronic RfV is not a desirable and protective value that can be applied to any exposure. Finally, EPA should explicitly state that less than lifetime values should be developed if data show that high exposures for short time periods do not cause harm over a lifetime so that the less-than-lifetime values are as protective as chronic reference values.

2. Please comment on the revised definitions for the reference values.

The technical panel's review of the current procedures for developing less-than-lifetime reference values was helpful. The suggestion for standardization appears to be consistent with the various program values already in use. The selection of the three less-than-lifetime exposure periods (24 hour, 30 days, or 7 years) could be presented in terms of the balance between the testing data that are available (i.e., minimizing the extent to which data from toxicity studies would need to be adjusted to match these exposure periods) and the application needs of programs (specific less than lifetime exposure scenarios used by programs).

It is not explicitly stated that the EPA would create *only* 24-hour acute, 30 day short-term, and 7 year longer-term RfVs. In the discussion of duration adjustment for acute reference values (D.2.c., 4-25), it is implied that 1- and 8-hour acute RfCs may be developed as needed. The EPA technical panel should clarify that that discussions related to other time points (i.e., other than 24 hour, 30 day, and 7 years) are illustrative or represent existing practices in various programs and are *not* recommended.

Wording for the definition:

1. It is not clear what "generally" means in the description "factors generally applied to reflect limitations of the data".

2. "Up to" emphasizes that the reference value applies to any lesser exposure duration. A good addition.

3. Does the sentence "The application of these factors is intended to provide an estimate centered within an order of magnitude" mean that once you divide by an uncertainty factor, the reference value is in the center of a new range for the value? What was the old range? It is difficult to understand how this is being said with any quantitative confidence—is this statement data driven?

4. The definition describes appreciable risk to the human population. What does the panel think about the extent to which RfDs and RfCs should protect hypersensitive individuals? The proportion of the human population that the RfV is intended to protect is left undefined.

Removal of the term noncancer:

RfVs are developed for carcinogens as well as noncarcinogens. The cancer risk associated with a chronic RfV exceeds 1 in 100,000 for the majority of chemicals that the MDH has reviewed. Most individuals and many public health programs will likely consider the cancer risk associated with the chronic RfV exposure to a nonthreshold carcinogen to be an "appreciable risk of adverse effects for a lifetime." Removing the term "noncancer" from the definition results in a false assurance of protection if the chemical is a nonthreshold carcinogen.

This false assurance of protection extends to less than lifetime RfVs as well because cancer studies rarely measure the cancer risk from less than lifetime exposure. The small amount of data that are available suggest that, for at least some carcinogens, a less than lifetime exposure to a particular dose rate early in life can be as potent as a chronic exposure to that dose rate.

One alternative is to describe the cancer risk associated with the RfV exposure. However, since little is known about the cancer risk associated with less than lifetime exposures, this may not be possible or only possible for a chronic RfV. There should at least be a cautionary statement that the RfV for a nonthreshold carcinogen carries a cancer risk—even for less than lifetime RfVs.

Standardized terminology:

While the term RfV is useful for purposes such as this—a collective referral to RfDs and RfCs—the RfD and RfC should be retained. An alternative would be to write, for example, acute oral reference value or acute inhalation reference value. However, the terms RfD and RfC are unambiguous, understandable, and merit separate designations. The most compelling reason to keep them separate is that each has different units, cannot be easily converted from one to another, and cannot be directly compared.

Using subscripts poses additional issues. Subscripts are difficult to type and to read. While RfD and RfC subscripted for exposure duration would be useful in some writing and in tables, the full description (acute RfD, longer-term RfC) is not burdensome to write out.

In summary, this one question for the peer review group brings up a host of assumptions that need to be articulated and, in some cases, reconsidered. It is important to reconsider the meaning of the assurance "without appreciable risk" for an RfV derived for a nonthreshold carcinogen. It is important to consider the interpretation of "including susceptible subgroups" when policy is lacking on the portion of the population (e.g., 99%) that is to be protected. It is also important to reconsider, in light of these imprecise portions of the definition, the false sense of quantitative certainty for the "estimate centered within an order of magnitude."

3. Please comment on the recommendation that endpoint specific reference values should not be derived.

The EPA technical panel has presented a very good rationale to NOT develop endpoint specific reference values. The rationale is consistent with the current viewpoint of a reference value offering protection from all measured effects. The material that EPA has presented is sufficient in explaining this. Creating multiple endpoint specific reference values would add confusion that could lead to misapplication of any particular value. The risk practitioner using these reference values should not have to ferret out and

compare all possible reference values based on all possible endpoints.

4. A review of current guideline study protocols and approaches to testing was conducted to determine what information is currently developed to support the derivation of less-than-lifetime reference values.

Life-stage approach taken in this review:
The review by life-stage is very useful. The organization of the review into the three categories of "gaps" (life stages, endpoints, duration/latency) was particularly useful.

In addition to the study description and data gap analysis found in 3.A., the reader would benefit from a sense of the extent to which testing according to these guidelines is actually carried out (who uses these guidelines and for what purpose) and the extent to which the EPA is developing reference values from studies that were not completed using the guidelines. For example, if a large proportion of RfDs and RfCs have been developed from studies conducted by academicians (for example) rather than registrants and contract labs, a recommendation might be to conduct outreach and education to toxicologists working in those areas in which guideline studies are not widely used. In order to get a better feeling for the importance of this review and these recommendations, it would be useful to see an analysis of the IRIS database (particularly those chemicals that are not registered pesticides). A temporal analysis of the source of the critical studies used to develop reference values and whether or not guidelines were followed in those studies would help determine where future efforts would be most productive.

Comments on data gaps:

a) It is not clear how some of the recommendations would be implemented. While it is laudable that the EPA would encourage collecting more data across life stages, route, duration, and timing of exposure, what are the incentives to produce these data? To date, it appears that the pesticide registration requirements are the only program for which there are incentives to improve testing.

b) New study protocols to evaluate children's health issues are essential. One of the most troublesome data gaps is the lack of information on neonates and young exposed prior to weaning.

c) Of all of the data gaps, the need to develop information on latency appears most important in developing lifetime reference values. This is the data set most critical to developing valid less than lifetime reference values. In addition, the lack of data on aged animals is a great concern.

Chapter 3 included conclusions about pharmacokinetic modeling, which were not based on the text of the review document. The review would be improved if the discussion on pharmacokinetics that led to those conclusions were included in the review. There is no information presented on exactly what data would be most useful in improving the

derivation of reference values, although the most specific concerns appears to be related to developing dose adjustment factors. There is also no explanation of the work that would be necessary to develop these data.

Recommendations for new testing strategies
Stop-exposure studies were not explicitly discussed. The only explicit recommendation for studies incorporating latency was the 14-day post exposure observation period in acute studies. Stop-exposure studies might be particularly important in neurotoxicity, reproductive toxicity, immunotoxicity, and cardiovascular toxicity as, in each case, there is potential for early damage to late maturing systems or early damage that may only be apparent following tissue growth, maturation, or subsequent challenge. Unless stop-exposure steps and latency are incorporated into continuous dosing study design, it will not be possible to determine whether intermittent or early life stage dosing have effects equivalent to continuous dosing (which has large implications for applications). This comment applies to carcinogenicity as much as it does to derivation of reference values.

While the enhanced studies suggested by the EPA technical panel are highly desirable, there are dosing concerns for persistent bioaccumulative toxins (as was briefly pointed out in the review). Even if tissue concentrations were periodically measured, the animals would likely have varying internal doses over the duration of the study. Would reference doses for such chemicals be based on body burden (e.g., as in the recent dioxin review)?

