Report of the Peer Consultation Meeting on the Scientific Rationale for Deriving Database and Toxicodynamic Uncertainty Factors for Reproductive or Developmental Toxicants

September 19, 2005
METS Center
Northern Kentucky University

Peer Consultation Organized by
Toxicology Excellence for Risk Assessment
(http://www.tera.org/peer/)
Cincinnati, Ohio

Final Meeting Report
December 2, 2005

Table of Contents

Executive Summary 3

1. Participants 6

2. Background 7

3. Introductions, Conflict of Interest, and Meeting Process 8

4. Author Presentation of Database Uncertainty Factor (UF_D) Data Compilation and Support 8

5. Panel Discussion of Database Uncertainty Factor (UF_D) Data Compilation and Support 9

5.1 Clarifying Questions 9
5.2 Public Comments 9
5.3 Panel Discussion of Charge Questions 10
5.3.1 Question 1. ..... *10*
5.3.2 Question 2. ..... *13*
5.3.3 Question 3. ..... *15*
5.3.4 Question 4. ..... *15*
5.3.5. Questions 5 and 6. ..... *16*
5.4 Additional Public Comments and Questions 17

6. Author Presentation of Comparing Animal and Human Toxicodynamics for Reproductive and Developmental Endpoints 17

7. Panel Discussion of Comparing Animal and Human Toxicodynamics for Reproductive and Developmental Endpoints 17

7.1 Clarifying Questions 17
7.2 Public Comments and Panel Responses 18
7.3 Panel Discussion of Charge Questions 19
7.3.1 Question 7. ..... *19*
7.3.2 Question 8. ..... *22*
7.3.3 Question 9. ..... *22*
7.3.3 Question 10. ..... *22*
7.3.4 Question 11. ..... *22*
7.3.5 Question 12. ..... *23*

8. Summary of Conclusions, Recommendations, and Suggestions 23

8.1 Database Uncertainty Factor 23
8.2 CSAF for Reproductive/Developmental Toxicity 24

9. References 25

Appendix A – List of Attendees A-1
Appendix B – Meeting Materials B-1
Appendix C – Sponsor Presentation Slides C-1

Executive Summary

An expert panel was assembled on September 19, 2005, to conduct a scientific peer consultation of a document titled "The Scientific Rationale for Deriving Database and Toxicodynamic Uncertainty Factors for Reproductive or Developmental Toxicants." The document and background information discussed at this peer consultation were prepared by scientists from *TERA*. The panel held in-depth discussions regarding key aspects of the document as highlighted in the charge to the panel (found in Appendix B of this meeting report). The deliberations of the panel were supplemented by input from the document authors and meeting observers.

Two separate compilations were provided in the document that formed the basis for the peer consultation. The first portion of the document addressed the approach for considering data gaps in the development of Reference Doses (RfDs), particularly the database uncertainty factor (UF_D). The document addressed the question of whether the current default UF_D values are appropriate in light of current knowledge, or whether modifications are needed. This issue was evaluated through the distribution of the ratios of NOAELs from five key study types, using data from the U.S. EPA's Integrated Risk Information System (IRIS) and from the Agency for Toxic Substances and Disease Registry (ATSDR). The second portion of the document addressed the development of chemical-specific adjustment factors (CSAFs) for the toxicodynamic component of the interspecies uncertainty factor for reproductive and developmental endpoints. Specifically, the document presented information that may allow the development of supplemental guidance to the IPCS (2001) CSAF methods for evaluating toxicodynamic uncertainty for reproductive and developmental endpoints based on consideration of feasibility as well as the comparability of the endpoints in humans and experimental animals.

The discussion of each portion of the document began with the document authors making a short presentation highlighting salient points. Following the authors' presentation, panel members asked clarifying questions and an observer presented a short statement to the panel. The panel then addressed the charge questions provided to them prior to the meeting.

Overall, the panel felt that both portions of the document reflected a substantial amount of work. The section on UF_D, in particular, was noted as being a good start at addressing an important issue.

A number of suggestions were made for improving the presentation and analyses in the section on UF_D. Panel members suggested that the introduction frame the questions being addressed, including limitations of the analysis and the degree to which the data sets used are representative of the universe of chemicals. They also emphasized that, in practice, development of a UF_D considers the entire database on the chemical and also information on related chemicals, not just the five specific study designs that were the focus of the analysis. They recommended that the document explicitly note this difference between the analysis and practical application, and that consideration be given to broadening the analysis to include other study designs. It was also noted that the apparent underlying assumption of the document (seeming to favor defined generic minimal datasets for chemicals irrespective of available chemical-specific information) is seemingly contrary to current trends toward tiered toxicity testing strategies to more quickly and

efficiently identify the effect(s) of concern while reducing animal use. The International Life Sciences Institute/Health and Environmental Sciences Institute (ILSI/HESI) work on more targeted testing for pesticides was recommended as a resource (available at: http://www.hesiglobal.org/Committees/TechnicalCommittees/ACSA/ImprovedTierTestingACSA.htm).

A number of panel members recommended further analyses based on mode of action (MOA) categories as a refinement to the current analysis. Any refinement to the defaults for UF_D would first require an analysis by MOA to help identify target tissues and lifestages of concern. Panelists also suggested consideration of whether the chemicals that lie at the tails of the distributions in Figure 5 of the document are related in some manner. Panel members recommended that the document address maternal toxicity in reproductive and developmental toxicity studies, both because of the potential impact of maternal toxicity on fetal effects, and because maternal toxicity can be the critical effect. Consideration of how the analysis relates to male reproductive toxicants was also recommended. It was noted that toxicity tests conducted on non-pesticides are based at least somewhat on the expectation of specific types of toxicities (toxicity triggers), and that this targeted testing may skew the results of the analysis, thereby increasing uncertainty in interpretation.

The panel discussed several alternative mathematical approaches for evaluating the data. For example, using benchmark doses (BMDs) instead of NOAELs has several advantages, including less sensitivity to dose spacing, but practical issues affecting the use of BMDs were noted. Another suggestion was to look at the slope and severity of the endpoint. Overall, however, panelists generally felt it would be more useful in further analyses to conduct additional analyses of specific more detailed aspects, rather than conducting the same analyses using alternative mathematical approaches.

Overall, most panelists felt that it is appropriate to retain the UF_D, although updating the approach to determining the UF_D value would be useful. However, panelists felt that the specific analyses in the document were insufficient for recommending specific values for UF_D. Some panelists recommended that if UF_D is too large, that the relevant data gap should be filled by doing the study, rather than using a larger UF_D.

With regard to extrapolating the results in the document, calculated for oral studies, to the inhalation route, panelists noted that oral reproductive and developmental toxicity data can inform the choice of UFs for development of an RfC, but that this is different from saying the specific UF applies across routes.

The panel discussion of the development of CSAFs for the toxicodynamic component of the interspecies uncertainty factor for reproductive and developmental endpoints focused on the framework, rather than the case studies. The panel strongly recommended building on the ILSI/RSI (Risk Science Institute) framework document (in press in *Critical Reviews in Toxicology*) for evaluating the human relevance of effects observed in experimental animals. The ILSI framework provides a systematic structure for ensuring that MOA is adequately considered in extrapolating from animals to humans.

Several panel members noted the distinction between endpoints that are *typically* evaluated in human studies and endpoints that *could be* evaluated in a properly-designed study, and they cautioned against excluding endpoints too soon. The panelists also made a number of specific suggestions regarding how to categorize individual endpoints, including consideration of mode of action. Panel members stated that concordance in the exact endpoint being measured is rare, except possibly for hormone measurements. Adverse effects in animals can predict adverse effects in humans, but the endpoints may not be the same.

It was recommended that the authors conduct case studies with known human developmental or reproductive toxicants that have known kinetics and metabolites, known MOA, and available PBPK models. This would allow the response at a given internal dose to be compared directly between the experimental animal specie and humans. The case studies in the ILSI framework document may also serve as useful starting points.

There was considerable discussion regarding the database that would be appropriate to support development of a CSAF for interspecies differences in toxicodynamics, since the availability of human data would usually mean that the human data would be used directly in developing the RfD. However, some conditions when the CSAF could be applied were suggested. One panelist stated that a CSAF for toxicodynamics based on *in vitro* surrogates is a more likely scenario than a CSAF based on *in vivo* endpoints.

1. Participants

Authors and Presenters

Bernard Gadagbui, Ph.D., Toxicologist
Toxicology Excellence for Risk Assessment (*TERA*)

Jay Zhao, Ph.D., M.P.H., Toxicologist
Toxicology Excellence for Risk Assessment (*TERA*)

Andrew Maier, Ph.D., DABT, *VERA* Program Manager/Toxicologist
Toxicology Excellence for Risk Assessment (*TERA*)

Michael Dourson, Ph.D., DABT, Director/Toxicologist
Toxicology Excellence for Risk Assessment (*TERA*)

Peer Consultation Panel Members

Sandra Baird, Ph.D.
Menzie-Cura & Associates, Inc.

John DeSesso, Ph.D., DABFE, DABFM
Mitretek Systems, Inc.

Abby Li, Ph.D.
Exponent, Inc.

Bette Meek, M.Sc.
Health Canada

Rebecca Parkin, Ph.D.
The George Washington University Medical Center

Jennifer Seed, Ph.D.
U.S. Environmental Protection Agency

Calvin Willhite, Ph.D.
State of California

Observers and Other Attendees

A list of observers and other attendees is found in Appendix A.

2. Background

This peer consultation meeting was organized by Toxicology Excellence for Risk Assessment (*TERA)*. *TERA* is an independent non-profit organization with a mission to protect public health through the best use of toxicity and exposure information in the development of human health risk assessments. *TERA* has organized and conducted peer review and peer consultation meetings for private and public sponsors since 1996. *TERA* is further developing the peer consultation concept with various projects, particularly those evaluating risks to children such as the Voluntary Children's Chemical Evaluation Program (VCCEP), and this consultation on "The Scientific Rationale for Deriving Database and Toxicodynamic Uncertainty Factors for Reproductive or Developmental Toxicants."

This peer consultation was both sponsored and organized by *TERA*. *TERA* scientists prepared the meeting materials that are the subject of the peer consultation, while other *TERA* staff organized the consultation. To maintain separation and independence between the development of the work product and the review of it, the *TERA* scientists who prepared the meeting materials did not participate in organizing the panel meeting (i.e., selecting panelists, developing panel charge, preparing this report). *TERA* used funds from an EPA Cooperative Agreement CX-82916801 for this peer consultation.

The peer consultation panel for this meeting consisted of seven members independently selected by *TERA* based on their technical expertise in areas relevant to the subject matter under discussion: reproductive and developmental toxicology, general and probabilistic risk assessment, and epidemiology. Each panel member disclosed information regarding potential conflicts of interest and biases related to the authors or to the subject matter of the submission. *TERA* evaluated these disclosures before selecting the panel members, and the disclosures were publicly presented at the beginning of the meeting (see Appendix B for the panelist disclosure statements). The panel members received a copy of the submitted document and references approximately one month before the meeting to provide them with adequate time to review the documents and prepare for the discussions. During the peer consultation panel meeting, the panel members did not attempt to reach consensus opinions; instead, the opinions expressed by the individual panelists were summarized for this report.

Members of the public were invited to attend the peer consultation meeting and were given the opportunity to provide brief oral or written technical comments on the submitted materials. The meeting also was available to the public via webcast, and those observing the webcast were invited to submit comments or questions during the meeting via email.

This report summarizes the authors' presentations and comments, the panel discussions, and comments from the public. The meeting report is a summary, not a transcript. Panel members reviewed and commented on this draft meeting report. The authors also reviewed the draft meeting report to confirm the accuracy of their presentations and comments. This final meeting report is available to the public on the Internet at http://www.tera.org/peer/UFD/UFDWelcome.htm.

This report is organized into sections corresponding to the charge questions addressed by the panel.

3. Introductions, Conflict of Interest, and Meeting Process

The meeting opened with a welcome by Dr. Dan Briggs of *TERA*. He described the background and purpose of the peer consultation and explained that this panel was convened to offer suggestions and to provide guidance for improving the authors' document, as well as suggestions for future research. He noted that four *TERA* scientists authored and sponsored the submitted materials for this peer consultation, while other *TERA* staff selected the panel members, prepared the Panel Charge, organized the meeting, and would prepare the meeting report. *TERA* maintained a separation between these two groups to assure that the authors did not influence the panelist selection or the conduct of the consultation meeting. Dr. Briggs described *TERA's* policy on conflict of interest (COI) and bias, stating that this policy had been followed in selecting the panel members. He noted that copies of panel members' biosketches and COI/bias disclosure statements were provided to all attendees (included in Appendix B) and would be included in the meeting report. He added that no written public comments had been submitted for the meeting, but one person had requested an opportunity to present oral comments to the panel during the morning and afternoon sessions. This person would address the panel at the times designated in the Meeting Agenda. (The Agenda is included in Appendix B).

The panel members introduced themselves and stated whether they had changes or additions to their disclosure statements. No panelists had changes or additions.

Ms. Bette Meek, the panel chair, described how the meeting would be conducted. She explained that the panel discussions would be based on the items found in the Panel Charge (located in Appendix B). She noted that all panelists would have the opportunity to state their own positions on the charge items, to ask one another clarifying questions, and to further discuss the issues. No attempt would be made to reach consensus positions on the charge items. She reminded the panel that the purpose of the peer consultation was to provide guidance and suggest improvements to the document.

Dr. Michael Dourson, an author of the report, briefly addressed the panel, informing them that their comments and suggestions regarding the document that was being discussed would not be solely for the benefit of the *TERA* authors, but would be for the entire community of risk assessors. He said that various groups (e.g., *TERA* scientists or others) may use the comments and recommendations of the panelists to guide their future work.