Another dosing concern that is not specifically addressed in these study designs is the dose to the neonate from nursing. This dose has direct application to the concerns that have been expressed about children's exposures. Some mention of how to quantify and model that dose would be useful.

Overall, the proposal for a comprehensive testing strategy is interesting and useful. It would of course be desirable to have all of these data for developing reference values. As mentioned above, the review would be strengthened by some analysis of the current IRIS database. It is not clear how many of the current reference values in IRIS are based on desired but rarely tested endpoints. It is also not at all clear from this review the extent to which such a study design would be required by EPA or, if voluntary, used by researchers.

OCHP proposal for an NAS study:

I must have missed the description of this proposal in the review paper.

5. Please comment on the recommendation to include duration–adjustment for inhalation developmental toxicity as for other health endpoints.

Section D.2. was instructive and of great interest. Minnesota Department of Health staff worked through an example from our files (six hours a day for six gestation days) in order to compare the CxT conversion to our current methodology. In doing so, it becomes obvious that the strategy for duration-adjustment that the EPA technical panel

presents is protective of peak exposures during fetal development. Section D.2. would benefit from an example from a developmental study to make this point explicit and obvious. We assume it would show that the exposure concentration from a developmental study with exposures that last only a few hours of a few days during gestation will be adjusted to a lower exposure concentration equivalent to 24-hour per day exposure. However, without the example using developmental data, it is not clear how the days per week adjustment shown on 4-23 is relevant to a developmental study involving a few days of gestation. We further assume that the example would show that the adjusted concentration is used to derive a single reference level that could be used for all four of the exposure categories. Note that the written description in D.2.a. and D.2.b. did not make clear whether the resulting value for development would be considered an acute, short-term, longer-term, and/or chronic value. But it was apparent from the example (Chem X, table page B-22) that this value would be considered for all four less than lifetime exposure durations.

Section D.2.c. was potentially confusing. The EPA technical panel appears to recommend dose adjustment downward to derive a 24-hour value from discontinuous exposures of 24 hours or less. Careful reading suggests that the EPA technical panel does not recommend dose adjustment upward from the single acute exposure scenario of 24 hours or more to derive a 24 hour value (no adjustment is recommended). Unfortunately, the recommendation of the EPA technical panel (5-5, recommendation 8) to utilize duration adjustment could be interpreted to suggest that, because some studies indicate adverse developmental effects are a function of area under the curve (page 4-24), the duration adjustment can be made from a low exposure concentration to a higher exposure concentration. This impression is reinforced by figure 4-1. EPA might want to clarify that the sentence "an acute reference value may be required for both a 1-hour duration and an 8-hour duration" (page 4-25) is not an endorsement to create such RfVs; the intention is to create 24-hour acute values. It is also not clear whether EPA would recommend upward duration adjustment to develop short-term and longer-term RfVs.

There would seem to be little public health concern to using the duration adjustment to calculate a lower dose when extrapolating from a study with a short exposure period to a longer exposure period. As the review points out, the reference value for the longer exposure period would be a concentration lower than that used in the original developmental study. This may not be a conservative (health protective) assumption if the study exposure period did not extend through all periods of developmental susceptibility, but that is an area of uncertainty and should not affect these calculations.

6. Please comment on the recommendations in the report for using a weight of evidence approach for hazard characterization (Chapter 4, section C.2.), and for expanding characterization of the extent of the database and using a narrative to describe the database rather than a confidence ranking. Is the discussion of weight of evidence clear in terms of how it would be used in characterizing the database? Also, please comment on the consideration of this information in the derivation of reference values.

The confidence ratings have not been particularly useful to the MDH. MDH staff recently considered our own confidence rating for the purposes of rule making and rejected it, believing that all RfVs would meet a minimal confidence level. In fact, we have a second tier list of provisional RfVs for those data-poor substances for which some risk manager required a number (similar to the concept of having HEAST values in addition to the IRIS values).

The weight-of-evidence approach is completely defensible (in fact, hopefully describes current practice in reviewing data for developing reference values). Determining the strength of the data set, critical endpoints, key studies, etc., has always been subject to judgment and experience. A recurring question (and the subject of discussion on page 4-9 of the review) is "What is adverse?" particularly with endpoints related to morphological or biochemical changes that have no known functional impacts. This review and its recommendations will not solve that fundamental problem. However, by making those judgments explicit in what promises to be lengthy narratives in IRIS files, at least the decisions will be available for discussion. This disclosure will be valuable and such decisions will be shared and used by others.

One concern is that it is already apparent that there is a lack of consistency now in the IRIS files with some authors producing very detailed descriptions of studies and extensive comment on the interpretation and selection of studies (e.g., the newest files such as Boron). The weight of evidence information may need to be standardized so that all authors are required to provide the same type of information, when it is available, and comment when it is not available.

7. Please comment on the recommendation for use of an exposure-response array and derivation of sample reference values to select the point of departure (POD) for the final reference value. Also, please comment on consideration of the nature of the effects for different endpoints, durations, timing, and routes of exposure in selection of the POD.

It is very useful to know all of the potential health effects associated with doses in the range of the RfV. One application for knowing about effects that occur at higher levels is when programs consider concurrent exposures to multiple chemicals with potentially additive effects, or when hazard indices are used to consider concurrent exposures to multiple chemicals. The tables for Chemical X were very useful (and more useful for this purpose than the graphic array).

It is not clear what the user will do with the information that there is a cancer effect level (CEL) associated with a chemical and no recommendation has been made.

Table B-4 was excellent and should be used with Table B-1 as Table B-4 alone does not portray the number of distinct studies that were used to develop the RfV options.

It is very difficult to understand how someone preparing an IRIS file could portray

information on steepness of the dose response curve—and interesting that it comes up often as an important consideration. However, the dose interval data shown in Table B-1 may be helpful.

8. Please comment on several recommendations concerning the application of uncertainty/variability factors. Are there additional data or analyses in the literature not cited here that can be used to strengthen the recommendations? Should other factors be considered in the application of uncertainty/variability factors?

Cap the Ufs at 3,000
Yes, do not create RfVs with UFs over 3,000, with some allowance for judgment. However, the total UF value at which to draw the line seems somewhat arbitrary. Why not uncertainty in four areas? Or uncertainty in certain combinations of areas? Whenever EPA staff determine that it is not possible to develop a reference level of some type (RfC but not RfD, for example), staff should describe in an IRIS file why the decision was made and summarize the data on which the decision was based.

The wording of the recommendation is ambiguous. The wording does not say, "The Technical Panel recommends that reference values be developed only when the UFs applied to a chronic reference value total 3,000 or less." Instead, the recommendation can be interpreted to mean that if one really wants that reference value to be developed, propose a total UF of 3,000 or less. There are opinions within EPA that there is sufficient conservatism in each separate UF so that a combined UF of 3,000 offers sufficient health protection even when there is full uncertainty in four or more areas. However, it is not clear whether or not this is a recommended practice and the proposed recommendation does not make this any clearer. The review should describe the current practice for determining the total uncertainty factor when there is uncertainty in four areas (for both cases, full or partial) and the EPA technical panel should comment on the opinion that a UF of 3,000 is sufficient for four areas of uncertainty that would otherwise be calculated as 10,000.

There was significant discussion (page 4-33) of whether or not $BW^{3/4}$ was a dosimetric adjustment. It is implied that this adjustment is only currently used to develop cancer potency slopes. Are there reference values on IRIS that were derived using a human equivalent adjustment of dose? Is there any impediment to applying the values listed in Table 4-3 and making a corresponding adjustment to the interspecies UF?

Reducing the UF for intraspecies when data support the reduction.

Concur, but, as suggested by the panel regarding the interspecies UF, maintain the flexibility for applying a factor greater than 10. Firmly acknowledging that a full 10-fold UF will be applied unless proven overprotective is compatible with the FQPA approach and concern. For this reason, it may be controversial and appear conservative. However, so little has been accomplished to determine where sensitive subpopulations (children,

elderly, compromised) fit into the range of human variability that it is defensible. The supportive arguments put forth by the panel are well crafted and defensible. Use of the full ten-fold factor is nothing startling.

The question of who should be protected will continue to come up. The issue of protecting hypersensitive individuals is particularly interesting. The EPA technical panel fails to address goals such as protecting the most sensitive portions of the population (hypersensitive?) or protecting most of the population (99.9 percent?, 95 percent?).

The question of whether or not the most sensitive portion of the population is included and reflected in the dose-response curve is also interesting. The panel's recommendation of further study should include (among other issues) an analysis of the relationship between variability within a study population and the point of departure response level (i.e., 1, 5, or 10 percent response), the power of the study to detect response, and the identification of the most susceptible portion of the population.

UF should depend on data (quality, quantity) and scientific judgment. Default recommended:
Concur. Of course.

The FQPA factor is not necessary:

Concur. MDH staff have maintained that the portion of the FQPA that could relate to concerns about toxicity is more appropriately considered as intraspecies and data deficiency UFs, per the panel's cautionary statement "if appropriately applied." Further questions should concern whether or not existing guidelines are appropriately applied. As mentioned above, EPA risk assessors should have the option of using uncertainty factors larger than 10 if needed.

The MF is not necessary:

Concur.

CSAFs should be developed with caution:

Concur.

Errata and other specific observations:

Page 3-7, A.1.a. Overview of tests has a sentence fragment inserted into the first paragraph. Delete "young adult animals with a 14-day post-exposure observation period. Other than mortality, the"

Page 3-8 same section. Sentence beginning "other than mortalty" may be incorrect.

Same page, same section. Sentence in next paragraph beginning "the second, expanded study…" may be incorrect.

Executive summary page xvi, top of page. "…the reference values should be derived for lifetime protection of all types of effects for that duration of exposure."

Page 4-24, D.2.b. Dosimeter is used in this portion of the report to refer to CxT adjustments, where is other portions of the report, dosimetric is used to refer to pharmacokinetic adjustments. The duration adjustment is also used almost interchangeably with dosimetric in reference to the CxT adjustments. Perhaps the word dosimetric (dose metric?) (or dosimetric adjustment or duration adjustment) should be added to the glossary.

Closing

This review is very welcome. The panel is congratulated for carrying out this work. This is a long overdue, comprehensive, and in-depth review of a difficult subject. The development of reference values has been an evolving process; something that is obvious to anyone who has used the IRIS database over a period of years. It is a difficult task to make sense of this evolution. The review reflects a comprehensive review of EPA's current and past practices, and a great deal of careful work of the EPA technical panel in critically evaluating these practices. The panel is commended for taking a frank look at the shortcomings and gaps in the body of work around reference values. The changes suggested by the panel reflect a great deal of work and consensus building around innovations in risk assessment.

The review has dealt with almost every issue that has been a concern in the Minnesota Department of Health's (MDH) recent rulemaking for air and water risk levels (air and water concentrations that are based on reference levels and cancer potency slopes). The issues concerning reference levels that have been the greatest concern to us have been the application of the FQPA factor, less than lifetime exposures to developmental/reproductive toxins, total uncertainty factors, and data base sufficiency for various endpoints, particularly those associated with children's health. Our concerns arose in part due to MDH rulemaking activities over the past few years that drew sharp contrasts between the development of RfCs and RfDs. In addition, staff have been working for years on multiple issues related to children's environmental health and attempting to reconcile various viewpoints and actions of EPA offices. Work in the area of children's environmental health increased when a state statute was passed in 2001 that required the MDH to ensure that rules for air and water toxics were protective of children's health. While one rule has been completed, without the benefit of the material in this review, another has begun and we will have many opportunities to work with the ideas presented in the review.

The MDH found that panel members, and other EPA staff, were valuable resources to us

over the past few years that we have been working on the same issues that appear in the review. We are pleased that the product of the panel is now a resource to all states.

Review by
Ellen Silbergeld
Johns Hopkins University

COMMENTS OF ELLEN K SILBERGELD

EPA Draft Review of the Reference Dose and Reference Concentration Processes
[revised after expert panel meeting, 6/19/02]

Overall, I remain concerned by two basic issues raised by this draft review, but not fully considered by the authors: (1) the substantial, and often impressive, changes that encourage flexibility and further elucidation of RfDs (in terms of different exposure durations and multiple endpoints) are made without recognition of the extraordinary paucity of data currently available even to support much simpler approaches to RfDs; (2) the commitment to reduce uncertainty factors in number and impact results in renaming concepts as, for instance, duration "adjustments" when these are really uncertainty factors; conversely, some issues are called uncertainty factors when they are really factors related to variability, which may be well characterized but nonetheless not suitably represented by point estimates. The end result of these changes, I fear, will be to generate more and more numbers with less and less actual science behind them, in order to convey the appearance of more complex and multidimensional evaluations.

a. Response to Charge Questions

1. Derivation of less-than-lifetime reference values in addition to chronic RfD and RfC values
   There is a clear need for guidance to the agency for evaluating exposures that do not persist for the lifetime. Particularly in the context of children's health concerns, it can no longer be assumed a priori that assuming lifetime exposures is the most protective default for exposure assessment in the risk assessment process. Limited exposures during critical periods in the lifespan may incur greater risks. However, the goal of harmonizing or standardizing different approaches to this issue fails to provide the proper context for these methods, such as HAs or AEGLs. Because of different contexts and purposes, it may not be advisable to push for total harmonization. Second, this discussion in the review fails to acknowledge that one of the most important driving forces for deriving less than lifetime RfDs relates to developmental toxicity. In its zeal to eschew endpoint specific activities, the reviewers have at times obscured the context in which their proposed revisions can best be evaluated.

2. Definitions for reference values
   A general definition of terms, such as acute, short-term, and longer-term (<chronic) is appropriate, although considerable flexibility should be encouraged when evaluating specifically timed exposures (e.g., during gestation). I have some problems with the definitions as stated on p 4 - what does "risk of adverse effects for a lifetime" mean? Is this intended to cover delayed or latent effects, that may appear long

after the less than chronic exposure ceases, or does this mean exposures that may each last less than a lifetime but recurring over a lifetime? Also, what is the rationale for selecting "up to approximately 7 years" for the definition of longer-term?

i. indicating that the RfD is "centered" within an order of magnitude is still less clear than could be desired, and in fact may be misleading. RfDs, if calculated with UFs to deal with uncertainties, probably represent the lowest estimate of an exposure that is without appreciable risk of adverse effect, rather than a "centered estimate". How could there be a lower number? These are not statistical inferences, in which a confidence interval might be meaningfully computed.

ii. substituting the word "adverse" for "deleterious" is more in line with current terminology in toxicology and public health; however, it does not resolve much in terms of the ambiguities in meaning for policymakers.

iii. avoiding the specifying term "noncancer" does not seem to provide any clarification, particularly if the goal is to move towards mode of action evaluations. It is not really possible to "harmonize" risk assessment approaches for cancer and noncancer effects, using current EPA methods, so it may be better to retain the separation. I agree that at least equal public health significance should be assigned to either type of effect.