4. Author Presentation of Database Uncertainty Factor (UF_D) Data Compilation and Support

Dr. Bernard Gadagbui presented a series of slides (included in Appendix C), providing an overview of the first part of the document (pp. 6-29), titled "Updating the Scientific Basis for the Database Uncertainty Factor." His presentation covered the history and development of database

uncertainty factors, the purpose of these factors, current approaches, analysis of the supporting data, and suggestions for possible future work.

5. Panel Discussion of Database Uncertainty Factor (UF_D) Data Compilation and Support

5.1 Clarifying Questions

Asked why this document was prepared and what hypothesis was being tested, the authors responded that some scientific groups (e.g., the National Academy of Sciences [NAS], (NRC,1993a), in their work related to the Food Quality Protection Act [FQPA]) appeared unaware that current U.S. Environmental Protection Agency (EPA) methods include an uncertainty factor for an incomplete database, including the lack of studies that test young animals (database uncertainty factor, UF_D), and that a more general approach to data gaps is used by Health Canada. The authors described how the U.S. EPA developed the UF_D concept in the 1980s based upon pesticide data and the use of default factors of 3 or 10, depending on which studies are missing. The authors also described the purpose of the current work as compiling data that could be used to evaluate the UF_D concept and be the basis of further research. The authors hypothesized that the current default UFD value of 1, 3 or 10, based on the results of key toxicity studies for 69 pesticides, is justified in light of additional knowledge on other chemicals. This hypothesis was tested by collecting similar toxicity information on an additional set of approximately 80 chemicals with different modes of action, and by varying the size of the data base considered to be "complete."

The panelists asked several questions concerning the methods used to carry out the study. In response to panelist questions, the authors stated that their compilation did not always include the principal study for developing the reference dose (RfD). Their compilation only included the five study types identified as a minimal data set, and so if the principal study was a specialized design (e.g., acetylcholinesterase inhibition), it was not included. The authors also noted that only developmental or reproductive toxicity data were compiled for the young experimental animals; data concerning maternal health status were not included in the current investigation.

5.2 Public Comments

Dr. John Lipscomb, U.S. EPA, NCEA, presented several oral comments, noting that the comments were his own and did not represent any U.S. EPA policy or position. His comments were as follows:

- The document is technically thorough and provides a sound analysis.
- Be careful not to confuse chronic toxicity data with systemic toxicity data.
- It is important to determine whether fetal effects observed in reproductive toxicity studies result from the chemical's toxicity to maternal systems or from direct interaction with the fetus or placenta.

- The document could be improved by providing the technical definition of the UF_D and by giving more explanation of the intent of the default UF_D values.

- Although the document mentions the use of benchmark dose (BMD) methods, it fails to acknowledge the impact of dose spacing on No Observed Adverse Effect Level (NOAEL) values. Considering both BMD and dose spacing may improve the data analyses.

- The document's Figure 4 indicates that NOAELs from reproductive and developmental toxicity studies are less sensitive than NOAELs from chronic rat studies. There were two analyses that had been conducted. The first (4B) contained 62 data sets, while the second (5A) contained more than twice that number, due to the addition of data sets from ATSDR. There was concern that adding these data sets may skew, or change, the results. However, in comparing Figures 4B and 5A of the document, it appears that the addition of ATSDR values does not skew the results, as the chronic effects are less sensitive than the reproductive effects in about 25-30% of the cases. The addition of the ATSDR data increased the confidence that the reproductive effects are generally comparable to or less sensitive than chronic effects.

- The analyses presented seem to justify the UF_D values, but the set of chemicals presented may not be truly representative of the chemical universe. The selected database of chemicals needs to be justified.

- Slide 10 presented by Dr. Gadagbui showed that the default UF_D values covered over 80 or 90% of chemicals. This appears to support the use of the default values. The term "coverage" is misleading in that it refers to the universe of chemicals and not to the percent of people who are covered. The document should discuss this subject more, including whether the U.S. EPA has ever specified the percent of the population intended to be protected by the selected UF_D value.

5.3 Panel Discussion of Charge Questions

5.3.1 Question 1

The data compilation presented in this report indicates that refinement of the default UF_D may be appropriate when certain animal or human toxicity data are available. Is the ratio-of-NOAELs approach (and criteria used in the report) appropriate to conduct the UF_D analysis? What are the strengths and limitations to this approach? What other approaches would you suggest might be useful to consider (e.g. Brand et al. bootstrap techniques to estimate imprecision of NOAEL ratios, benchmark dose analysis, Evans and Baird regression approach)? What are the strengths and weaknesses of these other approaches?

In beginning the panel discussion of the charge questions, the Chair observed that several of the questions are overlapping. She asked that the panel members not be distracted by this, but recommended that they focus on addressing the big issues. She noted that the public comments offered by Dr. Lipscomb raised the issue of whether the conclusions presented in the document are broadly applicable. The panel should consider this point and also consider if the results

shown in the document are chemical-specific or generic. Using default UF_D values in the absence of information may be necessary; however, risk assessors usually have at least some information on the chemical that might allow replacement of the default uncertainty factors.

Several panelists complimented the authors on the document, stating that it represented a major effort and was an important first step in a long process. They also echoed the chair's comment that, in practice, risk assessors have at least some chemical-specific information (e.g., based on structure, knowledge from other related chemicals, endpoints observed in other studies) that is used in a weight of evidence evaluation of the UF_D, and that this sort of contextual information is not acknowledged in the current analysis. Better "framing" of the objectives and nature of the analysis was suggested to address this issue, clearly describing the relevant datasets. Panelists asked how maternal toxicity was addressed, both because fetal toxicity can be secondary to maternal toxicity, and because maternal toxicity in developmental toxicity studies is sometimes the most sensitive endpoint. Panel members also requested additional explanation of the analysis method used for chemicals for which a specialized study (e.g., acetylcholinesterase inhibition for organophosphates) was used as the basis for the RfD. Panel members suggested that it could be useful to identify the study that was the basis for the RfD for each chemical and to explicitly identify the limitations of the analysis (e.g., what was and was not included) in the introduction. It was also noted that Brand et al. (2001) described the effect of the number of animals per dose and dose spacing on calculated ratios, information which might be helpful for consideration in the *TERA* compilation.

Several panel members noted that, by focusing on studies in isolation, an underlying assumption of the document appears to be that the toxicology database on each chemical needs to be greater. Such an assumption is the opposite of the current trend towards considering whether one can back off from the core dataset to a more refined dataset that will more quickly and efficiently identify the effect of concern. Panel members suggested that the "3 Rs" of toxicology testing (reduce, refine, replace) be acknowledged in the document. Work by the International Life Sciences Institute/ Health and Environmental Sciences Institute (ILSI/HESI) on pesticides was cited as an example of the current trend toward minimizing the size of toxicology data sets. One panelist stated that the HESI study is on the web (available at: http://www.hesiglobal.org/Committees/TechnicalCommittees/ACSA/ImprovedTierTestingACSA.htm), and has been accepted for publication. The purpose of the HESI study was to determine how to refine pesticide testing based on exposure information. Several panelists noted the current move toward problem formulation at an early step in the risk assessment process. Problem formulation/issue identification includes identifying the purpose of the evaluation, the population of concern (e.g., children, adults, or everyone), and the exposure scenarios, as part of determining data needs. Several panelists also noted the move by U.S. EPA, Health Canada, and others toward "intelligent testing strategies." Documentation is not yet available from the first two agencies, but a panelist recommended reviewing the documentation for the Registration Evaluation and Authorisation of Chemicals (REACH) initiative of the European Union (TAPIR, 2005). One panelist recommended that the *TERA* report note up-front that data needs were considered without the context of exposure information. Another suggested that the *TERA* analysis could be used together with the HESI document in evaluating the "expected value" of whether a new study will provide useful new information.