3. <u>No longer derive or develop endpoint specific RfDs</u>

I strongly disagree with this recommendation. For many reasons, as will be discussed below, there are reasons to identify DT-specific RfDs, or issues that relate to DT concerns. Moreover, the endpoint upon which the final RfD is based must be specified, even in a POD type approach as discussed within the report. It should also be made clear that the RfD may not have been based on the most sensitive effect, if data are not sufficient upon which to derive an RfD. Transparency on this point can help stimulate further research as well as preserving the appropriate sense of caution. Switching to a nonspecified RfD is potentially confusing to the public and does not reflect the scientific basis for an RfD. It is still important to focus policymakers' attention on DT and certain carcinogens, especially to draw attention to those chemicals for which there are no data on these endpoints. It may be possible to craft a linguistically satisfying means of accomplishing these objectives, but I propose the notion of a RfD+DT, to denote an RfD in which DT data were available [even if they did not form the basis for calculating the RfD] and RfD-DT, to note those RfDs in which no DT data were available. This serves to signal to the public that an important element in hazard identification, relevant for considering

less than lifetime exposures, is in fact missing.

Moreover, this recommendation completely obscures the purpose of the FQPA. How does one develop non endpoint specific RfDs? Every toxicology test has a set of specified endpoints, and the LOEL/NOEL derived from well conducted tests can only be interpreted in the context of the endpoints measured. These two recommendations make no sense. By the way, this conclusion makes no sense in the context of the next question, where the group recommends research to develop guidelines to assess specific endpoints.

4. Use of life stage approach and alternatives to support less than lifetime and chronic RfDs
   It is possible that some existing test guidelines can be modified, as suggested in the text, to increase the amount of information available at interim points in the study design to provide data on life stage specific events. However, it must be kept in mind that such modifications will greatly increase the size and complexity of such tests. Having experienced the almost exponential effect of attempting such modifications (see Biegel et al (1997) Tox Sci), I caution against the assumption that this will be simple or even desirable.

   It is wholly unclear to me how information on pharmacokinetics will answer these questions, since *life stage specific events have more to do with mode of action and the physiological processes occurring during specific life stages, rather than age-dependent differences in absorption, metabolism, or distribution*. Throughout this document there appears to be a failure to draw the connection between the questions in urgent need of scientifically based answers, and the opportunities to accomplish this through research focused on understanding mode of action.

   Understanding mode of action requires much more than simply adding interim sacrifices to study designs or generating data for PBPK modeling. In considering these innovations, the EPA might ask the NAS to advise in terms of strategies to encourage submission of other than routine tox testing data, especially related to understanding mode and/or mechanism of action.

5. Duration adjustment for inhalation developmental toxicity and other endpoints
   I do not recommend accepting this as a general rule for inhalation or other routes of exposure. Experience with many developmental and reproductive toxicants indicates that timing of dose, rather than

simple adjustment, is critical. Intermittent exposures are different from continuous. As noted below, I recommend that this issue be kept as an uncertainty factor, rather than pretending that a generic approach (adjustment) is other than a gloss over the lack of chemical- or endpoint-specific data.

6. Weight of evidence approach
   I recommend general approach, particularly to accomplish the goal of transparency [once again, this is best shown n endpoint specific tables, as evidenced in the example for Chemical X]. Of course, it is possible to provide confidence rankings and to assign values for weight of evidence, using Bayesian approaches. See the analysis done by Neutra et al for the California EPA assessment of RF risks. It would be interesting to inquire as to the public reaction to this approach - has it helped in understanding how to interpret RfDs?

7. Exposure-response arrays for derivation for the POD
   The exposure array (otherwise known as a dose/response table for multiple endpoints) is an advance on clarity; it must be arranged in a fashion that clearly reveals both data gaps as well as actual data. A consistent template is recommended, rather than just for "relevant endpoints" [how defined?] so that the reader can readily identify these gaps and datapoints. From the example provided, for chemical X, it is not exactly clear how the presentation of a "full array" relates to the calculated RfD; the table does not help the user understand how one dataset was selected for this purpose. And if the RfD is based on only one dataset, then what is the purpose of presenting the (unused) datasets and endpoints? As a member of the "December Group" of dioxin experts, I am reminded of my frustrating experience of drawing together a comprehensive presentation, and discussion, of all the noncancer effects of dioxin, only to have this information completely ignored in the risk assessment/characterization process. Another thing for the NAS to cogitate upon: how can toxicology do meta-analysis, so that more than one study/dataset can be used?

8. Uncertainty factors
   Greater clarity in defining and presenting UFs is to be encouraged. However, the recommendations in this document have the effect of hiding uncertainties in the guise of "adjustment factors". These are really defaults and as such are highly uncertain. That is, as described, the same metric will be utilized for all chemicals/endpoints, which is my definition of an uncertainty factor. I would like to see complete clarity in revealing when adjustment factors have been used for exposure assessment or for inferring other than chronic endpoints.

I also recommend the cautious approach in setting RfD values when 4 or more areas are uncertain. An upper bound of 3000 is arbitrary and does not serve the same useful guidepost for indicating that the amount of uncertainty - e.g., datagaps - may preclude setting a meaningful RfD. It may be appropriate to re-open the issue of how to incorporate multiple UFs in the calculation of an RfD. Should it always be multiplication - what if the sources of uncertainty overlap?

The further point, concerning the 10X UF, seems to be expressed backwards. The intent of this UF was to cover lack of data; however, in the statement on p 11, it seems that data must be advanced in order to support any UF to cover children's health risks. As written this recommendation provides a perverse incentive not to gain real information on susceptible populations, including but not necessarily limited to children. Obviously, sound scientific judgment should be exercised in applying UFs; so what else is new? Remember, however, UFs are meant to cover datagaps, not data.

I do not accept the general assumption that "current traditional UFs will be adequate in most cases to cover concerns about children's health risks" (p 12). Some examples must be presented to demonstrate this; the references in the review are not to primary data for the most part. If the EPA has undertaken this review elsewhere, it should at least be summarized here. Inter- and intraspecies extrapolations, LOAEL to NOAEL, database gaps, and duration extrapolations are not sufficient to substitute for lack of DT data. Lack of information on potential developmental endpoints is different in kind and must be considered explicitly in terms of children's hazard identification and risk assessment. Please reconsider the literature on DES, lead, methyl mercury, solvents as teratogens, etc. UFs are not limited to pharmacokinetic and pharmacodynamic issues, but include mode of action as well.

CSAF approaches require data richness; of course, when data are available, CSAF methods are preferred. I agree that there is an important need to develop explicit guidance as to how such information can be evaluated and then used; I have heard too many complaints from industry that mode of action information is largely ignored in risk assessment.

Review by
Alan Stern
New Jersey Department of Environmental Protection

1. Please comment on the recommendation to derive less-than-lifetime values in addition to the chronic RfD an RfC> Is the need for these values adequately justified in the report?

While acute and longer-term guidelines (HA's), are routinely derived by the EPA's Office of Water, it can be argued that drinking water is a special case since there are real-world situations where drinking water sources become contaminated, and there is an interim period prior to the availability of a new drinking water source during which the contaminated source may or may not be used. No such rationale was presented for the need for non-chronic RfDs in other contexts. The rationale for acute and non-chronic RfCs is easier to envision (e.g., short term air emissions resulting from temporary construction, fumigation, remediation procedures), however, this case is not made explicitly either. Clear guidance on the appropriate uses for acute and short-term reference values should be provided. A concern with the derivation of acute and non-chronic RfDs, and RfCs is that the dose-response curves for acute and short-term effects are often steeper than for chronic effects. Thus, for such shorter-term effects, the implications of uncertainty in the risk assessment process are more weighty. This point, and the appropriate approaches for dealing with it, should be addressed.

2. Please comment on the revised definition for reference values.

For the most part, the revised definitions (apart from the substantive changes involving the non-chronic reference values) are attempts at clarification. In themselves, these clarifications are not unreasonable. However, these attempts at clarification are being made within a basic definition of the RfD/RfC which is ambiguous and does not lend itself to practical quantitative description. This is most specifically the case for the term "susceptible subgroups," and for the term adverse (formerly "deleterious"). In practice, the inclusion of "susceptible subgroups" has not meant even the most susceptible individuals (e.g., those with extreme sensitivities, those with rare genetic diseases and/or polymorphisms, those whose health is otherwise significantly compromised). Thus some individuals who might reasonably be argued to fall into characterizable subgroups are not necessarily being protected. At one level, there is no clear definition of "susceptible subgroups." At a more fundamental level, however, the issue is less about the definition of such groups, than it is about translating such intentions into a percentile of the population for which exposure at the RfD/RfC is, in fact, truly intended to be without adverse effect. This is done, more or less explicitly, in the derivation of BMDs, where the BMR (e.g., 1%, 5%, 10% etc.) defines the fraction of the test population which is experiencing adverse effects at the BMD. In the absence of such a quantitative definition, attempts to provide quantitative clarity to the RfD/RfC derivation process (including the UF adjustment process) are without context. The change in the terminology from "deleterious" to "adverse" seems innocuous (if unnecessarily hair-splitting), but it begs the question of what is intended by either term. This *concept* should be made more explicit, and should delineate adaptive changes from those resulting in reduced quality of life and/or reduced function and/or reduced survival. Additionally, the concept should include the rationale for including some impacts on the quality of life (e.g., skin discoloration, dental mottling) while excluding others (e.g., allergic contact dermatitis).

The elimination of "non-cancer" from the RfD/RfC definition is a fairly crude attempt at policy forcing. This omission now removes any definitional barrier to the application of the RfD/RfC approach to carcinogens. The use of the reference dose approach for the setting of exposure guidelines for carcinogens is , arguably, a reasonable step. However, it is not the only logical approach to harmonization, and not necessarily a universally accepted approach. The opening of such a door should be done explicitly rather than implicitly and surreptitiously.

The terminology "Reference Value" is introduced on page 4-5, but does not appear to be carried through elsewhere in the report. It is not defined and its use and intent are unclear.

3. Please comment on the recommendation that endpoint-specific reference values should not be derived.
This seems to be a semantic, rather than technical, issue. It is clear that developmental toxicity as an endpoint needs to be addressed. It is also clear that developmental toxicity is not, strictly speaking, a chronic endpoint. As long as the RfD/RfC structure is designed to explicitly require consideration of developmental endpoints, and it is clear that the non-chronic nature of such endpoints will not prevent them from taking precedence over RfD/Cs based on higher chronic doses, there is no reason to derive endpoint-specific reference values.

4. ... Please comment on the life-stage approach taken in this review, as well as the recommendations for alternative testing approaches and strategies for developing a targeted testing strategy to support setting less-than-lifetime reference values as well as chronic values. Also, please comment on a proposal from the Office of Children's Health Protection to request a study by the National Academy of Sciences to take a fresh look at toxicity testing approaches and strategies based on this and other reports (e.g., ILSI, 2001).
The life-stage approaches taken in this review will definitely provide a significant amount of useful information not previously available on a regular basis for risk assessment purposes. In particular, the ability to obtain information on effects in ageing animals, information on time to effect in chronic studies, and more integrated testing during pre-adult stages will, if put into practice, eliminate considerable uncertainty currently inherent in RfD/C derivation. However, limitations appear to still exist within this approach. There does not appear to be a provision for obtaining information on sequelae, and/or latent effects uncovered with ageing resulting from dosing during development and pre-adult stages. Dosing at such stages appears to be either continuous through sacrifice, and/or sacrifice in some protocols (e.g., developmental) occurs before ageing effects might reasonably be observed. The timing of developmental events in rodents some of which occur post-nataly, but whose corresponding stages in humans occur *in utero*, continues to be an area of uncertainty particularly with respect to cross-placental transport, and maternal-fetal interactions. In addition, there is a strong emphasis on the generation of pharmacokinetic (PK) data. Such data are often quite useful, however, PK data should not be seen as end in themselves. As such the generation of PK data should not necessarily be seen as critical to the derivation of

RfD/Cs except perhaps where lactational transfer is hypothesized to play a key role in exposure and effect.. In most cases, proper adjustment of the dose at the target organ will alter the estimated human NOAEL or BMD by less than a factor of 10, and often by a factor of 2-3 or less. The overall goal of the lifestage approach would appear to be mostly a qualitative refinement aimed at ensuring that no significant effect occurring at a particular lifestage or resulting from exposure at a particular lifestage is overlooked Thus, while the magnitude of differences in the ultimate RfD/C which are likely to result from the routine inclusion of PK assessment are worth pursuing, the absence of such data should not (given the inherent uncertainty in the ultimate determination) greatly diminish the integrity of the overall testing approach.

The expanded study designs (i.e., the expanded chronic/cancer study design, and the unified screening study design) will provide additional useful information, and (particularly the unified screening study design) are logical and elegant. However, given the considerable additional effort and costs involved, the application of such study designs may have an effect opposite to that intended. That is, in the interest of economy, and efficient use of scarce resources, testing of individual chemicals may be avoided except in cases where there is already suggestive evidence that particular effects exist. Thus, chemicals for which no *a priori* information of particular lifestage effects exist would not be tested under such a protocol and unexpected findings would be missed.

In the section on immunotoxicity testing, there appears to be some confusion between the usefulness and predictive value of hypersensitization challenge studies and infection challenge studies. These, in fact, reflect different, and in some ways opposite, functions of the immune system – the former reflecting the development of a generally unnecessary and deleterious response, and the latter reflecting the competency for a necessary beneficial response. It is not clear what goal is being sought here.

5. Please comment on the recommendation to include duration adjustment for inhalation developmental toxicity as for other health endpoints.

Whether peak exposure concentration or AUC-based concentration is more predicative of developmental effects, is likely to be more predicative of developmental effects will likely depend on the particular chemical and the particular mechanism of developmental toxicity. Those substances which interfere with a specific developmental process during a limited window of vulnerability would be most likely to have their effects predicted by peak exposure concentration, while those having more global effects through (e.g., effecting large scale metabolic processes) would more likely have their effects predicted by AUC. In the absence of specific-mechanistic information, and given the inherent protectiveness of using time-weighted rather than peak concentration, the recommendation to include duration adjustments for inhalation developmental toxicity is reasonable.

6. Please comment on the recommendation in the report for using a weight-of-evidence approach for hazard characterization and for expanding characterization of the extent of the database and using a narrative to describe the database rather than a confidence ranking. Is the discussion of weight of the evidence clear in terms of how it would be used in characterizing the database? Also, please comment on the consideration of this information in the derivation of reference values.

The use of a weight-of –evidence approach for hazard characterization is a sound approach. However, it should be noted that, conceptually, such an approach has generally been used in practice in the RfD/C derivation process. That is, the usefulness, and appropriateness of the available data has been evaluated to determine whether they will support a risk-based guideline. In cases where the data are unreliable or sparse, RfD/Cs have generally not been derived. Thus the approach proposed in the report is not a conceptual departure from the current practice. Rather it's usefulness lies in its explicitly providing a formal template against which studies can be evaluated . It important that the guidance in the report emphasizes that the specifics of the weight-of-evidence approach not be used as a checklist., but that significant professional judgment is required. Thus, strong dose-response data in animal models for highly significant effects might reasonably be considered sufficient even in the absence of other information suggesting significance for humans. With the possible exception of clear evidence from pharmacokinetics that a particular target organ effect is *not* applicable to humans, it is likely that factors other than dose-response will be used to resolve the appropriateness of the database for substances with unclear hazard identification.