The panel also noted the limitations of determining ratios of NOAELs, and compared these limitations with those of determining ratios of BMDs. Several panelists stated that an advantage of comparing BMDs is that BMDs are less dependent on dose spacing. One panelist said, however, that the BMD approach is not insensitive to dose spacing, and that each approach has its own limitations. One panelist commented that the critical effect based on a BMD may be different from that based on a comparison of NOAELs. Others noted that, in practice, one often must determine the critical effect from a mixture of BMDs and NOAELs. For this reason, several panelists considered a comparison of NOAELs to be adequate. One panel member suggested that using a ratio of LOAELs for comparable effects might be useful since this approach normalizes across severity, but others questioned whether additional analyses based on LOAELs would be useful.

One panel member noted that the histograms presented in the document had large groups in the center, and smaller groups at the tails. She suggested that it would be of interest to focus on the chemicals in the tails, identifying whether there are specific modes of action (MOAs) or chemical characteristics that lead to certain types of ratios. Several other panelists also supported looking at chemical classes, but one advised caution in defining chemical classes, noting that "solvents" and "pesticides" described chemical *uses*. The different chemicals included within these usage categories might exhibit vastly differing MOAs. This issue was addressed in greater detail in the context of Charge Questions 2 and 4.

One member said that focusing on the 95^{th} percentile mixed risk assessment and risk management. He suggested presenting analyses based on the 90^{th}, 95^{th}, and 99^{th} percentiles, so that risk managers can decide which value to use.

One panelist noted that the document did not appear to use the human equivalent dose (HED). She suggested that the analysis calculate the HED values for the various studies prior to calculating the NOAEL ratios, noting that this adjustment would have the largest impact on the mouse and dog studies.

The Chair summarized the panel's discussion of Charge Question 1 with the following points:

- Frame the scope of the document in the Introduction, noting the connection between identifying a UF_D based upon the limitations of the data set and identification of where to invest efforts on testing.

- Note how risk assessors take the entire database into account in developing the UF_D, and flag the limitations of the analyses in the report. Framing the analyses in the context of how RfDs are actually developed may also inform how the analyses in the report are conducted.

- The objective of the work can be defined as looking generically at the available data to inform the default UF_D. Address the types of chemicals being considered and the nature of the effects.

- Note maternal toxicity in reproductive and developmental toxicity studies, both in terms of the effect of maternal toxicity on developmental toxicity, and because maternal toxicity may be the critical effect.

- Discuss the advantages and limitations of using different approaches, such as NOAELs vs. BMDs. The objectives of the exercise and particularly the nature of the dataset will dictate whether meaningful BMDs can be developed, and which approach is most appropriate.

- Evaluate the tails in the histograms and consider what additional work is needed to evaluate specific classes of substances. Since detailed evaluations are not possible for every chemical, it is important to look at chemical classes to inform assessments.

5.3.2 Question 2

Given the information compiled here, what conclusions can be drawn regarding continued use of the current UF_D that is based on the 1980s pesticides?

Several panel members responded to this question by stating they would prefer not to rely on the existing default UF_D values, and they would not do so if additional generic or chemical-specific information were available. Others added that, in some instances, using the default values might provide enough information to determine whether meaningful hazards existed or if further work was needed; therefore, they would be reluctant to entirely abandon use of the UF_D default values. They noted that rough calculations may be sufficient to inform those doing tolerances whether it is in their interest to do more and better research.

One member said she could not support *any* generic approach because she needed to look at the chemical-specific issues when developing an RfD. She added that if the data set were missing reproductive and developmental toxicity studies, the UF_D would become too large, so she would not accept a data set without those studies. She was not sure she can accept the current approach to UF_D.

Another panelist stated that the results in the document indicate that the risk assessment community needs to update its thinking on uncertainty factors, but she would need to think more carefully before recommending specific values. She noted that the table in Slide 10 of Dr. Gadagbui's presentation would be useful in addressing the question of whether a missing study would give lower NOAEL. This table provides information on what the "surprise factor" would be if a reproductive or developmental toxicity study were missing from the data set. Other members agreed that the document had helped inform their thinking regarding what studies were most needed and might be conducted initially, should developmental/reproductive studies be desirable, based on available data. Another pointed out that the population most exposed to the chemical often drives the life-stage of interest, which in turn defines the toxicity of concern and the type of study needed.

Although not necessarily supporting the current approach to UF_D values, one member concluded that nothing presented in the document's analyses compelled him to abandon the current default values. He said the core issue was FQPA, even though some people consider the issues raised by

the FQPA to be primarily political, or related to exposure differences and thus not affecting the RfD. Additional references regarding the history and use of UFs were noted (NRC, 1993b, 1994, 2001).

One panelist characterized the issue as a question of whether the human variability uncertainty factor (UF_H) was adequate for addressing different life stages. Another wondered if including both an uncertainty factor for both database gaps and human variability (i.e., separate factors for UF_D and UF_H) would result in "double counting" of uncertainty factors. An author clarified that uncertainty factors for identifying the correct endpoint and consideration of human variability would not necessarily include double-counting[1].

Several panelists emphasized the importance of understanding a chemical's MOA, with one member saying it was impossible to address Charge Question 2 without addressing the MOA. MOA information helps to identify the target tissues/organs and lifestages of concern. It also helps determine the sufficiency of the toxicology database, which, in turn, helps decide on the best approach for developing the UF_D. In the absence of MOA data, chemicals can be grouped based on structural similarities.

The Chair summarized the panel discussion of Charge Question 2 with the following points:

- In absence of chemical-specific data, the types of analyses conducted in the document being reviewed (based on ratios of NOAELs for specific study types) will not fully inform the choice of UF_D.

- If one must develop a UF_D factor because of limited data, using a default value of 10 seems reasonable.

- While the report provides some information regarding the relative importance of endpoints and study types, further consideration of the use of chemical-specific data and of maternal toxicity considerations is needed.

- Moving from default values to more informed UF_D values requires chemical or group-specific data on MOA.

[1] In post-meeting comments, the author clarified that double counting would not be an issue where the question is the potential for certain types of toxicity to be detected. (e.g., systemic toxicity studies do not measure developmental toxicity.) However, the author noted that there is the potential for double counting when considering systemic toxicity of adolescent rats, since this variability is covered by UF_H, but also may be considered in the context of UF_D.

5.3.3 Question 3

To what extent can uncertainty factors based upon the oral route of administration be applied to dermal or inhalation routes of exposure? Are separate data compilations needed for different routes?