While the shift toward description of databases using narratives is appropriate, the recommendation that the database be characterized using only two categories "minimal" and "robust" would appear to provide little benefit over the current high, medium. low approach. The use of a "medium" or intermediate category, is useful for indicating that additional information could make a difference in the overall assessment, but that the current assessment is (in the absence of significant new data) supportable without a critical need for additional data gathering. In practice, the characterization of the overall strength of the database is mostly used to gauge with what level of certainty or uncertainty to interpret the potential for public health impact given exceedance of the RfD/C. Collapsing the categories provides less guidance for that purpose. In addition, although it is useful to provide an overall characterization for the database (i.e., for the RfD/C), the suggestion (pg. 4-18) that such an overall characterization should preclude separate descriptions of individual endpoints would seem to decrease useful information in at least some cases. Thus, if a database is relatively robust for e.g., developmental effects, but the RfD is based on a more general endpoint e.g., weight loss, or organ weight changes which yields a slightly lower RfD whose certainty is characterized as minimal, it would be necessary to choose an overall characterization as either minimal or robust. In truth, neither would adequately describe both endpoints. When specific information is available for a given endpoint, which is relevant at or near the POD, it should be described independently. The critical endpoint should also be described independently. This does not preclude providing an overall characterization for the RfD/C, and in fact, such a characterization would be helpful, particularly in those cases

where the critical effect is not the most robust. The factors for evaluation of the weight-of-evidence provide a good basis for categorizing the overall database. In the end, such considerations are mostly useful for communicating information about confidence. Such information would tend not to be have great practical import in the derivation of reference values, because such values will (and should) continue to be derived on the basis of the adverse effect which meets minimal criteria for supporting a reference value.

7. Please comment on the recommendation for the use of an exposure-response array and derivation of sample reference values to select the point of departure (POD) for the final reference value. Also, please comment on consideration of the nature of the effect for different endpoints, duration , timing and routes of exposure in selection of the POD.

The use of exposure arrays in the presentation of the overall picture of the database for a given reference value is a good idea. It not only provides background for those not well acquainted with risk assessment, but also provides useful perspective for practitioners. Such approach is not all that different from the summary information provided by ATSDR in it Toxicological Profiles. Where feasible, however, it would also be useful to show not only the sample reference value for each endpoint but also the actual dose-response curves. This would provide useful information on steepness of the response for each endpoint, and would thus provide useful information on the implications of uncertainty in POD selection both within and across endpoints. Given the possibility of different endpoints for different routes of exposure, duration of exposure etc., the ability to compare different endpoints and their relationship to the POD is quite useful. However, caution needs to be exercised so as not to engender apples-and-oranges comparisons.

8. Please comment on several recommendations concerning the application of uncertainty factors. Are there additional data or analyses in the literature not cited here that can be used to strengthen the recommendations? Should other factors be considered in the application of uncertainty/variability factors?

Issue A - There seems to be some confusion in this recommendation between the total number of UFs which constitute unacceptable uncertainty for RfD/C derivation, and the total UF product which constitutes unacceptable uncertainty. This is stated both as maximum of four categories of adjustment, and as a total UF product of 3,000. There is clearly nothing objective or absolute about either criterion. Conceivably, there could be a substance for which there UFs are required in five categories, but all or most of these are UFs of 3. In such cases, the resulting RfD/C might not be considered inappropriate. On the other hand a total UF product of 10,000 (i.e., four categories of 10) does seem excessive and likely to overwhelm the information provided by any POD. Such a UF would, indeed seem inappropriate for the derivation of a RfD/C.

Issue B - I concur with the recommendation of the Technical Panel.

Issue C - Although quantitative analyses have generally been limited in both scope and sources of data (i.e., pharmaceuticals), such studies have generally found that for LOAEL-NOAEL, subchronic-chronic, interspecies, and intraspecies, categories a factor of 10 represents an approximate 90th-99th percentile. Thus, the use of a default factor of 10 for each UF category continues to be reasonable. Nonetheless, the recommendation of the Technical Panel that specific consideration be applied in each case is sound.

Issue D - I concur that, if properly applied, the current scheme for application of UFs addresses those cases where children may be considered more sensitive than the general population. Specifically, the UF for intraspecies (inter-individual) sensitivity, should explicitly consider the potentially greater sensitivity of children than the test population. Therefore, there does not appear to be any scientific necessity for a separate UF intended specifically to address the sensitivity of children. The FQPA UF addressing children's sensitivity, is a reflection of concern and intent to be cognizant of the sensitivity of children, which has perhaps been overlooked in the past, rather than a necessary addition to the RfD methodology.

Issue E - The database UF would seem to provide opportunity to address all areas of uncertainty which would otherwise be addressed by the "Modifying Factor." It should be noted, however, that there is an inconsistency on this account in the background material provided on the IRIS website. The background document provided to explain the RfD process (http://www.epa.gov/iris/rfd.htm), does not list a database UF, but does list the modifying factor. The difference between the two appears to be entirely semantic.

Issue F - The analysis and approach of Renwick, which divides interspecies and intraspecies variability into two factors of 3 each for toxicokinetic and toxicodynamic factors appears, on the basis of the current limited data for substances outside the pharmaceutical literature, to be reasonable (see for instance Stern et al., 2002 2). As a default, this approach leads to the current defaults of 10. Chemical-specific adjustment factors (CSAFs), when properly derived on the basis of adequate data would be preferable to the default approach. However, I concur with the recommendation of the Technical Panel that such caution should be exercised in departing from the default approach in order to assure that the basis for that departure is, indeed appropriate

2 Stern, A.H., Clewell, H.J., and Swartout, J. An Objective Uncertainty Factor Adjustment for Methylmercury Pharmacokinetic Variability. *Human Ecol. Risk Assessment 8(4), 2002.*

Review by
Lauren Zeise
California Environmental Protection Agency

Comments on *A Review of the Reference Dose and Reference Concentration Processes*
EPA May 2002 External Review Draft EPA/630/P-02/002A

Lauren Zeise, Office of Environmental Health Hazard Assessment, California Environmental Protection Agency

1: Comment on recommendation to derive less than lifetime reference values

The provision on IRIS of less than lifetime reference values in addition chronic values is justified to the extent that such values can be reliably derived. However, when derived from a limited database consisting of studies that have not sufficiently evaluated exposures during critical life stages and on the suite of endpoints and systems, such values can be quite speculative. Under the proposed definitions, the less than lifetime values are represented as likely to be without an appreciable risk – for all possible health outcomes – in the human population, including susceptible groups. The guidelines should provide detailed criteria to help define databases that are minimal and for which such derivations to this end cannot be made. Also the cap for the combined uncertainty/variability factor should be reconsidered with regard to marginal databases that may be used in such derivations. In some instances a large factor is warranted and can be justified, for example, when large differences in susceptibility exist among people exposed. With regard to the cancer endpoint, there are a relatively large number of chemicals known to produce cancer from acute exposures, indicating importance of addressing the cancer endpoint in establishing less than lifetime reference levels. Methodology for calculating risk from less than lifetime exposures will be needed to adequately address this endpoint. For carcinogens, absent an explicit consideration of risks from short term exposures, short term reference values should not be discussed as protective of the cancer endpoint.

Other issues to be addressed in derivations of less than lifetime reference values include chemical persistence, the recurrence of short term exposures, and concomitant and subsequent exposures to chemicals contributing to the same toxicological process as the chemical under evaluation.