One panelist said this question could not be answered generically, because the report does not address the inhalation route; others said the key issues to be considered were toxicokinetics (particularly absorption), MOA, and toxicodynamics. Another said the ability to extrapolate uncertainty factors from one route of exposure to another depended on the particular chemical and also on the species of animal in which the oral study was conducted.

Panelists acknowledged that there is a desire to stay within a specific route where possible, but route-to-route extrapolation is done in the context of RfC development, where a complete database for the oral route can fill data gaps for the inhalation route. They noted the importance, however, of consideration of MOA in doing such extrapolations. One member said that although data from oral reproductive/developmental toxicity studies can inform the RfC, and generating reproduction/developmental toxicity data for each route is not necessary, this is different from saying the specific UF_D applies across routes. Others stated that the specific analyses (and specific ratios) calculated in the document should not be extrapolated to the inhalation route, but existing data from other routes can be considered in addressing UF_D for the inhalation route.

5.3.4 Question 4

Would it be useful to further refine the data compilation to address specific chemical groups (e.g., pesticides, solvents) or durations (e.g., acute, chronic)? Can the analysis be improved in other ways to help better define the UF_D?

As discussed above in the context of Charge Questions 1 and 2, panelists agreed that it would be useful to further refine the data compilation to address specific chemical groups, but they emphasized that the chemical groups should be defined by their MOAs, rather than by their usage (as suggested in this question [i.e., pesticides, solvents]). One member noted that some chemicals have multiple MOAs. Another said that, for feasibility reasons, setting priorities for large numbers of chemicals requires that structurally-related chemicals be evaluated as a group, rather than individually. Therefore, obtaining more data on chemical groups would be helpful. One panel member suggested that it would be useful to determine whether well-studied members of different MOA classes are representative of other members of the class with regard to the calculated ratios and importance of different study types. Another noted the importance of triggers in evaluating the need for additional studies, using a tiered testing scheme.

A panelist suggested that the analysis could be improved by analyzing the ILSI/HESI data set and comparing the results with the *TERA* analysis. Other suggestions were to compare the results of different study durations that were used in the TERA analysis (e.g., consider the impact of using 1-year vs. 7-year dog studies on the calculated ratios), and to take into account the different endpoints evaluated in different study types (e.g., 1-generation vs. 2-generation), or in different generations of a 2-generation study.

5.3.5. Questions 5 and 6

Question 5

Are the suggestions for future work and analyses that might be done by investigators in this area reasonable? What other suggestions should be considered to extend this compilation?

Question 6

What other comments or recommendations do you have regarding the rationale for selection of the database uncertainty factors for reproductive and development endpoints?

The panel addressed the last two charge questions together. One panelist discussed the use of alternative computational approaches to further the current work. She noted that Evans and Baird (1998) looked at three approaches (regression, constrained regression, and ratios), and these approaches exhibited both similarities and differences from the work presented in the *TERA* document. Strengths and weaknesses of the different approaches were discussed in the Evans and Baird paper, and the importance of characterizing uncertainty was highlighted. Another member recommended that, rather than spending the time on different computational approaches, it would be more informative to conduct additional analyses based on the nature of effect. She suggested evaluating the feasibility and utility of comparing BMDs by considering the nature of the data set, since older studies are less likely to have data that can be analyzed using the BMD approach.

A member said that while the recommendations for future work made in the document were reasonable based on the data presented, the authors likely will have different recommendations after hearing the panel discussions. Another added that this *TERA* document provides a database for addressing many issues. These analyses could include (1) evaluation by MOA; (2) determining the impact of various study types after consideration of all relevant studies on the chemical (including study sub-types, such as 1-generation vs. 2-generation reproductive toxicity studies); and (3) adding considerations of maternal toxicity. The Makris et al. (1998) paper on developmental neurotoxicity may provide a nice template for how to organize the data. Consideration of the critical effect for shorter durations would be relevant for the determination of acute RfDs. The panelist recommended that the analyses should identify the endpoint and study (as well as year, since study design changed with time) that drove the RfD/RfC. In light of the U.S. EPA (2002) guidance document, it was also recommended that different exposure durations be evaluated, particularly since developmental toxicity is often a driver for acute exposure.

Other recommendations included exploring high concentration, short duration exposures, and consideration of the critical study type for male reproductive toxicants. Panel members also suggested comparing slopes of dose-response curves at the BMD, in addition to comparing NOAELs. Consideration of events occurring at the critical effect level that are unrelated to the critical effect could be useful, as well as consideration of effects of different severities occurring in reproductive toxicity studies. One member recommended first looking at all data on the subject chemical, then at all data on related chemicals, including information from testing beyond the five core studies and information on exposure. Overall, it was recommended that the authors present their ideas for a defensible minimal dataset, but in a tiered testing context.

In summarizing these suggestions, the Chair cautioned that the ultimate issue is the comparability of data. She advised:

- Be careful about conducting numerous analyses unless they truly add value.
- Keep in mind the goal of adding value to risk assessments and to priority setting.
- Avoid over-analyzing the data beyond what they can really support.

5.4 Additional Public Comments and Questions

An observer raised the issue of the potential for the data being skewed by the realm of chemicals for which data existed, particularly if testing is triggered by structural or other concerns. A panelist agreed that this is a general issue, but added that this did not apply for the pesticide data, since pesticide testing is based on checklists of required studies. Panel members and the observer stated that this potential for skewing should be acknowledged as a potential issue. It was also noted that the more in-depth evaluation of groups of chemicals should include consideration of testing triggers and their impact.

6. Author Presentation of Comparing Animal and Human Toxicodynamics for Reproductive and Developmental Endpoints

Dr. Jay Zhao used a series of slides (included in Appendix C) to provide an overview of the second part of the document (pp. 30-46), titled "Supplemental Guidance to the IPCS CSAF Methodology for Evaluating Toxicodynamic Uncertainty for Reproductive and Developmental Endpoints." He presented an overview of Chemical Specific Adjustment Factors (CSAFs) as they relate to interspecies and intraspecies variability in toxicokinetics and toxicodynamics, considerations regarding how to use information on toxicodynamic variability for developmental and reproductive endpoints for development of CSAFs, and draft charts categorizing reproductive and developmental endpoints by whether CSAFs for variability in toxicodynamics can be derived.

7. Panel Discussion of Comparing Animal and Human Toxicodynamics for Reproductive and Developmental Endpoints

7.1 Clarifying Questions

Panel members made several specific comments regarding the categorization of endpoints in frames 7 and 8 of the presentation (Figures 6 and 7 of the document). One panelist clarified that resorption and post-implantation loss are the same events. He also noted that the estrous cycle in rats is not comparable to the human menstrual cycle, although it is possible to make inferences to humans based on extrapolations of the rat data. Asked why the onset of puberty could not be measured in the epidemiology studies, the presenter responded that not taking this measurement

is an issue of feasibility, not possibility. Puberty is not commonly measured in epidemiology studies unless there is a cohort leg that is included for that purpose and followed over a long time period. Several panel members noted the distinction between endpoints that are typically evaluated in human studies and ones that could be evaluated in a properly-designed study. The presenter agreed that the categorization was based on typical availability of data. Other endpoints that were listed as "no" but which can be evaluated in humans were noted, including the fertility index (e.g., has been studied by the U.S. EPA) and ovary size. One panelist suggested separately evaluating comparability and feasibility of endpoints.

Panel members asked the authors about the relationship of the second part of the document (presented by Dr. Zhao) to the document's first part (presented by Dr. Gadagbui). They wanted to know the hypothesis for the second part of the document and said a better explanation was needed to relate the two parts of the document because they appear to be separate topics. The authors explained that the two parts of the document really are separate topics, although they are related. The second part of the document, which presents a discussion of *toxicodynamics* in interspecies extrapolation, is a continuation and expansion of a presentation on *toxicokinetics*, which was the subject of a previous peer consultation meeting with a different panel held on March 31, 2005 (see http://www.tera.org/peer/adultchildtk/actkwelcome.htm for information about this meeting, the members of the panel and the subject presented and discussed). The authors acknowledged that this fact should have been explained more clearly.

7.2 Public Comments and Panel Responses

Dr. John Lipscomb, U.S. EPA, NCEA, presented several oral comments, as he had during the morning session. He noted that these comments were his own and did not represent any U.S. EPA policy or position. His comments were as follows:

- The document represents a large amount of work and will be useful in stimulating further discussion.

- In discussing effects on sperm (pp. 33-34), clarify what is meant by direct comparison of the effective dose. Does an effective dose equate to a 10% reduction in sperm count?

- In the Case Study on lead (p. 39 and the following pages), justification of the dose metric should include consideration of pharmacokinetics. The CSAF for interspecies differences in toxicodynamics for the effects of lead on sperm count is defensible and appropriately compared an internal dose metric (lead in blood) at a fixed response level. If lead occurs in the testes, then reliance on blood lead levels in animals and humans assumes that distribution from blood to testes is the same between the species. This is likely a valid assumption.

- Developmental effects may be the result of impacts on the maternal system. The document should address whether the observed effects were direct (in the fetal/placental compartment) or indirect (in the maternal system).

Following the Public Comments and building on them, panel members made several additional remarks. The chair clarified with the authors that they were seeking input primarily on the framework for addressing toxicodynamics. The purpose of the case studies was to illustrate the issues and principles in the framework, but the authors requested that the discussion focus on the framework, rather than on the examples.

One member informed the group that an ILSI/RSI (Risk Science Institute) framework document describes an approach for using MOAs determined in animals to evaluate the human relevance of tumors (Meek et al., 2003). The framework also can apply to non-cancer effects, particularly consideration of different life-stages. ILSI developed a number of case studies under the non-cancer framework, many of which had life-stage-specific components (Seed et al., in press). The frameworks address principally qualitative considerations, in extrapolating from animal data to humans in the hazard characterization component of risk assessment. However, the outcomes of these framework analyses have important implications for subsequent more quantitative consideration of inter- and intra-species differences in the dose-response analysis, including development of CSAF, where possible. Chemical-specific data on humans are not necessarily required to address these questions because one can use knowledge of basic biology, general variability, "natural knockouts," etc. to determine whether the animal information is relevant to humans. If the MOA is known to be the same in animals and humans, the animal effects usually would be expected to predict human effects (at appropriate doses). The MOA can also provide information on what outcomes to expect in light of the known interspecies differences in prenatal/postnatal timing. Another panelist said that use of CSAFs requires sufficient understanding of MOA to provide confidence that the endpoints being measured are relevant to humans and are relevant to the toxic endpoint of interest. A systematic approach, as described by ILSI, is needed to ensure that MOA is adequately considered in development of CSAFs. Case studies that address both aspects might be helpful in informing additional development of frameworks both for MOA and CSAF.

A third panelist noted issues in extrapolating from animals to humans, stating that concordance in the exact endpoint being measured is rare. He recommended that case studies be conducted with chemicals with known kinetics and metabolites, and known MOA, where the response at a given internal dose can be compared between species. As examples, he would use chemicals with fully-developed PBPK models, such as well-characterized pharmaceuticals (e.g., retinoic acid, valproic acid), for which published data are available for humans and experimental animals. One could use the PBPK model and available toxicokinetic data, including data on trans-placental transfer. Retinoic acid could be a particularly valuable case study, since it *does* have concordance across species. In addition, not only is MOA known for retinoic acid, but the mechanism of action is known, including identification of the nuclear receptors and the genes that the receptors control.

7.3 Panel Discussion of Charge Questions

7.3.1 Question 7

Which reproductive and developmental endpoints are appropriate to use, based on feasibility of collecting the needed human data? Which reproductive and developmental endpoints are appropriate to use based on comparability of endpoints between experimental animals and humans? Is it appropriate to use surrogate endpoints?

Answering the last of these questions first, a panelist responded that all endpoints are surrogates. Since development is such a complicated process, concordance should not be expected or assumed, except possibly for hormone measurements. Adverse effects in animals can predict adverse effects in humans, but the endpoints may not be the same. For example, decreased fetal body weight in rodents is predictive of many adverse effects in humans, not only decreased birth weight. The panelist also noted that estrus in animals is not analogous to menstruation in humans, and animals have multiple corpora lutea while humans do not. Some chemicals are known to alter sex ratios in amphibians and fish, but none are known to do this in humans or in mammals. As an example of why effects might vary across species even though the MOA is the same, consider a chemical that causes cell death in rats and in humans. In rats, organogenesis is relatively short, so the result might be rat pups with three fingers instead of five. In humans, organogenesis is longer, so some repair or recovery from cell death might occur. The result might be infants without any missing fingers, but some other adverse effect(s) may have occurred.

Panelists and authors discussed the conditions under which a CSAF for toxicodynamics could be appropriately developed for reproductive and developmental endpoints. In particular, panelists questioned how there could be sufficient human data to develop a CSAF, but insufficient human data to develop the RfD directly from the human data. An author suggested that, while a robust set of human data could identify a Point of Departure (POD) for calculating an RfD, a less robust data set might serve as the basis for calculating a CSAF. He suggested that, if toxicokinetics of the chemical are understood, one could calculate the area under the time-concentration curve (AUC) for both the animal studies and the human data (based on limited dose-response data in humans), and plot AUC (or other appropriate dose metric) versus the percent response in both animals and humans. This way, even though the human data alone might not be sufficient to identify a POD, a CSAF could be calculated based on the relative animal and human AUC estimated to cause a specified response. A panelist responded that she did not think this proposed approach could replace the toxicodynamic component very often, but suggested that human data might be used to bound the data, even if the data are limited. She suggested using a "decision tree" approach, first looking at what is known about the MOA, then looking at what is known about relevance to humans (this information would not necessarily be chemical-specific), and then considering the quantitative dose-response data in animals and humans. She said this work would be challenging, and require knowing the entire chemical context along with MOA. It was also noted that a CSAF for toxicodynamics based on *in vitro* surrogates is a more likely scenario than a CSAF based on *in vivo* endpoints. In such cases, the biological basis for the endpoint should be discussed. It may also be appropriate to address the slope of the dose-response curves in the experimental animal and humans. Panel members referred the authors to the ILSI framework for evaluating relevance to humans, noting that the attempt to quantify interspecies differences would be an extension of the ILSI framework**.** One panel member also recommended that references on teratology (e.g., Schardein, 2000; or Shepard, 2001) be consulted for primary references comparing developmental toxicants in humans and laboratory animals, and that the primary references should be independently evaluated. This panelist also noted that lists of teratogens typically do not consider either magnitude of exposure or timing of exposure, and these are the critical elements in determining the likelihood of adverse outcome.