There is already a considerable backlog of chemicals needing the current, standard cancer and non-cancer evaluations. The derivation of less than lifetime levels will add to the workload. It is recommended that EPA develop a process for prioritizing the development of reference values, considering public health import, feasibility and other factors.

2: Revised definitions for reference values

Picking language less easily understood is unlikely to change the likelihood for reference values being treated as bright line levels. If the goal is providing information an estimate that is “centered within an order of magnitude” then methodology needs to be consistent with this goal, and perhaps ranges should be reported. Making terminology more vague is unlikely to help. The move in risk characterization is toward greater transparency and

clarity, not less (EPA Science Policy Handbook: Risk Characterization, EPA 100-B-00-002, Office of Research and Development, December 2000). Words that communicate information in a more straightforward, easily understood fashion to the experts and non-experts alike should be considered. While the meaning of "Reference values" for "acute," "short-term," "longer term," and "chronic" will be understood among toxicologists, these labels may not be readily understood by the risk manager and many other non-toxicologist public servants, and the public. It would be preferable to report "one day," "one month," "5 year," and "lifetime" values, with a clear explanation of what is meant. Also, instead of "reference values," "reference drinking water concentration," "reference air concentration," and "reference dietary dose" should be more easily understood. Thus, IRIS could report the more general reference values, but then convert them to, for example, "one day reference drinking water concentration." The replacement of "deleterious" with the more commonly used "adverse" is a move in the right direction.

It can be quite useful to the public and the risk manager to be aware of reference levels thought to be protective of specific endpoints (e.g., adverse development), particularly where exposures occur above the reference value (see #3 below).

3: Comment on recommendation not to derive endpoint specific reference values

This charge question asks for comment on the recommendation that endpoint-specific reference values should not be derived. This raises two important concerns, first, the need for the more specific information such as this, and second, the verity and reliability of the exercise of deriving values that cover all endpoints.

1) In the spirit of providing a full characterization of the risk and sufficient information for decision-making, detailed information on the degree certain exposures may or may not impact particular endpoints can be important. The risk manager, or for that matter, someone from the public, may behave differently when more specific knowledge is provided. Information on, say, whether an RfD_{DT} is considerably different from the RfD, or a clear statement that there is insufficient information on that endpoint to provide values, can be important in certain circumstances. EPA documents are major source of information on toxicity for a variety of audiences, including local health departments, state governments, and the expert and non-expert public. Documentation of more detailed information on endpoint specific values (that would be presumably calculated and considered in any event as part of the reference value derivation) that can be accessed via the internet (including via links in IRIS) would be a service to the user community.

2) Under the proposed definitions, the reference values are represented as likely to be without an appreciable risk in the human population, including susceptible groups, for all possible health outcomes. When endpoints have not been sufficiently studied, for example, when cancer bioassays have not been run or available data are inadequate for a supportable carcinogenicity evaluation, the assertion that the reference value is likely to be without appreciable risk can be speculative and without foundation. Also, certain endpoint determinations may involve considerable resources to complete, for example,

the carcinogenicity evaluation for a chemical with a complex mechanism of action. Rather than wait for the completion of all aspects of the assessment on a chemical, toxicity information on non-cancer endpoints could be made available to the public on IRIS. Also, clear statements regarding where data gaps exist for endpoints and judgments on the likelihood that reference values may significantly change if filled would be a service to the public.

As the source of the Agency's consensus position regarding chemical exposure levels considered protective of public health, IRIS should provide the critical information on hazards and risk levels for endpoints of concern. This approach is consistent with the Agency's risk characterization guidelines (EPA 100-B-00-002, full citation given above) and the National Research Council (NRC) recommendations on risk characterization (Understanding Risk: Informing Decisions in a Democratic Society, NRC Committee on Risk Characterization, National Academy Press, 1996).

4: Comment on the life stage approach, recommendations for alternative testing, and proposal for NAS to look at related toxicity testing strategies

A Review of the Reference Dose and Reference Concentration Processes provides a good review of current guidelines study protocols and approaches to testing and was quite useful as basis for considering the type of information currently developed to support derivation of less than lifetime reference values. A number of thoughtful suggestions were presented for alternative testing approaches and strategies for developing a targeted testing strategy to support less than lifetime reference values. It is anticipated that the details of the suggestions will receive considerable vetting elsewhere. However, of note here is the lack of explicit consideration of study power, critically important in evaluating study adequacy. It is recommended that the power calculations be addressed in the next version of the *Review* and that these be carefully considered developing study protocols. A study by the National Academy of Sciences on toxicity testing approaches and strategies for addressing toxicity at different life stages would be valuable. Considerations for new strategies include 1) generation of data critical for risk assessment, 2) testing that contributes to development of generic knowledge that can be applied to characterizations for specific chemicals, and 2) establishment of a research environment that provides adequate testing of hypotheses and their alternatives from a variety of research perspectives.

5: Comment on the inclusion of a duration adjustment for inhalation developmental toxicity

The discussion of this issue is incomplete in the absence of discussing how the corresponding environmental exposures are to be calculated for the comparison with the reference value. In this sense, the peer review group was not given sufficient documentation to address this issue adequately. The point is made that use of C × T to correct duration is more health protective but this is unclear. If C × T is being used to adjust for intermittent exposures in the study, it may also be used in assessing environmental exposure. If the life stage timing and magnitude of in the study is not

comparable to the environmental exposure scenario, the C × T correction is not necessarily "health protective." Fifteen minutes of 2400 ppm environmental exposure may lead to considerably more toxicity than 24 hours of 25 ppm exposure in a study, or 6 hours of 100 ppm exposure, yet the assumption is the response will be the same. While a number of agents may be a function of AUC, others appear to be more of peak concentration, and the health impact of making such an assumption in cases where it does not hold can be large. This holds for routes of exposure other than inhalation as well. During the peer review meeting the proposed adjustment was discussed as occurring only during a single day. This is inconsistent with the calculations given in the *Review of the Reference Dose* document.

6: Weight of evidence hazard characterization and narrative description of database rather than confidence ranking

The peer review group was asked to comment on the recommendation in the report for using a weight of evidence approach for hazard characterization. Weight of evidence evaluations are the foundation for hazard characterizations. Several factors were stated that decreased or increased weight regarding the likelihood of effect in humans in general (see e.g., Table 4-1), and for susceptible populations (e.g., Table 4-2). In making weight of evidence calls it is important to remember that lack of evidence due to insufficient study is not the same as negative evidence. If studies have not been conducted, or if studies drop out because for example the mode of action data indicate the species studied is unlikely to be relevant to humans, then there can be a data gap that precludes a determination of endpoint hazard - data are simply not sufficient for making a judgment as to whether the endpoint may occur in humans. It is recommended that for a series of critically important endpoints an approach be adopted that characterizes the database as sufficient or insufficient to judge (as EPA has adopted for developmental and reproductive toxicity endpoints). A clear listing of areas where data are insufficient for making a judgment about important endpoints should accompany reference values presented in the IRIS and other EPA documents. If there are significant inconsistencies in study results that lead the analyst to doubt the findings or data gaps, then this should be reflected in the characterization of uncertainty and the uncertainty factor adopted for the reference value calculation.

Also, important cautionary notes regarding data on similarity of effects, mode of action, and pharmacokinetics are not given in the report. Lack of site concordance across species for effects can be common but quantitatively the measures of potency or activity can be quite similar. A finding of one endpoint in the animal can be predictive of a different toxicity endpoint in humans (e.g., (zymbal gland in rodent, relevant cancer site in humans). The report appears to presume that mode of action and the critical aspects of pharmacokinetics for the endpoint in question are well enough understood to make a weight of evidence call, without warnings as to technical difficulty and guidance elsewhere in EPA documents.