The Chair reminded the panel that they are trying to replace the default uncertainty factor of 10 for interspecies extrapolation with data, and that the differences are divided into toxicokinetic and toxicodynamic components to facilitate this effort. A member observed that the first part of the document asks how often a factor of 10 was sufficient to cover missing data, while the second part asks how to use the data to address whether a factor of 3 for interspecies differences in toxicodynamics is adequate. Another member suggested taking several data-rich examples and looking at the ratio of effective doses based on the delivered dose. An author responded that this has been done by Renwick and colleagues for systemic toxicity, but not for reproductive and developmental endpoints. These analyses were the basis for dividing the uncertainty factors for interspecies extrapolation and human variability into subfactors for kinetics and dynamics. A panelist noted that there were more data to support the default kinetic subfactor than the default dynamic subfactor; the default dynamic subfactor is what was left over to ensure consistency with the existing default uncertainty factors of 10. However, the panelist also noted that one should not aspire to the default factor of 10; rather, chemical-specific data should be used when possible and the separation of kinetics and dynamics provides a way to think about the information. Another panelist raised questions about separating kinetics and dynamics, particularly when the animal species are more sensitive, but was assured that factors less than 1 (indicating greater sensitivity of the animal model compared to humans) are acceptable. A panel member noted that it is rare that CSAFs can be developed for toxicodynamics. She emphasized that in order to develop such a CSAF, one would need to apply the ILSI framework, by determining the MOA, and ensuring that the *in vitro* endpoint being evaluated is relevant, on the pathway for the MOA, and quantitatively meaningful. It would be rare that human data are sufficient to develop a CSAF, but not sufficient to use the human data for the POD. Others noted that using the ILSI framework would help address the issue of concordant findings.

Recognizing that the case studies in the document were somewhat contrived, since reproductive endpoints did not form the critical effect, an author asked an observer (Dr. John Lipscomb) to summarize the EPA boron RfD, which was derived using CSAFs. Dr. Lipscomb responded that for boron:

- The critical effect is a developmental toxicity endpoint.
- The chemical is not metabolized.
- The toxicokinetic and toxicodynamic determinants are known in animals and in humans.
- The rate-limiting determinant for elimination is the glomerular filtration rate (GFR).
- The ratio of the GFR was used as a surrogate for ratio of clearance, and these ratios were used to replace the kinetic component for both interspecies variability and human variation. The resulting chemical-specific subfactors were 3.3 and 2 for interspecies differences in kinetics and human variability in kinetics, respectively.
- The default value of the dynamic component was used for both interspecies extrapolation and human variability.

- The total composite UF was 66, based on combining the default subfactors and CSAF values.

7.3.2 Question 8

In Figures 6 and 7, is the categorization of endpoints into Yes/No boxes for application of CSAFs for toxicodynamics appropriate? Are there better approaches for making the interspecies comparisons? Is additional information needed to support putting an endpoint into a YES/NO box?

There was no panel discussion of this Charge Question because the panel members had addressed it during their discussions of previous Charge Questions.

7.3.3 Question 9

What other factors need to be considered in comparing toxicodynamics of reproductive effects between experimental animals and humans? For example, in comparing the differences in toxicodynamic sensitivity between animals and humans, is it appropriate to compare the dose (using the correct dose metric) causing a 10% change in sperm count?

A panelist advised against using sperm count for the CSAF because of the widely differing rates of production and reservoirs among the species. What really should be measured is the effect on the rate of sperm production. Another panelist noted that there are implications to the differences in approach used to measure sperm count in rats (based on cutting open the epididymides) and in rabbits (based on ejaculate).

An author asked if a CSAF could be done when receptor binding is the key event, but produces a different toxic endpoint. One panelist responded that the ratio would still be based on the key event. Another emphasized that all key events need to be identified, and the entire pathway to the toxic endpoint needs to be understood, including both qualitative and quantitative differences, in order to develop a CSAF. Depending on the details of the pathway, the EC10 for receptor binding in this situation may or may not be appropriate for calculating a CSAF.

7.3.3 Question 10

How well do the case studies illustrate the application of the approach? Are there other chemicals that might better illustrate the concepts?

There was no panel discussion of this Charge Question because the panel members had addressed it during their discussions of previous Charge Questions.

7.3.4 Question 11

Are the suggestions for future work and analyses that might be done by investigators in this area reasonable? What other suggestions should be considered to extend this compilation?

Panelists suggested the authors consider the ILSI framework document and add more case studies. Specifically, one member recommended reviewing the list of human teratogens and reproductive toxins and look for chemicals that have PBPK models (as noted in Section 7.2). Possible chemicals include retinoic acid, valproic acid, ethanol, dibromochloropropane (DBCP),

and methotrexate, all of which have data that can be used to determine delivered dose. Anti-folates, phthalates, 5-fluorouracil, and isopropanol were also mentioned as possibilities, although issues were raised for some of these chemicals regarding the complexity of the MOA (anti-folates), whether complete PBPK models are available (5-fluorouracil), and whether the MOA is sufficiently characterized (isopropanol). The utility of looking at chemicals from different classes was also noted.

7.3.5 Question 12

What other comments or recommendations do you have regarding the development and use of CSAFs based on comparative toxicodynamics for reproductive and developmental endpoints?

Panel members did not have any additional general comments on the approach to CSAFs, but several ideas were raised about sources of data. One panel member noted a number of issues that make pooling data across states problematic. First, there is variability across registries in defining and classifying birth defects. Second, variations in funding registries have resulted in different levels of completeness and timeliness of data (e.g., sometimes the lag between birth and registration has been 5 years in some states, compared with the lag of up to 2 years that is preferred for registries). Because of these variations, she recommended using individual state registries that have achieved and sustained high quality data over an extended time period (e.g., California and Texas).

Another panelist recommended the authors consult the reproductive and developmental toxicity risk assessment guidelines prepared by the U.S. EPA (1991, 1996) and WHO/IPCS (2001). One panelist also noted that fetal death was not well characterized as written on p. 37.

8. Summary of Conclusions, Recommendations, and Suggestions

8.1 Database Uncertainty Factor

1. Frame the questions being asked at the beginning of the document. Include discussion of the limitations of the analysis and the degree to which the presented data set represents the universe of all chemicals. This framing should distinguish what is known about the relationships and what has typically been assumed in order to be pragmatic and develop values for use in human health risk assessment. Framing can also consider the exposure scenarios and exposed populations (including lifestages) of interest. Address objectives in the context of recent developments in intelligent or tiered testing strategies, which take into consideration potential exposure and predictive methodology for hazard.
2. In practice, risk assessors consider the entire database when choosing the UF_D, and this holistic consideration should be reflected in the analysis.
3. Consider the ILSI evaluation of the minimum set of animal toxicity testing needed for pesticides (depending on exposure), as well as the general trend of the "3Rs" of animal testing: reduce, refine, replace.
4. The uncertainty factor needed to cover multiple percentiles of the chemicals evaluated (e.g., 90^{th}, 95^{th}, and 99^{th} percentiles) should be presented, so that risk managers can decide which value to use.