A narrative description of the database to provide an idea of the confidence for producing

a reference value will be an important component of the documentation. Some of the elements recommended include clear statements regarding the sufficiency of the database to support weight of evidence determinations for specific endpoints and statements based on power calculations indicating ability to detect effect, as discussed elsewhere as well in these comments. The use of the generic term "Robust" as a label may mislead. It may be robust for setting reference values on the basis of certain endpoints or time periods but not others. Also, for the non-technical audience may understand the labels regarding confidence accompanying the narratives may aid the understanding. "Robust" may be taken to be a replacement label for "high confidence." If the "robust" label is to be used a clarification narrative should be given when data are insufficient for certain endpoints. Another option would be "Sufficient for setting reference level for protecting against (fill in) endpoints." Please note though that the confidence characterization is useful to some IRIS users in communicating values. The same concerns raised regarding endpoint hold though for the confidence ranking as well.

7: Exposure response array and derivation of sample reference values to select point of departure

The exposure response array provides a visual regarding the range of low and no effect levels among the tested species. However, it is of limited usefulness for considering comparability of findings when values presented are LOAELs and NOAELs and points plotted are not distinguished on the basis of severity and nature of toxicity. It would be preferable to plot the benchmark doses with 95% confidence bands, and to provide some way of distinguishing the severe from the less severe, particularly for related endpoints. The selection of the reference value after an evaluation of the relevant endpoints in terms of the reference values they generate and related uncertainty/adjustment factors seems a reasonable approach. Power calculations would be useful information to present in tabulations of NOELs.

8: Comments on several recommendations regarding uncertainty/variability factors

A: Comment regarding the number of areas of extrapolation and a cap on the total uncertainty/variability factor applied

The limit on the overall uncertainty factor should be considered in light of knowledge about the contributing factors. Some factors represent adjustments rather than uncertainty. First, the intraspecies factor represents variability rather than uncertainty, and is more properly named a variability factor. This factor will be employed when the degree of variability is adequately characterized as well as when it is not. This should be considered in any characterization of data base sufficiency. Note that values of one for the intraspecies factor do not reflect certainty but rather lack of realism. There may be knowledge of extensive human variability and a factor considerably larger than the typical factor of 10 may be derived from data. When there is knowledge about intraspecies variability, the factor should be removed from the total uncertainty factor cap calculation, and factored in later. The total uncertainty factor cap of 3000 applies to inhalation and oral equally. The inhalation adopts a factor of 3 for interspecies

extrapolation to address pharmacodynamic differences since levels are expressed as human equivalent concentration. The factor of 10 for oral can in certain cases (e.g., parent compound active) just represent a pharmacokinetic adjustment to the dose. In such cases it is not an uncertainty factor, but rather a dose adjustment. When this is known, it should be treated as a dose adjustment to correct dose to animals to the human equivalent dose and not factored in to the total uncertainty factor calculation, but later, or earlier, in the process.

B: Reduction of the intraspecies factor

I agree that the reduction of the intraspecies factor should only be considered if data are sufficient to support the conclusion that the data set on which the POD is based is representative of susceptible populations. Because there can be multiple such populations with the basis for the susceptibility differing among them, care must be taken to ensure that those most susceptible are taken into account.

Regarding the intraspecies factor applied to inhalation data, it is important to recognize that the factor is being applied to a concentration rather than dose and that there is variability in breathing rate, due for example to differences in physical activity and metabolism/age. In selection of the intraspecies factor this additional source of variability should be considered, particularly for acute exposures. Use of a default greater than 10 for the intraspecies factor should be considered, especially for inhalation exposures. Another issue in the evaluation of the magnitude of this factor has to do with the extent there are exposures to chemicals in the background that contribute to the same toxicological process as the chemical under evaluation. Individuals that are farther up on the dose response curve (due to background exposures) may be more susceptible, even where there are no inherent differences in susceptibility at the same exposure level.

C: Choice of uncertainty/variability factor dependent on study quality, database available, and scientific judgment

Agree. It is important to note that the independent derivation of oral and inhalation reference values can lead to an illogical set of reference values as well. Consider a systemic effect evaluated by the two routes with studies of differing sensitivity, even due to dose spacing. Good pharmacokinetic data may indicate that reference values for the two routes in terms of mg/kg dose should differ by say a factor of two. The differing study quality and data availability may result in a considerably greater difference. In such case, it may be more scientifically sound to derive reference values for both routes from the most sensitive and reliable study and apply a pharmacokinetic adjustment to estimate dose for the second route.

D: Factor to protect the young

While conceptually the factors noted an cover uncertainties about children's health risks, it is unclear whether in practice there will be adequate coverage. Clear guidance to ensure that the young are adequately is needed. To address this issue directly it seems prudent to increase the default intraspecies factor, and provide for its reduction for where data are sufficient to justify. For reference concentrations derived on a concentration basis differences in breathing rate alone may result in a factor contribute a factor of three greater sensitivity to the infant for certain chemicals (median adult breathing rate versus median rate for neonate).

E: Comment on the modifying factor

One can envision situations where it would be useful to have a modifying factor. For example, if there is good reason to believe on the basis of structure activity, mechanistic data and other information that a chemical is likely to be a low dose linear carcinogen, but are insufficient to treat the chemical as a carcinogen. In this case it may be prudent and appropriate to set a value lower than would be established using standard methodology.

F: Caution should be used in development of chemical specific adjustment factors

Agree. With regard to the division of the UFs into pharmacokinetic and pharmacodynamic, it appears that the overall factor for interspecies adjustment can be too small then when applied to oral exposures. It leaves a factor of 3 to cover pharmacokinetic differences, when size and metabolic differences indicate the factor should be greater. With regard to developing human equivalent doses and concentrations, the methodology appears to assume these calculations are exact and not uncertain.

Other comments

1. It is recommended that a power calculation accompany any dataset being considered for the establishment of a reference value, or considered important in weight of evidence determinations. For studies of low power, such as some of the subchronic dog studies, findings may not be statistically significant but carry biological significance, particularly in light of other relevant data. This should be given careful consideration in the reference value setting exercise.

2. Although recognized in the *Review* it does not appear sufficiently appreciated in the Chapter 4 that the possibility that a non-cancer endpoint has a low dose linear dose response relationship should be given careful consideration. An important issue in this regard is the extent that exposures to the chemical at hand and others operating via the same mechanism are present at background levels.

3. A group of scientists and statisticians has been assembled to address the eight

charge questions. A number of questions address the type of information that will be provided to by the Agency, in documents and on IRIS, a database widely accessed by both the expert and non-expert public in addition to the Agency. Several of the charge questions fully or in part deal with policy and other issues that have a strong non-science component. If it has not done so already, the Agency should consider vetting proposed changes to expert and non-expert users of IRIS and EPA documents and specialists in risk communication, perception and public policy.

4. The terminology "non-linear" for threshold or threshold-like dose response relationships and "linear" for non-threshold-like, continues to be used. Previously, the EPA Science Advisory Board has commented that these terms are misnomers. Most dose response relationships fit to data used in risk assessment are non-linear, though in the low dose range they are linear. Further there are a various cases where the dose response relationship is highly upward curving but the mode of action and pharmacokinetics support the assumption of low dose linearity. It would lead to less confusion to adopt a term different from "non-linear," such as "threshold-like." For a discussion see Science Advisory Board comments on the various drafts of the EPA Carcinogen Guidelines.