5. Be clear about what datasets were included in the analysis. In particular, explain how the analysis was conducted for chemicals for which the RfD was based on a study other than one of the five core study types (e.g., acetylcholinesterase inhibition).
6. Address maternal toxicity in reproductive and developmental toxicity studies.
7. Consider use of Human Equivalent Doses (HEDs).
8. Consider the implications of changes over time in standard study designs. These changes affect the sensitivity of a given study type conducted in different years.
9. Refinement of the analysis based on classes of MOA would be useful.
10. Consider in further detail the chemicals that fall in the tails of the distribution of ratios of NOAELs. Evaluate whether the chemicals in the tails differ from the rest of the distribution, and if so, why.
11. Discuss the issue of using different approaches: NOAELs vs. BMDs or other approaches.
12. Panelists generally felt it would be more useful to conduct additional analyses of specific more detailed aspects, rather than repeating the analysis using other mathematical approaches.
13. Most panelists felt that it is appropriate to retain the UF_D (including the default of 10) although updating would be useful. However, panelists felt that the specific analyses in the document were insufficient for recommending specific values for UF_D.
14. Some panelists recommended that if UF_D is too large, then the relevant data gap should be filled by doing the study, rather than using a larger UF_D.
15. With regard to extrapolating these results to the inhalation route, panelists noted that oral reproductive and developmental toxicity data can inform the choice of UFs for development of an RfC, but that this is different from saying the specific UF applies across routes.
16. The Makris et al. (1998) paper on developmental neurotoxicity was suggested as a possible template for organizing the data.
17. Some alternative analyses were suggested, including looking at slope and severity of the endpoint, but the general sense was that it would be more useful to drill down into the details of the chemical-specific study data.
18. Consider how the analysis relates to male reproductive toxicants.
19. Consider the potential for skewing of the results due to testing for non-pesticides being triggered by structural or other concerns.

8.2 CSAF for Reproductive/Developmental Toxicity

1. The two parts of the document are separate analyses and should be presented as two separate papers.
2. Consider separating the analysis of the endpoints by comparability and by feasibility.
3. Some endpoints that were not considered feasible because they are not measured routinely can be measured if studies are appropriately designed.
4. The assumption of concordance is not likely valid for developmental endpoints.
5. Errors in comparability of specific endpoints were noted.
6. Use the ILSI/RSI framework in evaluating MOA and considering relevance to humans.
7. Only rarely will data be appropriate for developing a CSAF for interspecies differences in toxicodynamics. If adequate *in vivo* human data are available, they would generally be

used directly to determine the POD. However, some conditions when the CSAF could be applied based on *in vivo* data were suggested.

8. Case studies were suggested using data-rich chemicals that are known human developmental or reproductive toxicants, and for which PBPK models are available. The ILSI framework may also provided useful potential case studies.
9. In developing a CSAF, all key events and the entire pathway need to be considered.
10. Use the EPA and IPCS guidelines on reproductive and developmental toxicity to enhance the consideration of endpoints.
11. Birth defects registries are generally not consistent enough to provide useful information, with the exception of California and Texas, and registry information should not be pooled across states.

9. References

Brand, KP; Catalano, PJ; Hammit, JK; Rhomberg, L; Evans, JS. (2001). Limitations to empirical extrapolation studies: The case of BMD ratios. Risk Anal 21(4):625-640.

Evans, JS; Baird, JS. (1998). Accounting for missing data in noncancer risk assessment. HERA 4(2):291-317.

IPCS (International Programme on Chemical Safety). (2001). Environmental Health Criteria 225: Principles For Evaluating Health Risks To Reproduction Associated With Exposure To Chemicals. World Health Organization, Geneva. Available at: http://www.inchem.org/documents/ehc/ehc/ehc225.htm

Makris, S; Raffaele, K; Sette, W; Seed, J. (1998). A Retrospective Analysis of Twelve Developmental Neurotoxicity Studies Submitted to the U.S. EPA Office of Prevention, Pesticides, and Toxic Substances (OPPTS). Available at: http://www.epa.gov/scipoly/sap/1998/index.htm

Meek, ME; Bucher, JR; Cohen, SM; et al. (2003). A framework for human relevance analysis of information on carcinogenic modes of action. Crit Rev Toxicol 33:591-654.

NRC (National Research Council). (1993a). Pesticides in the Diets of Infants and Children. Commission on Life Sciences. National Research Council, National Academy Press, Washington, DC. pp. 9, 361.

NRC (National Research Council). (1993b). Issues in Risk Assessment. Commission on Life Sciences, National Research Council, National Academy Press, Washington, DC.

NRC (National Research Council). (1994). Science and Judgment in Risk Assessment. Commission on Life Sciences, National Research Council, National Academy Press, Washington, DC.

NRC (National Research Council). (2001). Standing Operating Procedures for Developing Acute Exposure Guideline Levels for Hazardous Chemicals. National Academy Press, Washington, DC. pp. 62-92.

Schardein, JL. (2000). Chemically Induced Birth Defects. 3rd ed., Revised and Expanded. Marcel Dekker, New York, NY.

Shepard, TH. (2001). Catalog of Teratogenic Agents. 10th ed. The Johns Hopkins University Press, Baltimore, MD.

Seed, J; Carney, E; Corley, R; Crofton, KM; et al. (2005) Overview: Using mode of action and life stage information to evaluate the human relevance of animal toxicity data. Crit Rev Toxicol 35: in press.

TAPIR. (2005). Three point three – A Project for the Information Requirements of REACH. Lot 3: Scoping study on the Development of a Technical Guidance Document on Information requirements on Intrinsic Properties of Substances (RIP 3.3-1). Final TAPIR Report RIP 3.3-1. Submitted to European Chemical Bureau, Joint Research Centre, European Commission under Contract No. – 22553-2004-12 F1SC ISP BE. August.

U.S. EPA (U.S. Environmental Protection Agency). (1991). Guidelines for Developmental Toxicity Risk Assessment. Fed Reg 56(234):63798-63826. December 5. Available at: http://www.epa.gov/iris/backgr-d.htm

U.S. EPA (U.S. Environmental Protection Agency). (1996). Guidelines for Reproductive Toxicity Risk Assessment. Fed Reg 61(212):56274-56322. Available at: http://www.epa.gov/iris/backgr-d.htm

U.S. EPA (U.S. Environmental Protection Agency). (2002). A review of the Reference Dose (RfD) and Reference Concentration (RfC) processes. Risk Assessment Forum, Washington, DC. EPA/630/P-02/002F. December. Available at: http://www.epa.gov/iris/backgr-d.htm